



HAL
open science

Analyse des séries chronologiques à mémoire longue dans le domaine des ondelettes

Olaf Kouamo

► **To cite this version:**

Olaf Kouamo. Analyse des séries chronologiques à mémoire longue dans le domaine des ondelettes. Statistiques [math.ST]. Télécom ParisTech, 2011. Français. NNT : . pastel-00565656

HAL Id: pastel-00565656

<https://pastel.hal.science/pastel-00565656>

Submitted on 14 Feb 2011

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

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



Thèse

présentée pour obtenir le grade de docteur

de TÉLÉCOM ParisTech

Spécialité : Signal et Images

Olaf KOUAMO

Analyse des séries chronologiques à mémoire

longue dans le domaine des ondelettes

Long Memory Time Series Analysis using the Wavelet Domain

Soutenue le 28 Janvier 2011 devant le jury composé de

Didier Dacunha-Castelle
Anne Philippe,
Jean François Coeurjolly
Jean-Marc Bardet
Céline Levy-leduc
Eric Moulines
François Roueff

Président
Rapporteurs

Examineurs

Directeurs de thèse

Table des matières

1	Introduction	12
	Introduction	12
1.1	Contexte	12
1.2	Problématique.	13
1.3	Contribution de la thèse	16
1.3.1	Test de détection de rupture dans la variance des coefficients d'ondelettes.	16
1.3.2	Inférence sur une série à longue mémoire pas nécessairement stationnaire dans le domaine des ondelettes.	17
1.3.3	Estimation robuste du paramètre de mémoire d dans le domaine des ondelettes.	18
1.3.4	Liste des publications	19
2	Processus à longue mémoire et transformée en ondelettes.	20
2.1	Processus à mémoire longue	20
2.1.1	Série chronologique stationnaire	20
2.1.2	La mémoire longue.	23
2.1.3	Définition des $M(d)$ Processus	24
2.1.4	Quelques exemples de processus à mémoire longue	26
2.2	Analyse par ondelette	32
2.2.1	Cadre général.	32
2.2.2	Transformée en ondelette discrète à temps discrets	33
2.2.3	Analyse multirésolution.	35
2.2.4	Choix de l'ondelette	40
2.2.5	Covariance et densité spectrale des coefficients d'ondelettes	41
2.2.6	Calcul récursif de la matrice de covariance des coefficients d'ondelettes	44
2.3	Estimation semi-paramétrique du paramètre de mémoire d	46
2.3.1	L'estimateur local de régression en ondelettes	47

2.3.2	Estimateur local de Whittle en ondelettes	48
2.3.3	Le mouvement Brownien fractionnaire généralisé	49
2.3.4	Approximations asymptotiques.	50
2.3.5	Estimateurs robustes du paramètre de mémoire d	53
2.3.6	Estimateur semi-paramétrique par maximum de vraisemblance local du paramètre de mémoire d	57
2.4	Test de détection de rupture	59
2.4.1	Résultats principaux	59
2.4.2	Estimation de Γ_{J_1, J_2}	62
3	Change points detection in wavelet domain	69
3.1	Introduction	69
3.2	Asymptotic distribution of the W2-CUSUM statistic.	71
3.2.1	The wavelet spectrum and the scalogram.	72
3.2.2	The single-scale case	72
3.2.3	The multiple-scale case	79
3.3	Test statistics	88
3.4	Power of the W2-CUSUM statistics	89
3.4.1	Power of the test in single scale case	89
3.4.2	Power of the test in multiple scales case	93
3.5	Some examples	95
4	Inference of a generalized long memory process in the wavelet domain	101
4.1	Introduction	101
4.2	Wavelet setting.	103
4.2.1	Multiresolution analysis	103
4.2.2	pyramidal algorithm based on approximation coefficients increments	104
4.3	Second order properties in the wavelet domain	104
4.3.1	Assumptions and notation	104
4.3.2	Iterative formula in spectral domain	106
4.3.3	Iterative formula in temporal domain	107
4.4	Application	107
4.4.1	Likelihood estimation	108
4.4.2	Local likelihood wavelet (LLW) estimator of d	109
4.4.3	Change point test statistic.	113
4.4.4	Large scale multiple change detection for stock market data	117
4.5	Postponed proofs	121

5	Central Limit Theorem for the robust wavelet regression estimator	125
5.1	Introduction	125
5.2	Definition of the wavelet-based regression estimators of the memory parameter d	127
5.2.1	The wavelet setting	127
5.2.2	Definition of the robust estimators of d	129
5.3	Asymptotic properties of the robust estimators of d	131
5.3.1	Properties of the scale spectrum estimators	131
5.3.2	CLT for the robust wavelet-based regression estimator	132
5.4	Numerical experiments	134
5.5	Application to real Data	135
5.5.1	Nile River data	136
5.5.2	Internet traffic packet counts data	137
5.6	Proofs	139
5.7	Technical Lemmas	153

Table des figures

2.1	1000 observations d'un FGN(0.4) avec un bruit blanc Gaussien $\mathcal{N}(0, 1)$. (a) correspond aux observations (b) correspond à l'auto-corrélogramme empirique et (c) correspond à l'auto-corrélogramme partiel empirique.	28
2.2	1000 observations d'un processus ARFIMA(1, d , 0) avec un bruit blanc Gaussien $\mathcal{N}(0, 1)$, $d = 0.4$ et $\phi = 0.5$. (a) correspond aux observations, (b) correspond à l'auto-corrélogramme empirique et (c) correspond à l'auto-corrélogramme partielle empirique.	31
2.3	Calcul de la transformée en ondelette décimée en utilisant l'algorithme pyramidal où h et g sont les filtres miroirs quadratures et I désigne l'identité.	41
2.4	Fonction d'échelle ϕ (gauche/rouge) et fonction d'ondelette ψ (droite/bleu). La paire (ϕ, ψ) de haut en bas correspond à l'ondelette de Haar, ondelette de Daubechies avec $M = 2$ et $M = 4$ moments nuls.	42
2.5	Boxplots des estimateurs \hat{d}_{J_1, J_2}^{LW} et \hat{d}_{J_1, J_2}^{LWW} sur des processus ARFIMA(0, d , 0), $d \in \{-0.8, -0.4, 0, 0.2, 0.6, 1, 1.6, 2, 2.6\}$. 60	
3.1	(d) corresponds to the original signal obtained by the concatenation of two AR(1) with $\phi = 0.5$ and $\phi = 0.8$ respectively. Figure (a), (b) (c) represent respectively the wavelet coefficients at scale $j = 3, 2, 1$ of the signal (d).	72
3.2	Between scale stationary process.	80
3.3	Pvalue under \mathcal{H}_0 of the distribution $D(J)$ $n = 1024$ for white noise and MA(1) processes and $n = 4096$ for AR(1) and ARFIMA(1,d,1) processes; the coarsest scale is $J = 4$ for white noise, MA and AR processes and $J = 3$ for the ARFIMA process. The finest scale is $J_1 = 1$	99
3.4	Empirical power of KSM($d = 4$) for white noise, AR, MA and ARFIMA processes. . .	99
4.1	Monte Carlo simulation to compare $\hat{d}^{LLW}(\ell)$ and $\hat{d}^{LWW}(\ell)$ in a semi parametric setting. (Top left) plot is an ARFIMA(1,-0.8,0), (top right) is an ARFIMA(1,0.6,0), (bottom left) is an ARFIMA(1,1,0) and the (bottom right) is an ARFIMA(1,2,0).	112
4.2	Monte Carlo simulation to compare $\hat{d}^{LLW}(\ell)$ and $\hat{d}^{LWW}(\ell)$ in a semi parametric setting. (Top left) plot is a DARFIMA(1,-0.4,0), (top right) is a DARFIMA(1,0.2,0), (bottom left) is a DARFIMA(1,2,0) and the (bottom right) is a DARFIMA(1,2.6,0).	113

4.3	Standard and Poor daily log returns r_t 03/01/1950-27/10/2010 (left) and Sample auto-correlation of $ r_t $ (right).	118
4.4	Figure (a) indicates the statistic $\tilde{T}_{J_1, J_2}^{k_1: k_2}(k)$ based on time index. The red dotted lines where $\tilde{T}_{J_1, J_2}^{k_1: k_2}(k)$ is maximal indicate years of changes. Figure (b) corresponds to the change points in the absolute log returns S&P 500 $ r_t $	121
5.1	Confidence intervals of the estimates $\hat{d}_{n,CL}$, $\hat{d}_{n,CR}$ and $\hat{d}_{n,MAD}$ of an ARFIMA(0, d , 0) process with $d = 0.2$ (left) and $d = 1.2$ (right) for $J_0 = 1, \dots, 7$ and $J_0 + \ell = 8$. For each J_0 , are displayed confidence interval associated to $\hat{d}_{n,CL}$ (red), interval $\hat{d}_{n,CR}$ (gren) and $\hat{d}_{n,MAD}$ (blue), respectively.	135
5.2	Empirical densities of the quantities $\sqrt{n2^{-J_0}}(\hat{d}_{*,n} - d)$, with $*$ = CL (solid line), $*$ = CR (dashed line) and $*$ = MAD (dotted line) of the ARFIMA(0,0.2,0) model without outliers (left) and with 1% of outliers (right).	135
5.3	Empirical densities of the quantities $\sqrt{n2^{-J_0}}(\hat{d}_{*,n} - d)$, with $*$ = CL (solid line), $*$ = CR (dashed line) and $*$ = MAD (dotted line) of the ARFIMA(0,1.2,0) model without outliers (left) and with 1% of outliers (right).	136
5.4	Packet counts of aggregated traffic every 1 second.	137
5.5	Confidence intervals of the estimates $\hat{d}_{n,CL}$, $\hat{d}_{n,CR}$ and $\hat{d}_{n,MAD}$ on the data Thu (left) and Sat (right) for $J_0 = 1, \dots, 14$ and $J_0 + \ell = 19$. For each J_0 , are displayed confidence interval associated to $\hat{d}_{n,CL}$ (red), interval $\hat{d}_{n,CR}$ (gren) and $\hat{d}_{n,MAD}$ (blue), respectively.	139

Liste des tableaux

3.1	Quantiles of the distribution $C(d)$ (see (3.45)) for different values of d	89
3.2	Quantiles of the distribution $D(d)$ (see (3.47)) for different values of d	89
3.3	Empirical level of KSM – CVM for a white noise.	96
3.4	Empirical level of KSM – CVM for a $MA(q)$ process.	96
3.5	Empirical level of KSM – CVM for an $AR(1)$ process.	97
3.6	Empirical level of KSM – CVM for an $ARFIMA(1, 0.3, 1)$ process.	97
3.7	Power of KSM – CVM on two white noise processes.	98
3.8	Power of KSM – CVM on a concatenation of two different MA processes.	98
3.9	Power of KSM – CVM on a concatenation of two different AR processes.	98
3.10	Power of KSM – CVM two $ARFIMA(1,d,1)$ with same AR and MA part but two different values of memory parameter d	98
3.11	Estimation of the change point and confidence interval at 95% in the generalized spectral density of a process which is obtain by concatenation of two $AR(1)$ processes.	100
3.12	Estimation of the change point and confidence interval at 95% in the generalized spectral density of a process which is obtain by concatenation of two $ARFIMA(1,d,1)$ processes.	100
4.1	Bias, standard deviation and root mean-square error for $\hat{d}^{LW}(L)$ and $\hat{d}^{LWW}(L)$ over 1000 replications for a time series generated from the Gaussian $ARFIMA(1, d, 0)$ model. The lowest RMSE among the the method and all scales considered appears in boldface.	111
4.2	Bias, standard deviation and root mean-square error for $\hat{d}^{LW}(L)$ and $\hat{d}^{LWW}(L)$ over 1000 replications for a time series generated from the Gaussian $DARFIMA(1, d, 0)$ model. The lowest RMSE among the two methods and all scales considered appears in boldface.	112
4.3	Empirical level of KSM – CVM on 15000 observations of different $ARFIMA(1, d, 0)$ processes using two estimators of the scalogram covariance matrix.	117
4.4	Change point in the S&P 500 absolute returns	119
5.1	Asymptotic relative efficiency of $\hat{d}_{n,CR}$ and $\hat{d}_{n,MAD}$ with respect to $\hat{d}_{n,CL}$ for $n = 2^{12}$ observations of an $ARFIMA(0, d, 0)$ process.	134
5.2	Estimators of d from $J_0 = 1$ to $J_0 = 17$ and $J_0 + \ell = 19$ obtained from Thu1300 and Sat1930. Here S.E denotes the standard error of $(\sqrt{n2^{-J_0}}(\hat{d} - d))$	138

Kelôk yî gâa pác hă hoûu mma mêu tiép sseuh mî di pác pác.
Le plantain qui doit mûrir, même jeter dans l'eau finira toujours par mûrir.

Some kẹ kaphẹ nzeu nep pồ.
Nul ne peut cacher les rayons du soleil.

Dédicaces

A mes Parents,

Mr et Mme KOUAMO

Papa, Maman

Vous m'avez appris le sens de la famille et des amis.

Vous m'avez appris le sens du travail et de la persévérance.

Vous m'avez toujours donné les conditions d'études dans la limite de vos moyens et un cadre familiale favorable.

Merci maman, merci papa pour tout, voyez en ce mémoire un des fruits de tous les efforts et sacrifices que vous ne cessez de faire pour moi.

A ma soeur, son époux et leurs enfants

Mr et Mme DJANANG

A tous mes petits frères

Mrs KOUAMO Stéphane, Cédric, Hermann, Ebénézer Owens et Hervé Kisito

Ce travail est aussi le fruit de vos efforts consentis. Par dessus tout, nous devons toujours rester unis comme nous le sommes aujourd'hui.

A ma seconde famille

Mr et Mme MOULINES

Eric, Corinne

Vous m'avez accueilli chez vous et fait pour moi pendant ces dernières années plus que je ne pouvais imaginer encore moins espérer.

Vous m'avez traité comme vous traitez vos propres enfants.

Je vous en serai éternellement reconnaissant et les mots me manquent pour vous remercier pour tout ce que vous avez fait pour moi.

Voyez en ce travail un dénouement positif à tous les efforts que vous avez et ne cessez de faire pour moi.

Nathalie, Vincent je vous considère simplement comme ma seconde soeur et mon cinquième frère.

A ma petite amie,

Mlle PENTKEU Armelle

Pour ton soutien et ta compréhension

Sans oublier ma grand-mère sans laquelle je ne suis rien

Remerciements

A mon directeur de thèse François Roueff

Pour avoir su raffermir ma passion pour la recherche en Statistique, pour l'intérêt qu'il porte à mes travaux, pour avoir toujours été disponible, d'oreille attentive et de conseils pointus pendant tout mon travail.

A Céline Levy-leduc

Pour son sérieux et sa disponibilité dans le travail. Pour sa collaboration qui a été très fructueuse dans mes travaux. Et enfin pour avoir accepté d'examiner ma thèse.

A Anne Philippe, Jean François Coeurjolly, Jean-Marc Bardet et Didier Dacunha-Castelle

D'avoir accepté de rapporter et d'examiner ma thèse malgré leurs nombreuses occupations.

Pour l'intérêt porté sur mes travaux.

A tout le personnel de Télécom Paristech

Pour l'accueil chaleureux qu'ils m'ont réservé pendant mon séjour dans leur site. Je pense principalement à **Gersende Fort** et **Maurice Charbit** avec lesquels j'ai eu des échanges fructueux pendant mes travaux.

A tous mes camarades doctorants et post-doctorants

Avec qui année après année on a forgé une amitié qui j'espère durera à jamais.

Enfin à Eric Moulines

Je suis confus sur le fait qu'il faille utiliser soit mon autre directeur de thèse ou mon tuteur ou simplement "**mon papa français**" comme lui même s'est surnommé.

Merci pour le soutien sans faille et la patience que tu as su m'accorder.

Je remercie toutes les personnes dont j'aurai oublié le nom et qui auraient contribué d'une manière ou d'une autre à la réalisation de ce travail

Chapitre 1

Introduction

1.1 Contexte

Une *série chronologique* est un ensemble d'observations qui est obtenu de façon séquentielle à des intervalles de temps réguliers ou non. La modélisation et la prévision des séries chronologiques est un enjeu important dans de nombreux domaines : l'économétrie financière, l'économétrie classique, les sciences sociales et médicales, la démographie, les sciences du climat, l'hydrologie, le traitement du signal.

Les objectifs pour l'étude des séries chronologiques peuvent être divisés en deux composantes principales :

- La description et la compréhension du mécanisme de production de la série, qui comprend l'analyse descriptive des données et la modélisation,
- la prévision des valeurs futures et l'estimation des risques extrêmes.

La nature intrinsèque d'une série chronologique est telle que les valeurs observées sont généralement dépendantes et l'objectif est d'identifier et de modéliser la structure de dépendance temporelle. Nous nous intéressons dans cette thèse principalement à des séries chronologiques linéaires $\{X_t, t \in \mathbb{Z}\}$, dans lesquelles l'observation à l'instant courant est supposée résulter du filtrage linéaire invariant dans le temps d'un bruit blanc stationnaire $\{Z_t, t \in \mathbb{Z}\}$ i.e $X_t = \sum_{k \in \mathbb{Z}} h_k Z_{t-k}$ avec $\sum_{k \in \mathbb{Z}} h_k^2 < \infty$. Nous considérons aussi des séries non stationnaires, mais pouvant être rendu stationnaire en différenciant la série un nombre suffisant de fois.

Basée sur la vitesse de décroissance des coefficients de la réponse impulsionnelle du filtre, les séries chronologiques linéaires peuvent être divisées en deux principales catégories. Les **processus à mémoire courte** et les **processus à mémoire longue**. Nous dirons qu'un processus linéaire est à mémoire courte si sa fonction d'auto-covariance est absolument sommable i.e $\sum_{h \in \mathbb{Z}} |\gamma(h)| < \infty$ où $\gamma(h) \stackrel{\text{def}}{=} \text{Cov}(X_0, X_h)$. On dit que le processus est à longue mémoire si $\sum_{h \in \mathbb{Z}} |\gamma(h)| = \infty$. Les processus à mémoire courte sont large-

ment utilisés dans de nombreux domaines et de multiples modèles ont été proposés : AR, ARMA, (voir Yule [1927], Walker [1931], Box and Jenkins [1962], Box and Jenkins [1976], Azencott and Castelle [1984] Brockwell and Davis [1996], Shumway and Stoffer [1999] et les références qui s’y trouvent). En revanche, les modèles permettant de représenter les processus à mémoire longue ne sont apparus qu’au début des années 50, historiquement pour l’étude du comportement inhabituel des niveaux du fleuve Nil en Égypte (voir Hurst [1951], Mandelbrot [1965], Mandelbrot and Van Ness [1968], Mandelbrot and Wallis [1968], Mandelbrot and Wallis [1969]). Les processus à mémoire longue occupent une place de plus en plus importante dans la littérature des séries chronologiques (voir Granger and Joyeux [1980], Fox and Taquq [1986], Beran [1992], Beran [1994], Samorodnitsky and Taquq [1994], Baillie [1996], Doukhan et al. [2003], Palma [2007]). En effet, les processus à mémoire longue s’avèrent plus adaptés à l’étude des séries chronologiques issues par exemple de l’hydrologie, l’économie, la climatologie, les traffics informatiques et l’économétrie financière ; (voir Hurst [1951], Lawrence and Kottegoda [1977], McLeod and Hipel [1978], Veitch and Abry [1999], Gençay et al. [2002]).

1.2 Problématique.

Dans ce travail de thèse, nous nous sommes intéressés principalement à une classe de processus à mémoire longue dont la fonction d’autocovariance varie régulièrement à l’infini avec un exposant d appelé **paramètre de mémoire**.

L’exposant d encore appelé coefficient de mémoire est lié au **coefficient de Hurst** par la relation $d = H - 1/2$. (Voir [Doukhan et al., 2003, p.5-38]).

Nous nous sommes intéressés en particulier à l’estimation du paramètre de mémoire d . Il existe plusieurs méthodes pour estimer le paramètre de mémoire d . Ces méthodes peuvent être divisées en deux principales catégories suivant les informations dont on dispose sur le processus : les méthodes paramétriques et les méthodes semi-paramétriques. Dans un cadre paramétrique, la loi du processus est supposée dépendre d’un nombre fini de paramètres. Le plus classique des modèles utilisés dans un cadre paramétrique est le modèle ARFIMA(p, d, q), qui est une extension des modèles ARIMA ; ces modèles possèdent un pôle fractionnaire au point $z = 1$ qui introduit une structure de vraisemblance à longue portée. Une autre classe de modèle, beaucoup moins populaire, sont les FEXP, qui étendent dans un cadre fractionnaire les modèles exponentiels de Chatfield. Pour estimer les paramètres de ces modèles, l’approche la plus classique est celle du maximum de vraisemblance proposée par Granger and Joyeux [1980], Hosking [1981] pour estimer le paramètre de mémoire d ainsi que les parties autoregressives et moyennes mobiles pour un modèle ARFIMA(p, d, q) stationnaire ($-1/2 < d < 1/2$). Ce résultat a été étendu par Beran [1995] au cas non stationnaire ($d \geq 1/2$). On peut également citer

la méthode du maximum de vraisemblance approchée développée dans Fox and Taqqu [1986] ou Dahlhaus [1989] par exemple, qui est basée sur une approximation spectrale de la vraisemblance Gaussienne (approximation de Whittle).

Les méthodes semi-paramétriques permettent d'estimer le paramètre de mémoire sans spécifier de façon complète la distribution du processus (et en particulier, la densité spectrale du processus). L'idée est de considérer l'estimation du paramètre de longue mémoire comme l'estimation d'un paramètre d'intérêt, en présence d'un paramètre de nuisance a priori infini-dimensionnel, qui est le spectre de la partie "courte-mémoire" du processus. La méthode la plus classique et la plus connue est celle de la régression du log-périodogramme introduite par Geweke and Porter-Hudak [1983], qui consiste à régresser le log-périodogramme par rapport au logarithme de la fréquence normalisée dans un voisinage de la fréquence nulle. La normalité asymptotique et la consistance de cet estimateur ont été établies par Robinson [1995a] pour des processus Gaussiens stationnaires. Elle a été étendue dans un cadre non-stationnaire par Hurvich and Ray [1995] (qui ont modifié le périodogramme en introduisant une différentiation et ont proposé d'appliquer une fenêtre de pondération), Velasco [1999]). Un autre estimateur couramment utilisé du paramètre de mémoire est l'estimateur local de Whittle proposé par Künsch [1987]. Au lieu de régresser le logarithme du périodogramme à la fréquence nulle, cette méthode consiste à utiliser une approximation locale de la vraisemblance Gaussienne. La consistance et la normalité asymptotique de cet estimateur ont été établies dans Robinson [1995b] pour des processus Gaussien stationnaires. Ces résultats ont été étendus par Velasco [1999] au cas des processus non stationnaire et par Phillips and Shimotsu [2004] au cas des processus racine unité et des processus présentant une tendance polynomiale. Une autre approche consiste à ajuster un modèle sur l'ensemble de la densité spectrale. Cette approche elle aussi semi-paramétrique mais globale a été considérée par Moulines and Soulier [1999] et Moulines and Soulier [2003].

Plus récemment, une nouvelle classe d'approches a été introduite, basée sur l'utilisation de la transformée en ondelettes. En effet, la *transformée en ondelette*, s'est avérée être un outil très important en statistique ; s'il a été utilisé principalement dans des problèmes de régression et d'estimation de densité pour des données indépendantes, il s'est avéré aussi intéressant pour le traitement des séries chronologiques, et en particulier pour l'analyse des processus asymptotiquement autosimilaire et/ou à mémoire longue ; voir par exemple Abry et al. [2000], Abry et al. [2003]. La transformée en ondelettes a été développée dès les années 80 (Meyer [1989]). Ces méthodes, basées sur une décomposition temps-échelle des observations permettent d'extraire des observations des informations à la fois temporelle et fréquentielle. La transformée en ondelette pour l'analyse des séries chronologiques à mémoire longue a été originellement utilisée par Flandrin [1992] pour l'étude du mouvement brownien fractionnaire et des processus Gaussiens asymptotique-

ment auto-similaires. Ces travaux ont suscité un vif intérêt et ont très rapidement été suivis d'un grand nombre de travaux de recherches dans ce domaine.

L'utilisation de la transformée en ondelette et plus précisément des scalogrammes (variance empirique des coefficients d'ondelettes) pour l'estimation du coefficient de mémoire de processus stationnaires dans un cadre semi-paramétrique, s'est imposée comme une alternative aux méthodes de Fourier. Ces méthodes ont été introduites dans Flandrin [1992] et Wornell [1993] et ont été popularisés par les travaux de Abry and Veitch [1998] et Veitch and Abry [1999], qui ont montré l'intérêt de ces méthodes d'ondelette pour l'estimation du télétrafic. L'analyse, dans un cadre semiparamétrique, de ces méthodes a été entreprise dans un travail de Bardet et al. [2000] (pour la transformée en ondelettes continues d'un processus gaussien longue mémoire lui-même continue) ; des résultats partiels pour la transformée en ondelettes discrètes ont été obtenus par Craigmile et al. [2005]. Une analyse exhaustive, dans un cadre semi-paramétrique similaire à celui utilisé pour l'analyse des estimateurs fréquentiels, a été entreprise dans une série de papiers Moulines et al. [2007a], Moulines et al. [2007b], Moulines et al. [2008] et Roueff and Taqqu [2009] : ces travaux établissent la consistance et la normalité asymptotique du paramètre de mémoire pour des processus gaussiens ou linéaires forts, dans un cadre stationnaire et non-stationnaire. Notons finalement que des résultats précis ont été obtenus, pour des processus gaussiens continus asymptotiquement autosimilaires et dans un cadre paramétrique par Coeurjolly [2001], Coeurjolly and Istas [2001], Bardet [2000] et Bardet [2002].

Par rapport aux méthodes de Fourier, les approches par ondelettes présentent un certain nombre d'avantages. D'une part, les méthodes par ondelettes se révèlent plus efficaces numériquement que les méthodes de Fourier (elles sont moins sensibles aux erreurs d'arrondis) et se prêtent bien à des calculs en ligne sur de grands ensembles de données (qui sont typiques dans l'analyse des séries à longue-mémoire). De plus, elles s'avèrent robustes à la présence de tendances déterministes et de certaines formes de non-stationnarité dans les observations. Les méthodes de Fourier pour être rendues robustes, doivent être précédées de pré-traitements des observations. Par exemple, pour éliminer la présence de tendances il est nécessaire de différencier la série initiale puis d'appliquer un fenêtrage. Ces pré-traitements ne sont pas toujours faciles à contrôler sur le plan pratique, ni à justifier sur le plan théorique (voir par exemple Hurvich et al. [2002], Hurvich et al. [2005]).

L'utilisation des scalogrammes s'est avérée très prometteuse pour l'estimation du paramètre de mémoire pour des processus intégrés non-stationnaires. Moulines et al. [2007a] ont montré que la vitesse de l'estimateur était optimale (pour une classe de mesures spectrales généralisées localement lipschitziennes dans un voisinage de l'origine) et satisfaisait un théorème de la limite centrale. Ces résultats permettent une comparaison rigoureuse des méthodes de Fourier opérant sous des conditions équivalentes (en particulier, dans un cadre semi-paramétrique). Néanmoins, un certain nombre de points d'intérêt à la fois

pratique et théorique restent à étudier.

Les travaux présentés dans ce mémoire de thèse s'inscrivent dans les problématiques suivantes

- i- Test de non stationnarité pour des processus à mémoire longue ou courte dans le domaine des ondelettes.
- ii- Calcul récursif de la matrice de covariance des coefficients d'ondelettes pour un processus à mémoire longue pas nécessairement stationnaire mais qui devient stationnaire après un nombre fini de différentiation. Ce calcul récursif a été réalisé pour proposer de nouvelles méthodes d'estimation du paramètre de mémoire, basées sur une approche de maximum de vraisemblance local.
- iii- Estimation robuste du paramètre de mémoire dans des séries chronologiques gaussiennes. L'idée est de développer des méthodes qui résistent à l'introduction de données parasites dans les observations.

Cette thèse a été effectuée en cotutelle entre l'Ecole Nationale Supérieure de Polytechnique Yaoundé et Télécom ParisTech.

1.3 Contribution de la thèse

1.3.1 Test de détection de rupture dans la variance des coefficients d'ondelettes.

Le danger de confondre la mémoire longue et la non-stationnarité a été mise en exergue par plusieurs auteurs. Trouver une réponse à cette question difficile est centrale dans de nombreux problèmes où l'existence même de phénomène à longue mémoire est sujet à de très vifs débats (en particulier pour la modélisation des débits hydrologiques d'un fleuve, des températures historiques dans l'étude des changements climatiques, mais aussi dans la modélisation des comptes des nombres de paquets dans l'étude du trafic de réseau).

Dans le chapitre 3, nous développons une procédure de test pour détecter la présence de non-stationnarité pour une classe de processus à mémoire longue (et potentiellement non stationnaires), appelés processus $M(d)$.

A la différence des méthodes proposées dans la littérature, notre test a la même distribution, que le processus soit à mémoire courte ou à mémoire longue. Cela signifie qu'il est capable de détecter la non-stationnarité pour des processus à mémoire longue et à mémoire courte sans avoir à formuler d'hypothèse a priori sur la structure de dépendance de la série. Notre procédure de test est formulée dans le domaine des ondelettes. Dans ce domaine, un changement dans la densité spectrale généralisée du processus entraîne un changement dans la variance des coefficients d'ondelettes à une ou plusieurs échelles. Notre procédure de test consiste donc à rechercher des ruptures dans la variance des coefficients

d'ondelettes aux différentes échelles de décomposition. De telles procédures de tests ont déjà été proposées par Whitcher et al. [2002]; néanmoins ces derniers ne prennent pas en compte dans leur procédure de test la dépendance des coefficients d'ondelettes à une échelle et entre plusieurs échelles. Par suite, la distribution asymptotique de leur procédure de test n'est pas correcte et, sous l'hypothèse nulle d'absence de changement (\mathcal{H}_0), le niveau de leur test est erroné. Nous avons proposé deux procédures de tests tout les deux utilisant un estimateur de la variance des scalogrammes à une ou plusieurs échelles. La distribution asymptotique du test sous \mathcal{H}_0 est rigoureusement justifiée. La puissance du test en présence d'un unique point de changement dans la densité spectrale généralisée est également étudiée.

1.3.2 Inférence sur une série à longue mémoire pas nécessairement stationnaire dans le domaine des ondelettes.

Dans ce mémoire, nous présentons des méthodes récursives donc de faibles complexités pour calculer les caractéristiques du second ordre des coefficients d'ondelettes d'une série chronologique. Nous n'exigeons pas que la série soit stationnaire. En effet, nous considérons ici des processus avec tendance polynomiale, les processus racine unité ou plus généralement, les processus à mémoire longue stationnaire ou non-stationnaires. En fait, l'algorithme que nous présentons ici s'applique à condition que

1. la transformée en ondelette considérée soit développée en utilisant l'algorithme pyramidal et
2. le nombre de moment nul M de l'ondelette soit tel que le différentié M fois noté $\Delta^M X$ de la série X soit un processus stationnaire. Dans ce cas, les coefficients d'ondelettes décimés entre les échelles sont conjointement stationnaires à condition que les coefficients d'ondelettes aux échelles les plus fines soient regroupées dans des blocs appropriés pour faire face à la décimation voir Moulines et al. [2007b].

Cette méthode est à notre connaissance, le premier algorithme rapide de calcul de la covariance des coefficients d'ondelettes, qui est un algorithme très utile pour étudier la dépendance des coefficients d'ondelettes à la même échelle et aux différentes échelles.

Nous considérons plusieurs applications de nos algorithmes.

- i- L'estimation par maximum de vraisemblance local du paramètre de mémoire d dans un cadre semi-paramétrique. L'intérêt d'un tel estimateur est de prendre en compte la dépendance des coefficients d'ondelettes à la même échelle et aux différentes échelles. Nous pourrions comparer notre estimateur à l'estimateur local de Whittle dans le domaine des ondelettes proposé par Moulines et al. [2008]. La méthode de vraisemblance locale approchée proposée par ces auteurs ne prend pas en compte la structure

de dépendance des coefficients d'ondelettes intra et inter échelles, à l'inverse de l'estimateur que nous proposons ici. Nous avons montré à travers des simulations Monte Carlo que l'erreur quadratique moyenne de l'estimateur par maximum de vraisemblance approché développé dans ce travail est en général plus faible que celle de l'estimateur de Whittle, et ce, pour une plage de valeur du coefficient de mémoire d allant au delà de l'intervalle de stationnarité $(-1/2; 1/2)$.

- ii- Le calcul d'un test de détection de rupture dans le domaine des ondelettes, en utilisant comme estimateur de la matrice de covariance des scalogrammes un estimateur obtenu par maximum de vraisemblance. Nous avons également appliqué notre procédure de test pour détecter plusieurs points de changements dans les valeurs absolues des données journalières des log retours de l'indice S&P500.

Ce travail fait l'objet du chapitre 4.

1.3.3 Estimation robuste du paramètre de mémoire d dans le domaine des ondelettes.

Dans le chapitre 5, nous proposons et étudions les propriétés asymptotiques d'estimateurs robustes du paramètre de mémoire d d'une série chronologique gaussienne. En effet, dans la plupart des données financières, climatiques et hydrologiques, on rencontre très souvent des valeurs aberrantes; d'où la nécessité de proposer des estimateurs qui soient insensibles à ces valeurs aberrantes. Nous considérons une série chronologique gaussienne X , pas nécessairement stationnaire de densité spectrale généralisée f . Cette densité spectrale généralisée est caractérisée par d et par une fonction f^* qui spécifie la dépendance courte mémoire du processus. Le cadre de notre étude est semi-paramétrique en ce sens que d et f^* sont inconnus et le paramètre de mémoire d est le seul paramètre d'intérêt. Il est estimé en régressant le logarithme d'un estimateur de la variance des coefficients d'ondelettes de la série initiale. Les deux estimateurs robustes de d que nous proposons sont basés sur deux estimateurs robustes de la variance des coefficients d'ondelettes; à savoir le carré de l'estimateur d'échelle proposé par Rousseeuw and Croux [1993] et le carré de l'*écart absolu moyen*.

Nous établissons un théorème de la limite centrale (TLC) pour ces 2 estimateurs robustes. Nous présentons également une autre méthodologie pour établir un TLC pour l'estimateur de d basé sur les scalogrammes proposés par Moulines et al. [2007a]. Nous déduisons de ces estimateurs des méthodes heuristiques permettant de choisir l'échelle.

Des simulations Monte Carlo sont ensuite présentées pour illustrer notre étude et comparer les performances de nos différents estimateurs. Les trois estimateurs sont également comparées sur des données réelles d'hydrologie du Nil et des données de trafic informatiques. Ces différentes analyses soulignent l'intérêt des estimateurs robustes pour l'esti-

mation du paramètre de mémoire des séries chronologiques gaussiennes.

1.3.4 Liste des publications

Chapitre de livre

1. "Testing for homogeneity of variance in the wavelet domain" *Lecture Notes of Statistics Springer Verlag*, pages 175-205, 2010 avec E. Moulines et F. Roueff.

Publication dans des actes de conférences internationales

1. "Long-memory versus non-stationarity : a multi-scale test procedure" *European meeting of Statistician* July 2009 Toulouse (France) avec E. Moulines et F. Roueff.
2. "Fast computation of wavelets coefficients covariance matrix of a generalized long memory process with application" *European meeting of Statistician* August 2010 Pireaus (Greece).

Articles soumis

1. "Inference of a generalized long memory process in the wavelet domain " avec M. Charbit, E. Moulines et F. Roueff, soumis à *IEEE Transactions on Signal Processing*
2. "Central limit theorem for the robust log-regression wavelet estimation of the memory parameter in the Gaussian semi-parametric context," avec C. Levy-Leduc et E. Moulines, soumis à *Bernoulli*.

Toutes les applications numériques ont été réalisées en R et Matlab et sont disponibles sur demande.

Chapitre 2

Processus à longue mémoire et transformée en ondelettes.

2.1 Processus à mémoire longue

Définition 2.1.1. *Processus aléatoire.* Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, T un ensemble d'indices et (E, \mathcal{E}) un ensemble mesurable. On appelle processus aléatoire une famille $X = \{X_t, t \in T\}$, de variables aléatoires à valeurs dans (E, \mathcal{E}) indexées par $t \in T$.

Le paramètre t ici représente le temps. Lorsque $T \subset \mathbb{Z}$, nous dirons que le processus est à temps discret ou encore que le processus est une série chronologique et lorsque $T \subset \mathbb{R}$, le processus est à temps continu. Dans la suite, on ne s'intéresse qu'aux séries chronologiques. Quant à (E, \mathcal{E}) , on considérera tout au long de ce mémoire $(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$ avec $p \geq 1$, où $\mathcal{B}(\mathbb{R}^p)$ est la tribu de Borel de \mathbb{R}^p .

2.1.1 Série chronologique stationnaire

Soit $X = \{X_t, t \in \mathbb{Z}\}$ un processus aléatoire.

Définition 2.1.2. *Stationnarité stricte.* Le processus aléatoire X est dit stationnaire au sens strict si pour tout $n \in \mathbb{N}$, pour tout n -uplet $(t_1, \dots, t_n) \in \mathbb{Z}^n$ et pour tout $h \in \mathbb{Z}$, les vecteurs $(X_{t_1}, \dots, X_{t_n})$ et $(X_{t_1+h}, \dots, X_{t_n+h})$ ont même lois.

Définition 2.1.3. *Stationnarité au second ordre.*

Un processus aléatoire X est dit stationnaire au sens large ou au second ordre si :

- X est un processus du second ordre, i.e. $\mathbb{E}[|X_t|^2] < \infty$ pour tout $t \in \mathbb{Z}$
- pour tout $t \in \mathbb{Z}$, $\mathbb{E}[X_t] = \mu$,
- $\mathbb{E}[(X_t - \mu)(X_{t+h} - \mu)] = \gamma(h)$ pour tout $(t, h) \in \mathbb{Z} \times \mathbb{Z}$. γ est appelée fonction d'auto-covariance du processus stationnaire X .

Une série stationnaire au sens strict telle que $\mathbb{E}[X_0^2] < \infty$ l'est également au sens large. La réciproque n'est pas vraie.

Définition 2.1.4. On appelle *bruit blanc (BB)* un processus aléatoire stationnaire au second ordre centré et de fonction d'auto-covariance

$$\gamma(h) = \begin{cases} \sigma^2 & \text{si } h = 0 \\ 0 & \text{sinon.} \end{cases}$$

On notera $Z = \{Z_k, k \in \mathbb{Z}\} \sim BB(0, \sigma^2)$. Si les variables aléatoires $\{Z_k, k \in \mathbb{Z}\}$ sont indépendantes et identiquement distribuées (i.i.d), de moyenne nulle et de variance σ^2 alors on dira que Z est un bruit blanc fort et on notera $Z \sim \text{IID}(0, \sigma^2)$.

Le théorème d'Herglotz montre que la fonction d'auto-covariance du processus stationnaire au sens large X notée $\gamma(\cdot)$ sont des coefficients de Fourier d'une mesure sur l'intervalle $[-\pi, \pi]$:

$$\gamma(h) = \int_{-\pi}^{\pi} e^{i\lambda h} dF(\lambda), \quad (2.1)$$

où $F(\cdot)$ est une fonction bornée, croissante, continue à droite sur $[-\pi, \pi]$ et $F(-\pi) = 0$. F est appelée fonction de distribution spectrale de γ . Si il existe une fonction $f : [-\pi, \pi] \rightarrow \mathbb{R}_+$ intégrable telle que pour tout λ dans $[-\pi, \pi]$ on ait

$$F(\lambda) = \int_{-\pi}^{\lambda} f(\nu) d\nu, \quad -\pi \leq \lambda \leq \pi,$$

alors f est appelée **densité spectrale** de $\gamma(\cdot)$. Ainsi, l'équation (2.1) peut se simplifier en $\gamma(h) = \int_{-\pi}^{\pi} e^{i\lambda h} f(\lambda) d\lambda$. Si $\sum_{h=-\infty}^{\infty} \gamma^2(h) < \infty$, alors la densité spectrale f existe et est donnée par

$$f(\lambda) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma(h) e^{-i\lambda h}.$$

La convergence de la série $\sum_{h \in \mathbb{Z}} \gamma(h) e^{-i\lambda h}$ doit être comprise au sens de $L^2([-\pi, \pi], \lambda^{\text{Leb}})$, où λ^{Leb} est la mesure de Lebesgue sur $[-\pi, \pi]$.

Si $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$ alors la fonction f est continue et bornée; c'est la densité spectrale associée à un processus courte mémoire. Dans le cas contraire, la densité spectrale f est en général non bornée, et sera la densité spectrale associée à un processus longue mémoire.

Sauf mention contraire, dans la suite, nous appellerons processus stationnaire les processus faiblement stationnaires ou stationnaire au sens large.

Filtrage linéaire

Soit X un processus stationnaire de fonction de distribution spectrale F_X et de fonction d'auto-covariance $\gamma_X(\cdot)$. Soit $h_n(\lambda) = \sum_{k=-n}^n h_k e^{-ik\lambda}$ une suite de Cauchy dans

$L^2([-\pi, \pi], F_X)$; on note $h(\lambda) = \sum_{k=-\infty}^{\infty} h_k e^{-i\lambda k}$ la limite de la suite $(h_n(\cdot))$ dans l'espace $L^2([-\pi, \pi], F_X)$. On définit $Y_t^n = \sum_{|k| \leq n} h_k X_{t-k}$. On vérifie sans aucune difficulté que pour tout $n \leq m$ appartenant à \mathbb{N} on a

$$\mathbb{E}[(Y_t^n - Y_t^m)^2] = \int_{-\pi}^{\pi} \left| \sum_{n+1 \leq |k| \leq m} h_k e^{-i\lambda k} \right|^2 dF_X(\lambda).$$

La suite (Y_t^n) est une suite de Cauchy dans $L^2(\Omega, \mathcal{F}, \mathcal{P})$ qui converge dans cet espace vers (Y_t^∞) . On a de plus

$$\mathbb{E}(Y_t^\infty Y_{t+\tau}^\infty) = \lim_{n \rightarrow \infty} \mathbb{E}(Y_t^n Y_{t+\tau}^n) = \lim_{n \rightarrow \infty} \mathbb{E}(Y_0^n Y_\tau^n) = \mathbb{E}(Y_0^\infty Y_\tau^\infty),$$

ce qui montre que le processus $Y^\infty = \{Y_t^\infty, t \in \mathbb{Z}\}$ est stationnaire. Sa fonction d'auto-covariance est donnée par

$$\mathbb{E}(Y_0^\infty Y_\tau^\infty) = \lim_{n \rightarrow \infty} \mathbb{E}(Y_0^n Y_\tau^n) = \lim_{n \rightarrow \infty} \sum_{k,l=-n}^n h_k h_l \gamma_X(\tau - k + l) = \int_{-\pi}^{\pi} |h(\xi)|^2 e^{i\xi\tau} dF_X(\xi).$$

Ainsi, la fonction de distribution spectrale du processus stationnaire Y^∞ est donnée par $F_Y(\lambda) = \int_{-\pi}^{\lambda} |h(\xi)|^2 dF_X(\xi)$.

Décomposition de Wold

Soit X un processus stationnaire et centré. On appelle $\mathcal{M} = \overline{\text{Sp}}\{X_t, t \in \mathbb{Z}\}$ l'enveloppe linéaire du processus (l'adhérence dans $L^2(\Omega, \mathcal{F}, \mathcal{P})$ des combinaisons linéaires finies $\sum_{i=1}^p a_i X_{t_i}$). On note de façon similaire $\mathcal{M}_n = \overline{\text{Sp}}\{X_t, -\infty < t \leq n\}$ et $\mathcal{M}_{-\infty} = \bigcap_{n=-\infty}^{\infty} \mathcal{M}_n$. Pour tout n , \mathcal{M}_n est un sous espace vectoriel fermé de l'espace de Hilbert \mathcal{M} . Le processus X est dit déterministe si et seulement si, pour tout $t \in \mathbb{Z}$, X_t appartient à $\mathcal{M}_{-\infty}$. Ces processus sont aussi appelés processus singuliers. Ainsi, un processus sera dit purement non déterministe ou régulier si et seulement si $\mathcal{M}_{-\infty} = \{0\}$

Théorème 2.1.1 (Décomposition de Wold). *Tout processus stationnaire X est la somme de processus régulier et de processus singulier; ces deux processus sont orthogonaux et la décomposition est unique. Ainsi, un processus stationnaire purement non déterministe peut s'écrire sous la forme*

$$X_t = \sum_{k=0}^{\infty} \psi_k Z_{t-k}, \quad (2.2)$$

où $\psi_0 = 1$, $\sum_{k \in \mathbb{Z}} \psi_k^2 < \infty$, $Z_t \sim BB(0, \sigma^2)$, et Z_t appartient à \mathcal{M}_t pour tout t . Les suites (ψ_j) et (Z_t) sont déterminées de façon unique par (2.2).

2.1.2 La mémoire longue.

Dans l'introduction, nous avons rappelé brièvement quelques caractéristiques des processus linéaires stationnaires présentant soit un comportement de courte mémoire soit un comportement de mémoire longue. Nous allons préciser ces deux notions dans ce paragraphe.

D'une manière générale, un processus est appelé processus à mémoire longue si ce processus est stationnaire au sens large et si sa fonction d'autocovariance n'est pas absolument sommable.

Même s'il semble y avoir un accord général selon lequel pour avoir la mémoire longue, les séries chronologiques doivent présenter une décroissance lente de la fonction d'autocovariance, la condition $\sum_{h=-\infty}^{\infty} |\gamma(h)| = \infty$ est trop générale pour pouvoir développer une théorie statistique pertinente. Certaines restrictions sont donc nécessaires sur la forme de la décroissance de la fonction d'autocovariance à l'infini. Ces restrictions peuvent porter soit directement sur la fonction d'autocovariance, soit sur les propriétés de la densité spectrale dans un voisinage de la fréquence nulle. Ces deux types de conditions ne sont pas équivalentes.

Définition 2.1.5. *Un processus stationnaire X de fonction d'auto-covariance γ_X est à mémoire longue au sens de la covariance si :*

$$\gamma_X(h) \sim L_1(h)h^{2d-1}, \quad \text{quand } h \rightarrow \infty, \quad (2.3)$$

où d est le paramètre de mémoire, $L_1(\cdot)$ est une fonction à variation lente à l'infini, i.e si pour tout $a > 0$, $\lim_{h \rightarrow \infty} L_1(ah)/L_1(h) = 1$ et $d < 1/2$.

Une autre définition de la longue mémoire très utilisée dans le domaine spectral est donnée par :

Définition 2.1.6. *Un processus stationnaire X de densité spectrale f_X est un processus à mémoire longue si*

$$f_X(\lambda) \sim |\lambda|^{-2d}L_2(1/|\lambda|) \quad \text{quand } \lambda \rightarrow 0 \quad (2.4)$$

où $L_2(\cdot)$ une fonction à variation lente à l'infini et $d < 1/2$.

On remarque donc que quand $d \geq 1/2$ la fonction f_X définie en (2.4) n'est pas intégrable et ne correspond donc pas à la densité spectrale d'un processus stationnaire. Ce cas sera exploré dans la suite à travers l'introduction des $M(d)$ processus et la notion de densité spectrale généralisée introduite par Yaglom [1958].

Une troisième définition est basée directement sur la décomposition de Wold du processus (2.2).

Définition 2.1.7. Soit X un processus régulier et soit $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ où $\{Z_t\}$ est un bruit blanc sa décomposition de Wold. Nous dirons que le processus X est à mémoire longue si

$$\psi_j \sim j^{d-1} L_3(j), \text{ quand } j \rightarrow \infty \quad (2.5)$$

où $L_3(\cdot)$ est une fonction à variation lente à l'infini et $d < 1/2$.

Ces trois définitions de la longue mémoire ne sont pas équivalentes à moins de rajouter des hypothèses supplémentaires.

Théorème 2.1.2. Soit X un processus stationnaire régulier de décomposition de Wold donné en (2.2)

$$X_t = \sum_{k=0}^{\infty} \psi_k Z_{t-k}$$

où $\{Z_t\}$ est une séquence de bruit blanc. Si on suppose que $0 < d < 1/2$ alors

- Si le processus X satisfait (2.5) alors il satisfait aussi (2.3).
- Si la fonction $L_1(\cdot)$ est quasi-monotone et à variation lente, alors (2.3) entraîne (2.4)

La preuve se trouve dans [Palma, 2007, Page 40]. Un ensemble de définitions alternatives sont obtenues en imposant des conditions supplémentaires sur le choix des fonctions à variations lentes $L_1(\cdot)$, $L_2(\cdot)$ et $L_3(\cdot)$. Voir Palma [2007] et les références qui s'y trouvent.

Nous opterons dans la suite pour une classe plus générale de processus pas nécessairement stationnaires mais qui peuvent être rendus stationnaires en leur appliquant un nombre fini de différentiation. Cette classe de processus comprend à la fois les processus à longue mémoire, les processus à racine-unité et les processus à courte mémoire. Nous appelons dans la suite ce type de processus $M(d)$ où d est le paramètre de mémoire.

2.1.3 Définition des $M(d)$ Processus

Soit $X = \{X_t, t \in \mathbb{Z}\}$ un processus à valeurs réelles pas nécessairement stationnaire. On note par I l'opérateur identité et par B l'opérateur retard définis respectivement par

$$[IX]_t \stackrel{\text{def}}{=} X_t \quad \text{et} \quad [BX]_t \stackrel{\text{def}}{=} X_{t-1}, \quad t \in \mathbb{Z}. \quad (2.6)$$

Nous définissons récursivement pour un entier $K \geq 1$, $B^K = B \circ B^{K-1}$. En outre, notons $[\Delta X]_t$ l'opérateur de différentiation :

$$[\Delta X]_t \stackrel{\text{def}}{=} X_t - X_{t-1}, \quad t \in \mathbb{Z}. \quad (2.7)$$

De même, nous définissons récursivement, pour un entier $K \geq 1$, la différentiation d'ordre K comme : $\Delta^K = \Delta \circ \Delta^{K-1}$.

Définition 2.1.8. *Un processus X est dit stationnaire au K -ième ordre si le processus $([\Delta^K X]_t)$ est stationnaire.*

Soit f une fonction positive $[-\pi, \pi]$ symétrique tel qu'il existe un entier K satisfaisant la condition, $\int_{-\pi}^{\pi} |1 - e^{-i\lambda}|^{2K} f(\lambda) d\lambda < \infty$.

Définition 2.1.9. *Le processus X admet comme densité spectrale généralisée f si $([\Delta^K X]_t)$ est stationnaire de densité spectrale*

$$f_K(\lambda) = |1 - e^{-i\lambda}|^{2K} f(\lambda) . \quad (2.8)$$

Cette classe de processus inclut non seulement les processus à mémoire courte et les processus à mémoire longue, mais également les processus racine unité et les processus racine unité fractionnaire.

Nous allons maintenant définir les processus admettant le paramètre de mémoire d plus simplement les *processus $M(d)$* (nous parlerons aussi de $M(d)$ -processus). On définira d'abord cette notion pour les processus stationnaires, où $d < 1/2$ et ensuite on étendra cette définition pour $d \in \mathbb{R}$.

Soit f^* une fonction positive, paire, bornée sur $(-\pi, \pi)$ et minorée par une constante strictement positive dans un voisinage de l'origine. Un processus stationnaire X est un processus $M(d)$ de paramètre de mémoire d , $-\infty < d < 1/2$ et de densité spectrale courte mémoire f^* , si sa densité spectrale est donnée par :

$$f(\lambda) \stackrel{\text{def}}{=} |1 - e^{-i\lambda}|^{-2d} f^*(\lambda), \quad \lambda \in (-\pi, \pi). \quad (2.9)$$

Pour $d \geq 1/2$, on considère les processus non-stationnaires et on étend la définition précédente, valide pour les processus stationnaires de la manière suivante.

Définition 2.1.10. *On dit que le processus X de paramètre de mémoire $d \in \mathbb{R}$ est un $M(d)$ -processus de densité spectrale courte mémoire f^* , si f^* est positive, paire, bornée sur $(-\pi, \pi)$ et minorée par une constante strictement positive dans un voisinage de l'origine et pour n'importe quel entier $K > d - 1/2$, son K -ième ordre de différentiation $([\Delta^K X]_t)$ est stationnaire de densité spectrale*

$$f_{\Delta^K X}(\lambda) = |1 - e^{-i\lambda}|^{2(K-d)} f^*(\lambda), \quad \lambda \in (-\pi, \pi) . \quad (2.10)$$

Nous excluons par cette définition la présence de composantes à variations lentes dans la densité spectrale à la fréquence nulle. Observons que $f_{\Delta^K X}(\lambda)$ dans (2.10) est intégrable si $-(K-d) < 1/2$. Observons qu'un processus $M(d)$ est stationnaire uniquement si $d < 1/2$. Pour $d > 1/2$, un processus $M(d)$ n'est pas stationnaire, mais nous pouvons néanmoins lui associer la densité spectrale généralisée

$$f_X(\lambda) = |1 - e^{-i\lambda}|^{-2d} f^*(\lambda) . \quad (2.11)$$

Le processus X sera dit :

- à longue mémoire si $0 < d < 1/2$,
- à courte mémoire si $d = 0$,
- à mémoire négative si $d < 0$; dans ce cas, le processus est non inversible si $d < -1/2$.
- Si $d > 1/2$, le processus est non stationnaire.

Remarque 2.1.1. *La définition des $M(d)$ -processus a été proposée par Hurvich and Ray [1995]. Une définition alternative pour les $M(d)$ -processus a été introduite par Robinson [1994] et utilisée plus récemment par d'autres auteurs (voir Tanaka [1999], Shimotsu and Phillips [2005], Shimotsu and Phillips [2006]). Leur définition ne repose plus sur le fait qu'il faille différencier plusieurs fois la série mais plutôt sur une troncature de l'opérateur d'intégration/différentiation fractionnaire. Les processus ainsi construits ne sont plus stationnaires mais asymptotiquement stationnaires. Les deux approches présentent des avantages et des inconvénients.*

Le paramètre de mémoire d joue un rôle très important dans la définition des $M(d)$ -processus car il caractérise le comportement de la densité spectrale généralisée $f(\lambda)$ aux basses fréquences. En effet, si on suppose que f^* est continue en zéro, alors (2.11) implique que $f(\lambda) \sim |\lambda|^{-2d} f^*(0)$ quand $\lambda \rightarrow 0$.

Permettre à d de prendre des valeurs non entières produit un changement fondamental dans la structure de corrélation d'un processus fractionnaire comparativement à la structure de corrélation d'une série chronologique standard telle que les modèles ARMA. L'étude des $M(d)$ -processus a récemment attirée l'attention parmi les théoriciens et les praticiens. En économétrie, les $M(d)$ -processus avec $d > 0$ fournissent des modèles pertinents pour de nombreuses séries macro-économiques (inflation, taux d'intérêt, taux de change, qui ne sont pas stationnaires ...) aussi bien que certaines séries financières (modélisation de la volatilité des rendements des actifs financiers, modèles de taux,...).

2.1.4 Quelques exemples de processus à mémoire longue

Dans cette section, nous étudions successivement le mouvement brownien fractionnaire (FBM), le bruit Gaussien fractionnaire (FGN) et le processus ARFIMA qui sont les modèles les plus simples présentant la longue mémoire.

Le Mouvement Brownien fractionnaire (FBM)

Le FBM noté $B^H(t)$ et défini sur \mathbb{R} est un processus Gaussien à trajectoires continues centré et de fonction d'auto-covariance

$$\mathbb{E} (B^H(t)B^H(s)) = \frac{\sigma^2}{2} (|t|^{2H} + |s|^{2H} - |t - s|^{2H}), \quad t, s \in \mathbb{R}$$

où $H \in (0, 1)$, est le paramètre de Hurst et σ^2 la variance d'innovation. La valeur de H détermine la nature du FBM. Ainsi,

- si $H = 1/2$, le processus est en fait un mouvement brownien standard ;
- si $H > 1/2$, les incréments du processus sont positivement corrélés ;
- si $H < 1/2$, les incréments du processus sont négativement corrélés.

Ce processus a été introduit par Mandelbrot and Van Ness [1968] pour expliquer le *phénomène de Hurst* (persistance des périodes de crue et de décrue des débits du fleuve Nil) observé par Hurst [1951] sur les données du Nil. Le FBM est un processus auto-similaire i.e $B^H(at) \stackrel{\mathcal{L}}{=} |a|^H B^H(t)$, où $\stackrel{\mathcal{L}}{=}$ signifie égalité des lois fini-dimensionnelles. Notons $(B^H(k))$, $k \in \mathbb{Z}$, les échantillons du FBM $B^H(t)$, $t \in \mathbb{R}$ aux instants entiers avec $H \in (0, 1)$. Les incréments du processus $B^H(k)$ sont stationnaires ; sa densité spectrale généralisée est donnée à une constante multiplicative près par (voir Samorodnitsky and Taqqu [1994])

$$f_{\text{FBM}}(\lambda) \stackrel{\text{def}}{=} \sum_{k=-\infty}^{\infty} |\lambda + 2k\pi|^{-2H-1}, \quad \lambda \in (-\pi, \pi).$$

On peut l'exprimer sous la forme (2.11),

$$f_{\text{FBM}}(\lambda) = |1 - e^{-i\lambda}|^{-2d} f_{\text{FBM}}^*(\lambda), \quad (2.12)$$

en posant $d = H + 1/2 \in (1/2, 3/2)$ et

$$f_{\text{FBM}}^*(\lambda) = \left| \frac{2 \sin(\lambda/2)}{\lambda} \right|^{2H+1} + |2 \sin(\lambda/2)|^{2H+1} \sum_{k \neq 0} |\lambda + 2k\pi|^{-2H-1}. \quad (2.13)$$

Le Bruit Gaussien Fractionnaire (FGN)

Un processus $X = \{X_k, k \in \mathbb{Z}\}$ est un FGN s'il se définit comme les incréments d'un FBM aux instants entiers à savoir

$$X_k = B_H(k+1) - B_H(k) \quad k \in \mathbb{Z},$$

où $(B^H(t))$ est un mouvement Brownien fractionnaire de paramètre de Hurst $H \in [0, 1)$. C'est un processus Gaussien centré stationnaire dont la fonction d'auto-covariance $\gamma(h) = \mathbb{E}(X_k X_{k+h})$ est donnée par

$$\gamma(h) \stackrel{\text{def}}{=} \frac{\sigma^2}{2} \{|h+1|^{2H} + |h-1|^{2H} - 2|h|^{2H}\} \quad h \in \mathbb{Z}, \quad (2.14)$$

où H et σ^2 sont définis comme précédemment. Le comportement asymptotique de la fonction d'auto-covariance γ d'un FGN est donné par la relation

$$\gamma(h) \sim \sigma^2 H(2H-1)h^{2H-2} \quad \text{as } h \rightarrow \infty, \quad (2.15)$$

quand $H \neq 1/2$. Si on note d le paramètre de mémoire, alors, $d = H - 1/2$; d'où dans le cas du FGN, d appartient à $(-1/2, 1/2)$ et de façon équivalente (2.15) se réécrit

$$\gamma(h) \sim \sigma^2 d(2d+1)|h|^{2d-1} \quad \text{as } h \rightarrow \infty.$$

Comme $\gamma(h) = 0$ pour $h \geq 1$ quand $H = 1/2$ ou de façon équivalente quand $d = 0$, le processus X_k est un bruit blanc dans ce cas. La densité spectrale du FGN est donnée par

$$f(\lambda) = C_H \left(2 \sin \frac{\lambda}{2}\right)^2 \sum_{k=-\infty}^{\infty} \frac{1}{|\lambda + 2\pi k|^{2H+1}} \sim C_H |\lambda|^{1-2H} \text{ quand } \lambda \rightarrow 0, \quad (2.16)$$

où C_H est une constante. Comme application, on représente 1000 observations simulées d'un FGN avec $\sigma^2 = 1$ et $d = 0.4$. On observe une décroissance lente de la courbe d'autocorrélogramme empirique comme attendu.

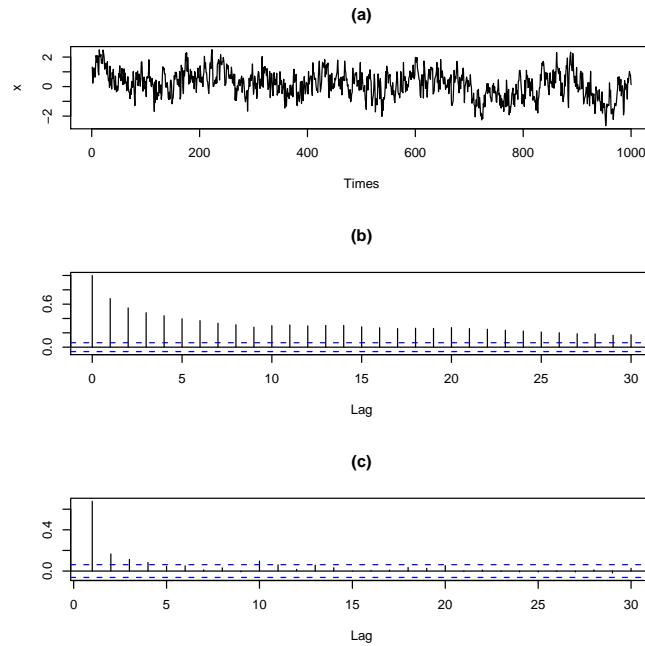


FIG. 2.1 – 1000 observations d'un FGN(0.4) avec un bruit blanc Gaussien $\mathcal{N}(0,1)$. (a) correspond aux observations (b) correspond à l'auto-corrélogramme empirique et (c) correspond à l'auto-corrélogramme partiel empirique.

Les processus ARFIMA

Un autre exemple de processus à mémoire longue très répandu est le processus auto-régressif fractionnaire intégré et à moyenne mobile dit processus ARFIMA(p, d, q) où p représente le nombre de coefficient de la partie AR, d le paramètre de mémoire et q le nombre de coefficients de la partie MA. Ce processus est devenu très populaire depuis son introduction au début des années 1980, principalement à cause de sa structure mathématique relativement simple et de ses nombreux points communs avec les modèles ARMA, qui restent le modèle prédominant pour l'analyse des séries chronologiques linéaires. Le processus a été originellement proposé par Granger and Joyeux [1980] et Hosking [1981] comme une extension des modèles ARIMA au cas où les valeurs de

d appartiendrait à $(-1/2, 1/2)$. Un processus ARFIMA(p, d, q) (Y_t) peut se définir par l'équation aux différences (fractionnaires)

$$\phi(B)Y_t = \theta(B)(I - B)^{-d}Z_t, \quad (2.17)$$

où $\phi(B) = 1 + \phi_1 B + \dots + \phi_p B^p$ et $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$ sont respectivement les opérateurs auto-régressif et moyenne mobile; $\phi(B)$ et $\theta(B)$ n'ont pas de zéros en commun et ont tous leurs racines en dehors du cercle unité. $(I - B)^{-d}$ est un opérateur fractionnaire de différentiation ($d \in [0, 1/2)$) ou d'intégration ($d \in \mathbb{R}^-$, défini par l'expansion binomiale

$$(I - B)^{-d} = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}, \quad (2.18)$$

pour $d < \frac{1}{2}$, $d \neq 0, -1, -2, \dots$. B est l'opérateur retard défini en (2.6) et (Z_t) est un bruit blanc de variance finie. Notons que les processus ARMA ou ARIMA peuvent être considérés comme des cas particuliers des processus ARFIMA avec $d = 0$ et $d = 1, 2, \dots$ respectivement.

Le théorème suivant qui est prouvé dans Palma [2007] établit l'existence et l'unicité d'une solution stationnaire du processus ARFIMA défini par (2.17), ainsi que sa causalité et son inversibilité. Pour les définitions de processus causal et inversible voir Brockwell and Davis [1996].

Théorème 2.1.3. *Considérons un processus ARFIMA(p, d, q) défini par (2.17). Supposons que les polynômes $\phi(\cdot)$ et $\theta(\cdot)$ n'ont aucun zéros en commun et que $d \in (-1, \frac{1}{2})$. Alors,*

- si les zéros de $\phi(\cdot)$ sont à l'extérieur du cercle unité $\{z : |z| = 1\}$, alors il existe une unique solution stationnaire pour (2.17) donnée par

$$Y_t = \sum_{j \in \mathbb{Z}} \psi_j Z_{t-j},$$

où $\psi(z) = (1 - z)^{-d} \theta(z) / \phi(z)$.

- Si les zéros de $\phi(\cdot)$ sont à l'extérieur du disque unité $\{z : |z| \leq 1\}$, alors la solution Y est causale.
- Si les zéros de $\theta(\cdot)$ sont à l'extérieur du disque unité $\{z : |z| \leq 1\}$, alors la solution Y est inversible.

Remarque 2.1.2. *Notons que dans le Théorème 2.1.3, la condition $d \in (-1, \frac{1}{2})$, étend la condition usuelle de stationnarité et d'inversibilité $d \in (-\frac{1}{2}, \frac{1}{2})$ des précédents résultats voir par exemple [Hosking, 1981, Théorème 2] et [Brockwell and Davis, 1987, Théorème 13.2.2 et Remarque 7]. [Bloomfield, 1985, Section 4] établit l'inversibilité d'un bruit fractionnaire (ARFIMA(0, d , 0)) pour $d \in (-1, \frac{1}{2})$ et l'extension de ce résultat aux processus ARFIMA(p, d, q) est due à Bondon and Palma [2007].*

D'après le théorème 2.1.3, si les racines des polynômes $\phi(B)$ et $\theta(B)$ sont à l'extérieur du disque unité et $d \in (-1, 1/2)$, alors le processus ARFIMA(p, d, q) est stationnaire causal et inversible. Dans ce cas, on a

$$Y_t = (I - B)^{-d} \phi(B)^{-1} \theta(B) Z_t = \psi(B) Z_t \quad (2.19)$$

et

$$Z_t = (I - B)^d \phi(B) \theta(B)^{-1} Y_t = \pi(B) Y_t. \quad (2.20)$$

Les coefficients ψ_j et π_j satisfont les relations asymptotiques suivantes.

$$\psi_j = \frac{\theta(1)j^{d-1}}{\phi(1)\Gamma(d)} + O(j^{-1}) \text{ et } \pi_j = \frac{\phi(1)j^{-d-1}}{\theta(1)\Gamma(-d)} + O(j^{-1}), \quad (2.21)$$

quand $j \rightarrow \infty$. Voir [Kokoszka and Taquq, 1995, Corollaire 3.1]. Les représentations (2.19) et (2.20) du processus sont respectivement appelées représentation MA(∞) et représentation AR(∞) du processus (Y_t).

Dans le cas où le processus ARFIMA(p, d, q) est stationnaire i.e lorsque $d < 1/2$, il admet une densité spectrale donnée pour tout λ dans $(-\pi, \pi)$ par

$$f(\lambda) = \frac{\sigma^2}{2\pi} |1 - e^{-i\lambda}|^{-2d} \frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2}. \quad (2.22)$$

Dans le cas où $d \geq 1/2$, f définie en (2.22) est appelée densité spectrale généralisée du processus ARFIMA. Voir Hosking [1981], Bloomfield [1985] Brockwell and Davis [1996], Palma [2007] pour plus de détails sur ces processus.

Le processus ARFIMA(0, d , 0)

Le processus ARFIMA(0, d , 0) noté (Y_t) encore appelé bruit fractionnaire FN(d, σ^2) où d est le paramètre de mémoire ou paramètre de différenciation fractionnaire et σ^2 la variance de l'innovation est une classe particulière de processus ARFIMA(p, d, q) dans laquelle les polynômes $\phi(B)$ et $\theta(B)$ des parties AR et MA sont réduits à l'opérateur identité I . Ainsi, il se définit par

$$Y_t = (I - B)^{-d} Z_t,$$

où $(I - B)^{-d}$ est défini en (2.18) et (Z_t) un bruit blanc de variance σ^2 . Les représentations MA(∞) et AR(∞) du processus ARFIMA(0, d , 0) sont données respectivement par

$$Y_t = (I - B)^{-d} Z_t = \psi(B) Z_t \quad \text{et} \quad Z_t = (I - B)^d Y_t = \pi(B) Y_t, \quad (2.23)$$

où les coefficients ψ_j et π_j satisfont les relations suivantes :

$$\psi_j = \prod_{k=1}^j \frac{k-1+d}{k} = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}, \quad \pi_j = \prod_{k=1}^j \frac{k-1-d}{k} = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}. \quad (2.24)$$

Pour $d \in [-\frac{1}{2}, \frac{1}{2})$, le processus ARFIMA(0, d , 0) a pour densité spectrale

$$f(\lambda) = \frac{\sigma^2}{2\pi} |1 - e^{-i\lambda}|^{-2d}, \quad \lambda \in (-\pi, \pi). \quad (2.25)$$

Pour $d \in (-1/2, 1/2)$, le FN(d, σ^2) est stationnaire et inversible, et est un bruit blanc pour $d = 0$. Pour $d = -1/2$, le processus est stationnaire mais non inversible. Si $d \geq 1/2$ dans (2.25), on obtient une classe de processus non stationnaires qui deviennent stationnaires après $\lfloor d + 1/2 \rfloor$ différentiation. Pour $d \in [-1/2, 1/2)$, la fonction d'auto-covariance est donnée par (Beran [1994] et Hosking [1981] pour le cas $d = -1/2$) :

$$\gamma(k) = \frac{\sigma^2}{2\pi} \frac{(-1)^k \Gamma(1 - 2d)}{\Gamma(1 - d + k) \Gamma(1 - d + k)}. \quad (2.26)$$

Comme illustration, on représente 1000 observations simulées d'un ARFIMA(1, d , 0) avec $d = .4$, $\phi = 0.5$ et la variance d'innovation $\sigma^2 = 1$. On observe comme dans le cas du

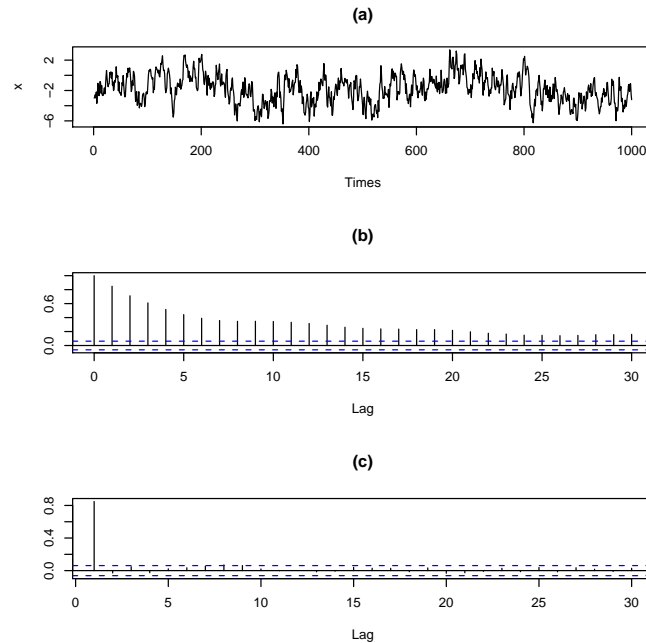


FIG. 2.2 – 1000 observations d'un processus ARFIMA(1, d , 0) avec un bruit blanc Gaussien $\mathcal{N}(0, 1)$, $d = 0.4$ et $\phi = 0.5$. (a) correspond aux observations, (b) correspond à l'auto-corrélogramme empirique et (c) correspond à l'auto-corrélogramme partielle empirique.

FGN une décroissance lente de l'auto-corrélogramme empirique.

2.2 Analyse par ondelette

2.2.1 Cadre général.

Dans cette sous-section, nous introduisons les concepts de base requis pour définir la transformée en ondelettes discrète ¹ Soit ϕ et ψ deux fonctions de $L^2(\mathbb{R})$

$$\widehat{\phi}(\xi) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \phi(t)e^{-i\xi t} dt \quad \text{et} \quad \widehat{\psi}(\xi) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \psi(t)e^{-i\xi t} dt .$$

Considérons les hypothèses suivantes:

(W-1) ϕ et ψ sont à support compact, intégrables, et $\widehat{\phi}(0) = \int_{-\infty}^{\infty} \phi(t) dt = 1$ et $\int_{-\infty}^{\infty} \psi^2(t) dt = 1$.

(W-2) Il existe $\alpha > 1$ tel que $\sup_{\xi \in \mathbb{R}} |\widehat{\psi}(\xi)| (1 + |\xi|)^\alpha < \infty$.

(W-3) La fonction ψ possède M moments nuls, *i.e.* $\int_{-\infty}^{\infty} t^m \psi(t) dt = 0$ pour tout $m = 0, \dots, M - 1$.

(W-4) La fonction $\sum_{k \in \mathbb{Z}} k^m \phi(\cdot - k)$ est un polynôme de degré m pour tout $m = 0, \dots, M - 1$.

La condition (W-1) implique que $\widehat{\psi}$ et $\widehat{\phi}$ sont infiniment différentiables. La condition (W-2) assure que la transformée de Fourier $\widehat{\psi}$ décroît rapidement vers 0 (condition de régularité). La condition (W-3) assure que ψ oscille et que son produit scalaire avec des polynômes de degré $m = 0, \dots, M - 1$ s'annule; ou de façon équivalente que les $M - 1$ premières dérivées de $\widehat{\psi}$ s'annulent à l'origine. Ainsi en utilisant un développement de Taylor, on obtient

$$|\widehat{\psi}(\lambda)| = O(|\lambda|^M) \quad \text{quand} \quad \lambda \rightarrow 0 . \quad (2.27)$$

Par [Cohen, 2003, Théorème 2.8.1, Page 90], (W-4) est équivalent à

$$\sup_{k \neq 0} |\widehat{\phi}(\lambda + 2k\pi)| = O(|\lambda|^M) \quad \text{quand} \quad \lambda \rightarrow 0 . \quad (2.28)$$

Comme indiquées ci-dessous, les conditions (W-3)-(W-4) impliquent que la transformée en ondelette des polynômes de degré $M - 1$ s'annule. Ces deux conditions sont souvent assimilées à la condition d'admissibilité. Adoptant la convention classique en traitement du signal qui stipule que les grandes valeurs de l'indice d'échelle j correspondent aux échelles grossières, (basses fréquences), on définit les familles $\{\phi_{j,k}, j \in \mathbb{Z}, k \in \mathbb{Z}\}$ et $\{\psi_{j,k}, j \in \mathbb{Z}, k \in \mathbb{Z}\}$ de fonctions translatées et dilatées .

$$\phi_{j,k}(t) = 2^{-j/2} \phi(2^{-j}t - k) \quad \text{et} \quad \psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k). \quad (2.29)$$

¹Le terme *transformée en ondelette* est un abus de langage, comme nous le verrons dans la suite.

2.2.2 Transformée en ondelette discrète à temps discrets

Ayant définies les fonctions ϕ et ψ , nous définissons maintenant ce que nous appelons transformée en ondelette à temps discrets et l'utilisons pour calculer les coefficients d'ondelettes et les coefficients d'échelles d'une série chronologique X .

Considérons une séquence de valeurs réelles $\mathbf{X} = \{X_k, k \in \mathbb{Z}\}$. Nous voulons construire un processus à temps continu à partir d'un processus à temps discret. En utilisant la fonction d'échelle ϕ , on commence par associer à la séquence \mathbf{X} les fonctions continues suivantes

$$\mathbf{X}_n(t) \stackrel{\text{def}}{=} \sum_{k=1}^n X_k \phi(t-k) \quad \text{et} \quad \mathbf{X}(t) \stackrel{\text{def}}{=} \sum_{k \in \mathbb{Z}} X_k \phi(t-k), \quad t \in \mathbb{R}. \quad (2.30)$$

La définition de la fonction \mathbf{X}_n n'exige que la connaissance des valeurs X_1, \dots, X_n ; à l'opposé, la définition de la fonction \mathbf{X} requiert la connaissance de toute la trajectoire du processus $\{X_k, k \in \mathbb{Z}\}$. Sans perte de généralité, nous supposons que les supports des fonctions d'échelle ϕ et d'ondelette ψ sont respectivement inclus dans $[-T, 0]$ et $[0, T]$, pour n'importe quel entier $T \geq 1$. Ceci implique que $\mathbf{X}_n(t) = \mathbf{X}(t)$ pour tout $t \in [0, n-T+1]$ et que les supports de $\psi_{j,k}$ et $\phi_{j,k}$ sont inclus dans les intervalles $[2^j k, 2^j(k+T)]$ et $[2^j(k-T), 2^j k]$, respectivement.

Les coefficients d'ondelettes $W_{j,k}$ et les coefficients d'échelles $A_{j,k}$ à l'échelle j et en position k en évaluant le produit scalaire de la fonction $t \mapsto \mathbf{X}(t)$, et des fonctions $t \mapsto \psi_{j,k}(t)$ et $t \mapsto \phi_{j,k}(t)$ respectivement

$$W_{j,k} \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \mathbf{X}(t) \psi_{j,k}(t) dt$$

$$A_{j,k} \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \mathbf{X}(t) \phi_{j,k}(t) dt$$

pour tout $j \geq 0$ et $k \in \mathbb{Z}$. Remarquons que

$$W_{j,k} = \int_{-\infty}^{\infty} \mathbf{X}_n(t) \psi_{j,k}(t) dt, \quad (2.31)$$

quand $[2^j k, 2^j(k+T)] \subseteq [0, n-T+1]$, *i.e.* pour tout $(j, k) \in \mathcal{I}_n$, où

$$\mathcal{I}_n \stackrel{\text{def}}{=} \{(j, k) : j \geq 0, 0 \leq k < n_j\} \quad \text{avec} \quad n_j = \lfloor 2^{-j}(n-T+1) - T + 1 \rfloor. \quad (2.32)$$

De même

$$A_{j,k} = \int_{-\infty}^{\infty} \mathbf{X}_n(t) \phi_{j,k}(t) dt, \quad (2.33)$$

quand $[2^j(k-T), 2^j k] \subseteq [0, n-T+1]$, *i.e.* pour tout $(j, k) \in I_n$, où

$$I_n \stackrel{\text{def}}{=} \{(j, k) : j \geq 0, T \leq k < m_j\} \quad \text{avec} \quad m_j = \lfloor 2^{-j}(n-T+1) + 1 \rfloor. \quad (2.34)$$

En d'autres termes, \mathcal{I}_n et I_n désignent l'ensemble des indices (j, k) pour lesquelles les coefficients d'ondelettes $W_{j,k}$ et les coefficients d'échelles $A_{j,k}$ dependent uniquement des

observations X_1, \dots, X_n . Si le nombre d'observations n augmente, ces coefficients restent inchangés et de nouveaux coefficients sont calculés. Par conséquent, la définition des coefficients d'ondelettes et d'échelle ne dépend pas du nombre d'observations disponibles. Seul le nombre de coefficients effectivement calculables dépend du nombre d'observations. Ceci est un avantage par rapport à l'analyse de Fourier, où la définition des coefficients de Fourier discrète à une fréquence donnée change quand le nombre d'observation augmente.

Mise en oeuvre pratique.

En pratique, les coefficients d'ondelettes $W_{j,k}$ et les coefficients d'échelles $A_{j,k}$ ne sont pas calculés en utilisant (2.31) et (2.33) mais par filtrage linéaire et décimation. En effet, les coefficients d'ondelettes $W_{j,k}$ et d'échelles $A_{j,k}$ peuvent se récrire sous la forme

$$W_{j,k} = \sum_{l \in \mathbb{Z}} X_l h_{j,2^j k - l} = (h_{j,\cdot} \star \mathbf{X})_{2^j k} = (\downarrow^{2^j} [h_{j,\cdot} \star \mathbf{X}])_k, \quad (j, k) \in \mathcal{I}_n, \quad (2.35)$$

et

$$A_{j,k} = \sum_{l \in \mathbb{Z}} X_l b_{j,2^j k - l} = (b_{j,\cdot} \star \mathbf{X})_{2^j k} = (\downarrow^{2^j} [b_{j,\cdot} \star \mathbf{X}])_k, \quad (j, k) \in I_n, \quad (2.36)$$

où, pour tout $j \geq 0$, les réponse impulsionnelles $h_{j,\cdot}$ et $b_{j,\cdot}$ sont respectivement définies par

$$h_{j,l} \stackrel{\text{def}}{=} 2^{-j/2} \int_{-\infty}^{\infty} \phi(t+l) \psi(2^{-j}t) dt, \quad l \in \mathbb{Z}, \quad (2.37)$$

et

$$b_{j,l} \stackrel{\text{def}}{=} 2^{-j/2} \int_{-\infty}^{\infty} \phi(t+l) \phi(2^{-j}t) dt, \quad l \in \mathbb{Z}, \quad (2.38)$$

où \star désigne la convolution des suites et \downarrow^{2^j} est l'opérateur de sous-échantillonnage par un facteur de 2^j défini, pour n'importe quelle suite $\{c_k\}_{k \in \mathbb{Z}}$, par $(\downarrow^{2^j} c)_k = c_{k2^j}$. Comme ϕ et ψ sont à support compact, les fonctions de transfert associées H_j et B_j sont respectivement les polynômes trigonométriques,

$$H_j(\lambda) = \sum_{l \in \mathbb{Z}} h_{j,l} e^{-i\lambda l} = \sum_{l=-T(2^j+1)+1}^{-1} h_{j,l} e^{-i\lambda l} \quad (2.39)$$

et

$$B_j(\lambda) = \sum_{l \in \mathbb{Z}} b_{j,l} e^{-i\lambda l} = \sum_{l=-(T-1)}^{T2^j-1} b_{j,l} e^{-i\lambda l}. \quad (2.40)$$

Pour tout $m = 0, \dots, M-1$, sous l'hypothèse (W-4), $t \mapsto \sum_{l \in \mathbb{Z}} \phi(t+l) l^m$ est un polynôme de degré m et (W-3) entraîne alors que, pour tout $j \geq 0$,

$$\sum_{l \in \mathbb{Z}} h_{j,l} l^m = 2^{-j/2} \int_{-\infty}^{\infty} \psi(2^{-j}t) \sum_{l \in \mathbb{Z}} \phi(t+l) l^m dt = 0. \quad (2.41)$$

Considérons maintenant $P_j(x) = \sum_{l \in \mathbb{Z}} h_{j,l} x^l$. On remarque aisément que la condition (2.41) entraîne que $P_j(1) = 0, P'_j(1) = 0, \dots, P_j^{(M-1)}(1) = 0$. D'où $H_j(\lambda) = P_j(e^{-i\lambda})$ peut se factoriser comme suit

$$H_j(\lambda) = (1 - e^{-i\lambda})^M \tilde{H}_j(\lambda), \quad (2.42)$$

où $\tilde{H}_j(\lambda)$ est également un polynôme trigonométrique. Les coefficients d'ondelettes (2.35) peuvent alors être calculer comme suit

$$W_{j,k} = (\downarrow^{2^j} [\tilde{h}_{j,\cdot} \star \Delta^M \mathbf{X}])_k \quad (2.43)$$

où $(\tilde{h}_{j,l})_{l \in \mathbb{Z}}$ sont les coefficients du polynôme trigonométrique \tilde{H}_j et $\Delta^M \mathbf{X}$ est le différencié d'ordre M de la séquence \mathbf{X} . En d'autres termes, l'utilisation d'une fonction d'ondelette satisfaisant les conditions (W-1)-(W-4) calcule implicitement le différencié d'ordre M de la série. Ainsi, on peut travailler avec un processus $M(d)$ au delà du régime stationnaire ($d > 1/2$) susceptible d'avoir été contaminé par une tendance polynomiale de degré K sans pré-traitement spécifique préalable du processus, à condition que $d - M < 1/2$ et $M \geq K + 1$. Ceci est en contraste avec les méthodes de Fourier, où la série doit être explicitement différenciée au moins K fois et une fenêtre doit être appliquée à la série différenciée pour éviter les fuites (*leakage*) dans le domaine fréquentiel (voir, par exemple Hurvich et al. [2002]).

Remarque 2.2.1. *Il est important de souligner ici que les coefficients d'échelle $A_{j,k}$ ne peuvent pas s'écrire sous la forme de (2.43) car la fonction d'échelle ϕ ne satisfait pas l'hypothèse (W-3). En général, on ne s'intéressera qu'aux coefficients d'ondelettes $W_{j,k}$, les coefficients d'échelles $A_{j,k}$ n'étant utiles que pour des calculs récursifs via l'algorithme pyramidal. Cet algorithme est obtenu pour des fonctions d'échelles ϕ et d'ondelettes ψ associées à une analyse multirésolution.*

2.2.3 Analyse multirésolution.

Dans cette sous section, on introduit l'analyse multirésolution (MRA). On considère une paire $\{\phi, \psi\}$ d'ondelettes *père* et *mère* définissant une analyse multirésolution, voir Cohen [2003]. C'est le cas pour les célèbres ondelettes de Daubechies. La MRA est très utilisée en pratique car elle permet de calculer de façon rapide la transformée en ondelette discrète en utilisant l'*algorithme pyramidal* voir Mallat [1998]. La complexité de cet algorithme est en $O(n)$ pour un échantillon de taille n . Nous passons maintenant à une présentation plus formelle de la MRA.

Analyse multirésolution

L'analyse multirésolution permet de décomposer une fonction $f \in L^2(\mathbb{R})$ à diverses échelles de résolutions par projection orthogonale sur différents espaces $\{V_j\}_{j \in \mathbb{Z}}$ emboîtés qui sont

des sous espaces fermés de $L^2(\mathbb{R})$.

Nous allons tout d'abord rappelé quelques définitions essentielles de l'analyse multirésolution.

Définition 2.2.1 (Multirésolution). *Une suite $\{V_j\}_{j \in \mathbb{Z}}$ de sous espaces fermés de $L^2(\mathbb{R})$ définit une MRA si les 6 propriétés suivantes sont satisfaites :*

$$\forall (j, k) \in \mathbb{Z}^2, f(t) \in V_j \Leftrightarrow f(t - 2^j k) \in V_j,$$

$$\forall j \in \mathbb{Z}, V_{j+1} \subset V_j,$$

$$\forall j \in \mathbb{Z}, f(t) \in V_j \Leftrightarrow f\left(\frac{t}{2}\right) \in V_{j+1},$$

$$\lim_{j \rightarrow \infty} V_j = \bigcap_{j=-\infty}^{\infty} V_j = \{0\},$$

$$\lim_{j \rightarrow -\infty} V_j = \overline{\left(\bigcup_{j=-\infty}^{\infty} V_j \right)} = L^2(\mathbb{R}).$$

Il existe θ tel que $\{\theta(t - n)\}_{n \in \mathbb{Z}}$ soit une base de Riesz de V_0 .

L'approximation de f à l'échelle j est définie comme la projection orthogonale $P_{V_j} f$ sur V_j . Pour calculer cette projection, on doit trouver une base orthonormée de V_j . Le Théorème 7.1 de Mallat [1998] orthogonalise la base de Riesz $\{\theta(\cdot - n)\}_{n \in \mathbb{Z}}$ et construit une base orthonormale de chaque sous espace V_j en translatant et dilatant la fonction d'échelle ϕ . Une approximation multirésolution est alors entièrement caractérisée par la fonction d'échelle qui génère une base orthonormale pour chaque sous espace V_j . Si $\{\phi(\cdot - k)\}$ est une base orthonormale de V_0 alors pour tout $j \in \mathbb{Z}$,

$$\phi_{j,k}(t) \stackrel{\text{def}}{=} 2^{-j/2} \phi(2^{-j}t - k)$$

est une base orthonormale de V_j . La projection orthogonale de f sur V_j est obtenue via une expansion dans la base orthonormale $\{\phi_{j,k}\}$

$$P_{V_j} f = \sum_{k=-\infty}^{\infty} \langle f, \phi_{j,k} \rangle \phi_{j,k}.$$

Les coefficients $a_j[k] \stackrel{\text{def}}{=} \langle f, \phi_{j,k} \rangle$ sont appelés *coefficients d'échelle* et la projection $P_{V_j} f$ fournit une approximation de f à l'échelle j (cette approximation est caractérisée par la donnée des coefficients d'échelles $(a_j[k])$, $k \in \mathbb{Z}$). Les propriétés de multirésolution imposent que les différents espaces d'approximation soient emboîtés: $V_{j+1} \subset V_j$. Cette condition implique, en particulier, que $2^{-1/2} \phi(t/2) \in V_1 \subset V_0$. Comme $\{\phi(\cdot - k)\}$ est une base orthonormale de V_0 , on peut donc décomposer

$$\frac{1}{\sqrt{2}} \phi\left(\frac{t}{2}\right) = \sum_{n=-\infty}^{\infty} h[n] \phi(t + n), \quad (2.44)$$

où $\{h[n]\}$ sont les coefficients du filtre définis par

$$h[n] \stackrel{\text{def}}{=} \langle 2^{-1/2}\phi(\cdot/2), \phi(\cdot + n) \rangle . \quad (2.45)$$

En prenant la transformée de Fourier de (2.44) on obtient

$$\widehat{\phi}(2\omega) = \frac{1}{\sqrt{2}}\widehat{h}(\omega)\widehat{\phi}(\omega) .$$

Les approximations de f aux échelles $j+1$ et j sont respectivement égales à leur projection orthogonale sur V_{j+1} et V_j . Par construction, $V_{j+1} \subset V_j$. Soit W_{j+1} le complément orthogonal de V_{j+1} dans V_j ,

$$V_j = V_{j+1} \oplus W_{j+1} .$$

La projection orthogonale de f sur V_j peut être décomposée comme une somme de projection orthogonale sur V_{j+1} et sur W_{j+1} ,

$$P_{V_j}f = P_{V_{j+1}}f + P_{W_{j+1}}f .$$

Le complément $P_{W_{j+1}}f$ fournit les détails de f qui apparaissent à l'échelle j .

Notons ψ la fonction dont la transformée de Fourier est donnée par

$$\widehat{\psi}(\omega) = \frac{1}{\sqrt{2}}\widehat{g}\left(\frac{\omega}{2}\right)\widehat{\phi}\left(\frac{\omega}{2}\right) ,$$

où ϕ est la fonction d'échelle et, h le filtre correspondant et

$$\widehat{g}(\omega) = e^{-i\omega}\widehat{h}^*(\omega + \pi) , \quad (2.46)$$

où $*$ désigne la transposée conjuguée. Notons

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k) .$$

Alors pour tout $j \in \mathbb{Z}$, $\{\psi_{j,k}\}_{k \in \mathbb{Z}}$ est une base orthonormée de W_j ; voir [Mallat, 1998, Theorem 7.3]. Comme $W_{j+1} \subset V_j$, en particulier $2^{-1/2}\psi(t/2) \in W_1 \subset V_0$. On peut ainsi décomposer $2^{-1/2}\psi(t/2)$ dans $\{\phi(\cdot - n)\}_{n \in \mathbb{Z}}$ qui est une base orthonormée de V_0 :

$$\frac{1}{\sqrt{2}}\psi\left(\frac{t}{2}\right) = \sum_{n=-\infty}^{\infty} g[n]\phi(t + n) , \quad (2.47)$$

avec

$$g[n] \stackrel{\text{def}}{=} \langle 2^{-1/2}\psi(\cdot/2), \phi(\cdot + n) \rangle . \quad (2.48)$$

Le Théorème 7.3 de Mallat [1998] assure que \widehat{g} est la série de Fourier de $g[n]$. Ainsi par transformée de Fourier inverse de (2.46) on obtient

$$g[n] = (-1)^{1-n}h[1 - n].$$

Ce filtre miroir joue un rôle essentiel dans l'algorithme de transformée en ondelette rapide. Une transformée en ondelettes rapide décompose successivement chaque approximation $P_{V_j} f$ en une approximation plus grossière $P_{V_{j+1}} f$ plus les coefficients d'ondelettes définissant la projection $P_{W_{j+1}} f$ sur le complémentaire orthogonal de l'espace V_{j+1} dans V_j .

Basés sur les filtres miroirs quadratures h et g définis respectivement en (2.45) et (2.48), on peut calculer récursivement les coefficients d'ondelettes et les coefficients d'approximation définis respectivement en (2.31) and (2.33) comme suit

$$A_{0,k} = X_k, \quad k \in \mathbb{Z}, \quad (2.49)$$

$$A_{j,k} = \sum_{n=-\infty}^{\infty} h[2k-n]A_{j-1,n} = [\downarrow^2 (h \star A_{j-1,\cdot})]_k, \quad k \in \mathbb{Z}, j \geq 1, \quad (2.50)$$

$$W_{j,k} = \sum_{n=-\infty}^{\infty} g[2k-n]A_{j-1,n} = [\downarrow^2 (g \star A_{j-1,\cdot})]_k, \quad k \in \mathbb{Z}, j \geq 1, \quad (2.51)$$

où $\{h[k]\}_{k \in \mathbb{Z}}$ et $\{g[k]\}_{k \in \mathbb{Z}}$ désignent les réponses impulsionnelles de h et g , \star désigne l'opérateur de convolution et \downarrow^2 l'opérateur de sous-échantillonnage par un facteur de 2. Pour comprendre la récursion ci-dessus, rappelons que $\phi_{j+1,p} \in V_{j+1} \subset V_j$, ainsi, peut être décomposée dans une base orthonormale $\{\phi_{j,n}\}_{n \in \mathbb{Z}}$ de V_j ,

$$\phi_{j+1,p} \stackrel{\text{def}}{=} \sum_{k=-\infty}^{\infty} \langle \phi_{j+1,p}, \phi_{j,k} \rangle \phi_{j,k}.$$

Avec le changement de variable $t' = 2^{-j}t - 2p$, on obtient

$$\begin{aligned} \langle \phi_{j+1,p}, \phi_{j,k} \rangle &= \int_{-\infty}^{\infty} 2^{-(j+1)/2} \phi(2^{-j+1}t - p) 2^{-j/2} \phi(2^{-j}t - k) dt \\ &= \int_{-\infty}^{\infty} 2^{-1/2} \phi(t/2) \phi(t - k + 2p) dt \\ &= h[2p - k]. \end{aligned}$$

Par conséquent, $\phi_{j+1,k} = \sum_{n=-\infty}^{\infty} h[2k-n] \phi_{j,n}$. Pour la fonction continue $\mathbf{X}_n(\cdot)$ définie en (2.30), $A_{j+1,k} = \langle \mathbf{X}_n, \phi_{j+1,k} \rangle$, et la relation précédente montre que

$$A_{j+1,k} = \sum_{n=-\infty}^{\infty} h[2k-n] \langle \mathbf{X}_n, \phi_{j,n} \rangle = \sum_{n=-\infty}^{\infty} h[2k-n] A_{j,n}.$$

Pareillement, $\psi_{j+1,k} = \sum_{n=-\infty}^{\infty} g[2k-n] \phi_{j,n}$ et $W_{j+1,k} = \langle \mathbf{X}_n, \psi_{j+1,k} \rangle$ peut s'écrire

$$W_{j+1,k} = \sum_{n=-\infty}^{\infty} g[2k-n] A_{j,n}.$$

Nous supposons que ψ est à support compact et notons M le nombre de moments nuls de l'ondelette ($M \geq 1$). Cela correspond à l'hypothèse suivante sur les filtres h et g .

Hypothèse 1. Les filtres h et g ont pour support $\{-T + 1, \dots, 0\}$ et g a M moments nuls i.e.,

$$\sum_{t=-T+1}^0 g[t] t^\ell = 0 \quad \text{pour tout } \ell = 0, 1, \dots, M-1. \quad (2.52)$$

L'hypothèse (1) sera considérée tout au long du chapitre 4. Il résulte de l'hypothèse de support compact des filtres qu'un nombre fini de coefficients d'ondelettes peut être calculé à partir de n observations X_0, \dots, X_{n-1} . Dans ce cas, l'algorithme pyramidal est appliqué avec $k = 0, \dots, n_0 - 1$ dans (2.49) et $k = 0, \dots, n_j - 1$ dans (2.50) et (2.51), où

$$n_0 = n \quad \text{et} \quad n_j = \{ \lfloor (n_{j-1} - T)/2 \rfloor + 1 \}_+, \quad j \geq 1, \quad (2.53)$$

où $\lfloor x \rfloor$ dénote la partie entière x et x_+ la partie positive de x , $\max(x, 0)$.

Algorithme pyramidal basé sur les incréments des coefficients d'approximation

La transformée en ondelettes discrète (2.49–2.51) peut être vue comme un algorithme admettant comme entrée les échantillons du processus $\{X_k, 0 \leq k < n\}$, fournissant en sortie les coefficients d'ondelettes aux différentes échelles $\{W_{j,k}, j \geq 1, 0 \leq k < n_j\}$. Dans la suite du travail, nous seront amené à travailler sur des processus non-stationnaires mais dont la différence K ième est stationnaire. Nous souhaiterions comprendre comment nous pourrions calculer les coefficients d'ondelettes mais en utilisant comme entrée $\{[\Delta^M X]\}_k$, $M \leq k < n$, le processus différencié K fois.

A cette fin, nous définissons les incréments des coefficients d'approximations comme suit: pour $\ell = 0, 1, \dots$ on note $A_{j,k}^{(\ell)} = [\Delta^\ell A_{j,\cdot}]_k$, i.e. $A_{j,k}^{(0)} = A_{j,k}$ et pour tout $\ell \geq 0$, $A_{j,k}^{(\ell+1)} = A_{j,k}^{(\ell)} - A_{j,k-1}^{(\ell)}$. Comme X peut ne pas être stationnaire mais devient stationnaire après un nombre suffisant de différenciation, il est très important d'obtenir un algorithme pyramidal dans lequel $A_{j,k}$ en (2.49) et (2.50) est remplacée par $A_{j,k}^{(\ell)}$ avec ℓ assez grand. Nous pourrons ainsi calculer les covariances de ces coefficients d'approximation différenciés, ainsi que les densités spectrales associées, puis utiliser ces différentes quantités pour calculer les coefficients d'autocovariance des coefficients d'ondelettes ainsi que leur densité spectrale.

L'hypothèse 1 stipule que on peut prendre $\ell = M$. Pour obtenir ce nouvel algorithme pyramidal, (le Lemme 2.2.1 ci-dessous), on introduit de nouveaux filtres \tilde{g} et \tilde{h} . Notons par

$$G(\lambda) = \sum_{k \in \mathbb{Z}} g[k] e^{-i\lambda k} \quad \text{et} \quad H(\lambda) = \sum_{k \in \mathbb{Z}} h[k] e^{-i\lambda k},$$

les séries de Fourier de g et h . D'après l'hypothèse 1, G est un polynôme trigonométrique de degré $T - 1$ qui peut être factorisé en $G(\lambda) = (1 - e^{i\lambda})^M \check{G}(\lambda)$, où \check{G} est un polynôme de degré $T - 1 - M$, i.e.,

$$\check{G}(\lambda) = \sum_{k=-T+M+1}^0 \check{g}[k] e^{-i\lambda k}. \quad (2.54)$$

D'où les suites g et \check{g} satisfont la relation suivante

$$g = (I - B^{-1})^M \check{g} = \mathbf{\Delta}^M (-B^{-1})^M \check{g} = \mathbf{\Delta}^M \tilde{g}, \quad (2.55)$$

où on pose \tilde{g} comme étant le filtre de longueur $(T - M)$ défini par

$$\tilde{g}[k] = (-1)^M \check{g}[k + M],$$

et de fonction de transfert

$$\tilde{G}(\lambda) = (-e^{i\lambda})^M \check{G}(\lambda) = (1 - e^{-i\lambda})^{-M} G(\lambda).$$

Définissons maintenant $\tilde{h} = (I + B)^M h$,

$$\tilde{h}[k] = \sum_{s=0}^M \binom{M}{s} h[k - s], \quad k \in \mathbb{Z}, \quad (2.56)$$

où $\binom{M}{s}$ désigne le coefficient binomial et considérons sa fonction de transfert $\tilde{H}(\lambda)$ définie par

$$\tilde{H}(\lambda) = \sum_{k \in \mathbb{Z}} \tilde{h}[k] e^{-i\lambda k} = \left(1 + e^{-i\lambda}\right)^M H(\lambda). \quad (2.57)$$

En utilisant les décompositions introduites ci-dessus, on peut obtenir un algorithme pyramidal pour calculer récursivement les coefficients d'approximations différenciés $A_{j,k}^{(M)}$ et les coefficients d'ondelettes $W_{j,k}$ en utilisant les filtres \tilde{h} et \tilde{g} définis en (2.56) et (2.55) respectivement.

Lemme 2.2.1. *Nous avons le schéma pyramidal suivant pour calculer $\{W_{j,k}, j \geq 1, k \in \mathbb{Z}\}$.*

$$A_{0,k}^{(M)} = [\mathbf{\Delta}^M X]_k, \quad k \in \mathbb{Z}, \quad (2.58)$$

$$A_{j,k}^{(M)} = \sum_{n=-\infty}^{\infty} \tilde{h}[2k - n] A_{j-1,n}^{(M)} = \left[\downarrow^2 \left(\tilde{h} \star A_{j-1,\cdot}^{(M)} \right) \right]_k, \quad k \in \mathbb{Z}, j \geq 1, \quad (2.59)$$

$$W_{j,k} = \sum_{n=-\infty}^{\infty} \tilde{g}[2k - n] A_{j-1,n}^{(M)} = \left[\downarrow^2 \left(\tilde{g} \star A_{j-1,\cdot}^{(M)} \right) \right]_k, \quad k \in \mathbb{Z}, j \geq 1, \quad (2.60)$$

La preuve est donnée en Annexe 4.5 du Chapitre 4.

2.2.4 Choix de l'ondelette

Les hypothèses (W-1)-(W-4) sont assez standard dans le cadre de l'analyse multirésolution (MRA). Dans ce cas, ϕ est la fonction d'échelle et ψ la fonction d'ondelette associée. (Voir par exemple, Cohen [2003]).

Dans une MRA, les coefficients (2.35) et (2.36) peuvent être calculer récursivement au fur et à mesure que j croît et il n'est pas nécessaire de calculer explicitement les filtres $h_{j,\cdot}$ et $b_{j,\cdot}$ défini en (2.37) et (2.38).

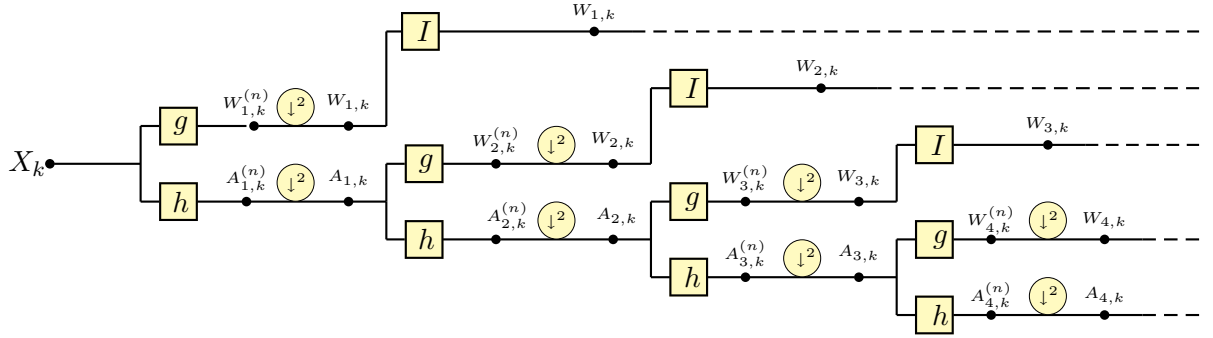


Figure 2.3: Calcul de la transformée en ondelette décimée en utilisant l’algorithme pyramidal où h et g sont les filtres miroirs quadratures et I désigne l’identité.

Les ondelettes de Daubechies et les ondelettes de Coiflets sont des exemples d’ondelettes orthogonales construites en utilisant la MRA.

Dans ce travail de thèse, sauf mention contraire, nous ne supposons pas dans nos analyses théoriques que les ondelettes sont orthogonales ni qu’elles proviennent d’une MRA. La propriété de MRA, très utile pour définir l’algorithme pyramidal de calcul des coefficients d’ondelettes, n’est en fait pas essentielle dans l’application considérée ici. Nous utiliseront pour nos analyses théoriques simplement des fonctions ϕ et ψ qui satisfont les conditions (W-1)-(W-4) mais qui ne définissent pas un MRA. Nous pouvons considérer par exemple, pour tout entier $N \geq 2$,

$$\phi(x) \stackrel{\text{def}}{=} \mathbb{1}_{[0,1]^{*N}}(x) \quad \text{et} \quad \psi(x) \stackrel{\text{def}}{=} c_M \frac{d^N}{dx^N} \mathbb{1}_{[0,1]^{*2N}}(2x), \quad (2.61)$$

où $\mathbb{1}_A$ est la fonction indicatrice sur l’ensemble A et pour des fonctions intégrables f , f^{*N} désigne la N -ième auto-convolution de f ,

$$f^{*N} = \underbrace{f \star \dots \star f}_{N \text{ fois}}.$$

et c_N est une constante de normalisation positive tel que $\int_{-\infty}^{\infty} \psi^2(t) dt = 1$. En utilisant (2.27) et (2.28), on vérifie aisément que les conditions (W-1)-(W-4) sont satisfaites avec M et α égal à N . Bien sur la famille de fonctions $\{\psi_{j,k}\}$ n’est pas orthonormale pour ce choix de la fonction d’ondelette ψ (et la fonction ϕ n’est pas associée à la MRA). Néanmoins avec un léger abus de terminologie, on appellera ϕ et ψ respectivement *fonction d’échelle* et *fonction d’ondelette*. Nous avons représenté graphiquement certains exemples d’ondelettes standard à la Figure 2.4.

2.2.5 Covariance et densité spectrale des coefficients d’ondelettes

Compte tenu du fait que les coefficients d’ondelettes à une échelle donnée sont obtenus par filtrage linéaire, calculer la covariance des coefficients d’ondelettes d’un processus dont le K -ième ordre de différentiation est stationnaire est un exercice relativement simple.

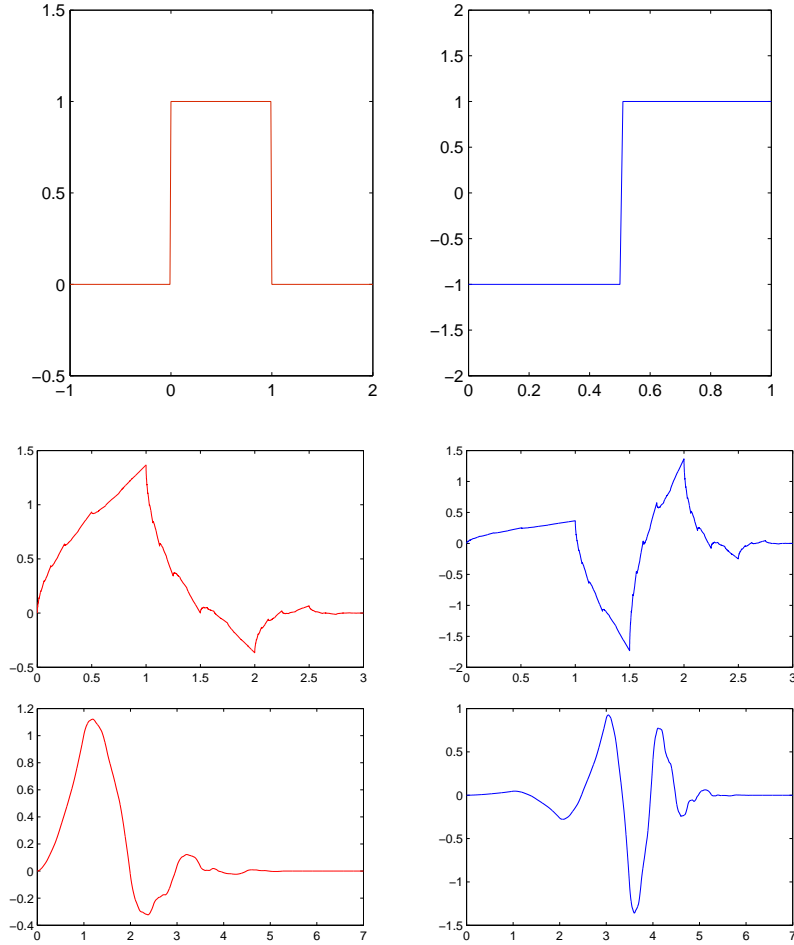


Figure 2.4: Fonction d'échelle ϕ (gauche/rouge) et fonction d'ondelette ψ (droite/bleu). La paire (ϕ, ψ) de haut en bas correspond à l'ondelette de Haar, ondelette de Daubechies avec $M = 2$ et $M = 4$ moments nuls..

[Moulines et al., 2007b, Proposition 1] donnent une expression intégrale de la covariance entre deux coefficients d'ondelettes à des échelles différentes exprimée en fonction de la fonction de transfert H_j définie en (2.42). On a alors

$$\text{Cov}(W_{j,k}; W_{j',k'}) = \int_{-\pi}^{\pi} e^{i\lambda(k2^j - k'2^{j'})} f(\lambda) H_j(\lambda) \bar{H}_{j'}(\lambda) d\lambda, \quad (2.62)$$

où les coefficients d'ondelettes $W_{j,k}$ sont définis en (2.31) et la densité spectrale généralisée f en (2.11).

Si le processus $\Delta^K \mathbf{X}$ est stationnaire, par (2.43), pour une échelle donnée j , le processus $\{W_{j,k}\}_{k \in \mathbb{Z}}$ est stationnaire. Mais pour deux échelles $j \neq j'$, le processus bidimensionnel $\left\{ (W_{j',k}; W_{j,k})^T \right\}_{k \in \mathbb{Z}}$, où T désigne la transposée n'est plus stationnaire à cause de la décimation. Ainsi, pour obtenir une séquence stationnaire, on doit considérer le processus

$$\left\{ [W_{j',2^{j-j'}k}; W_{j,k}]^T \right\}_{k \in \mathbb{Z}}, \quad j > j',$$

qui implique une séquence sous-échantillonné des coefficients à l'échelle fine j' . On peut aussi considérer le processus

$$\left\{ [W_{j', 2^{j-j'}k+v}; W_{j,k}]^T \right\}_{k \in \mathbb{Z}}, \quad j > j',$$

qui inclut une translation de l'indice de position de la première composante de v . Il s'avère que la méthode la plus commode consiste à fusionner les processus correspondant à $v = 0, \dots, 2^{j-j'} - 1$ et donc de considérer plutôt le *processus inter-échelle*

$$\left\{ [\mathbf{W}_{j,k}^T(j-j'), W_{j,k}]^T \right\}_{k \in \mathbb{Z}}, \quad (2.63)$$

où pour tout $u = 0, 1, \dots, j$,

$$\mathbf{W}_{j,k}(u) \stackrel{\text{def}}{=} [W_{j-u, 2^u k}, W_{j-u, 2^u k+1}, \dots, W_{j-u, 2^u k+2^u-1}]^T \quad (2.64)$$

est un vecteur de dimension 2^u composé des coefficients d'ondelettes à l'échelle $j' = j - u$ dans lequel toutes les translations possibles de l'indice de position $2^u k$ par $v = 0, 1, \dots, 2^u - 1$ sont prises en compte. L'indice u dans (2.64) désigne la différence $j - j' \geq 0$ entre l'échelle fine j' et l'échelle grossière j . Observons que $\mathbf{W}_{j,k}(0)$ ($u = 0$) est le scalaire $W_{j,k}$. On pourrait voir le processus inter-échelle (2.63) comme une paire composée du processus vectoriel $\{\mathbf{W}_{j,k}(j-j')\}_{k \in \mathbb{Z}}$ et du processus scalaire $\{W_{j,k}\}_{k \in \mathbb{Z}}$.

Comme le processus $\left\{ [\mathbf{W}_{j,k}^T(j-j'), W_{j,k}]^T \right\}_{k \in \mathbb{Z}}$ est stationnaire, il admet une densité spectrale [Moulines et al., 2007b, Corrolary 1] donnée par

$$\begin{aligned} \mathbf{D}_{j,u}(\lambda; f) \stackrel{\text{def}}{=} & \sum_{l=0}^{2^j-1} \mathbf{e}_u(\lambda + 2l\pi) f(2^{-j}(\lambda + 2l\pi)) 2^{-j/2} H_j(2^{-j}(\lambda + 2l\pi)) \\ & \times 2^{-(j-u)/2} \overline{H_{j-u}(2^{-j}(\lambda + 2l\pi))}, \end{aligned} \quad (2.65)$$

où pour tout $\lambda \in \mathbb{R}$,

$$\mathbf{e}_u(\lambda) \stackrel{\text{def}}{=} 2^{-u/2} [1, e^{-i2^{-u}\lambda}, \dots, e^{-i(2^u-1)2^{-u}\lambda}]^T$$

et $H_j(\lambda)$ est définie en (2.39). Le cas $u = 0$ correspond à la densité spectrale du processus *intra-échelle* $\{W_{j,k}\}_{k \in \mathbb{Z}}$ donnée par

$$\mathbf{D}_{j,0}(\lambda; f) \stackrel{\text{def}}{=} \sum_{l=0}^{2^j-1} f(2^{-j}(\lambda + 2l\pi)) 2^{-j} |H_j(2^{-j}(\lambda + 2l\pi))|^2. \quad (2.66)$$

Si on pose $j = j'$ et $k = k'$ dans (2.62) et si le processus $\mathbf{\Delta}^M X$ est stationnaire, on obtient

$$\sigma_j^2(d, f^*) = \text{Var}(W_{j,0}) = \int_{-\pi}^{\pi} |H_j(\lambda)|^2 |1 - e^{-i\lambda}|^{-2d} f^*(\lambda) d\lambda, \quad j \geq 0. \quad (2.67)$$

La suite $(\sigma_j^2)_{j \geq 0}$ est appelée *spectre d'ondelette* du processus X .

Si de plus $\Delta^M X$ est centrée, le spectre d'ondelette peut être estimé par les **scalogrammes** définis comme étant la moyenne empirique des carrés des coefficients d'ondelettes calculée à partir des observations X_1, \dots, X_n :

$$\hat{\sigma}_j^2 = \frac{1}{n_j} \sum_{k=1}^{n_j} W_{j,k}^2. \quad (2.68)$$

2.2.6 Calcul récursif de la matrice de covariance des coefficients d'ondelettes

Soit X un processus $M(d)$ de paramètre de mémoire d et de densité spectrale généralisée f . De la Définition 2.1.10, il existe K tel que $\Delta^K X$ est stationnaire. En augmentant K de façon à éliminer les tendances polynomiales, nous supposons en outre que $\Delta^K X$ est centré. En effet dans cette sous-section on n'utilise pas le fait que X est un processus $M(d)$ mais uniquement les hypothèses suivantes.

H-i Le processus différentié K fois $\Delta^K X$ est centré et stationnaire.

H-ii L'hypothèse (H-i) est obtenue avec $M \geq K$.

Soient i et j deux échelles dans la transformée en ondelettes. Définissons par $\mathbf{M}_{i,j} \stackrel{\text{def}}{=} (\mathbf{M}_{i,j}[k, k'], 0 \leq k \leq n_i - 1, 0 \leq k' \leq n_j - 1)$ la matrice de covariance des coefficients d'ondelettes aux échelles i et j , respectivement,

$$\mathbf{M}_{i,j}[k, k'] \stackrel{\text{def}}{=} \text{Cov}(W_{i,k}, W_{j,k'}). \quad (2.69)$$

Considérons les processus $\{A_{j,k}^{(M)}\}_{k \in \mathbb{Z}}$ et $\{W_{j,k}\}_{k \in \mathbb{Z}}$ définis récursivement en (2.58)-(2.60) et notons $\delta_{i,j}$, $\gamma_{i,j}$ et $\nu_{i,j}$ les fonctions de covariances définies par

$$\begin{aligned} \gamma_{i,j}[k] &\stackrel{\text{def}}{=} \text{Cov}(W_{i,k}, A_{j,0}^{(M)}), & \delta_{i,j}[k] &\stackrel{\text{def}}{=} \text{Cov}(W_{i,k}, W_{j,0}), & 1 \leq i \leq j, \\ \nu_i[k] &\stackrel{\text{def}}{=} \text{Cov}(A_{i,k}^{(M)}, A_{i,0}^{(M)}), & 0 \leq i. \end{aligned}$$

Quand $i = j$, on note $\gamma_{i,j} = \gamma_i$ et $\delta_{i,j} = \delta_i$. En utilisant la définition et la structure de stationnarité des coefficients d'ondelettes, on obtient

$$\mathbf{M}_{i,j}[k, k'] = \delta_{i,j}[k - 2^{j-i}k'], \quad 1 \leq i \leq j.$$

De même désignons par $\mathbf{D}_{j,u}$ la cross-densité spectrale du processus $\{[\mathbf{W}_{j,k}^T(u), W_{j,k}]^T\}_{k \in \mathbb{Z}}$ et par $\mathbf{B}_{j,u}$ la densité inter-spectrale du processus $\{[\mathbf{W}_{j,k}^T(u), A_{j,k}^{(M)}]^T\}_{k \in \mathbb{Z}}$.

Ainsi on a

$$[\delta_{i,j}[2^u k], \dots, \delta_{i,j}[2^u(k+1) - 1]]^T = \text{Cov}(\mathbf{W}_{j,k}(u), W_{j,0}) = \int_{-\pi}^{\pi} \mathbf{D}_{j,u}(\lambda) e^{i\lambda k} d\lambda \quad (2.70)$$

$$[\gamma_{i,j}[2^u k], \dots, \gamma_{i,j}[2^u(k+1) - 1]]^T = \text{Cov}(\mathbf{W}_{j,k}(u), A_{j,0}^{(M)}) = \int_{-\pi}^{\pi} \mathbf{B}_{j,u}(\lambda) e^{i\lambda k} d\lambda. \quad (2.71)$$

Par convention, $\mathbf{D}_{j,0}$ ($u = 0$) désigne la densité spectrale de $\{W_{j,k}, k \in \mathbb{Z}\}$. Finalement, soit \mathbf{A}_j la densité spectrale du processus $\{A_{j,k}^{(M)}\}_{k \in \mathbb{Z}}$. alors pour tout $j \geq 0$,

$$\nu_j[k] = \text{Cov} \left(A_{j,k}^{(M)}, A_{j,0}^{(M)} \right) = \int_{-\pi}^{\pi} \mathbf{A}_j(\lambda) e^{i\lambda k} d\lambda. \quad (2.72)$$

Observons que \mathbf{A}_0 et ν_0 sont respectivement la densité spectrale et la fonction d'autocovariance de $\mathbf{\Delta}^M X$. En particulier, si f est la densité spectrale généralisée de X définie en (2.11), alors

$$\mathbf{A}_0(\lambda) = |1 - e^{-i\lambda}|^{2M} f(\lambda) \text{ et } \nu_0(\tau) = \int_{-\pi}^{\pi} \mathbf{A}_0(\lambda) e^{i\lambda\tau} d\lambda$$

sont obtenues aisément à partir des propriétés du second ordre de X . On aimerait maintenant obtenir les propriétés de second ordre des coefficients d'ondelettes. Cela peut être obtenu soit en utilisant le domaine spectral en calculant $\mathbf{D}_{j,u}$ pour tout $j \geq 1$ et $0 \leq u \leq j$, soit en utilisant le domaine temporelle en calculant $\delta_{i,j}$ pour tout $1 \leq i \leq j$. Pour obtenir ces calculs récursifs, nous aurons besoin de la densité spectrale \mathbf{A}_j , de la cross-spectrale densité $\mathbf{B}_{j,u}$, ou des covariances $\gamma_{i,j}$ et $\nu_{i,j}$. On obtient alors les formules récursives suivantes

Domaine spectral

Proposition 2.2.1. *En utilisant les définitions précédentes, on a pour tout $j \geq 1$, $0 \leq u \leq j - 1$ et $0 \leq v < 2^{u+1}$,*

$$\begin{aligned} \mathbf{A}_j(\lambda) &= \frac{1}{2} \left\{ \mathbf{A}_{j-1} \left(\frac{\lambda}{2} \right) \left| \tilde{H} \left(\frac{\lambda}{2} \right) \right|^2 + \mathbf{A}_{j-1} \left(\frac{\lambda}{2} + \pi \right) \left| \tilde{H} \left(\frac{\lambda}{2} + \pi \right) \right|^2 \right\}, \\ \mathbf{D}_{j,0}(\lambda) &= \frac{1}{2} \left\{ \mathbf{A}_{j-1} \left(\frac{\lambda}{2} \right) \left| \tilde{G} \left(\frac{\lambda}{2} \right) \right|^2 + \mathbf{A}_{j-1} \left(\frac{\lambda}{2} + \pi \right) \left| \tilde{G} \left(\frac{\lambda}{2} + \pi \right) \right|^2 \right\}, \\ \mathbf{B}_{j,0}(\lambda) &= \frac{1}{2} \left\{ \mathbf{A}_{j-1} \left(\frac{\lambda}{2} \right) \tilde{G} \left(\frac{\lambda}{2} \right) \tilde{H}^* \left(\frac{\lambda}{2} \right) \right. \\ &\quad \left. + \mathbf{A}_{j-1} \left(\frac{\lambda}{2} + \pi \right) \tilde{G} \left(\frac{\lambda}{2} + \pi \right) \tilde{H}^* \left(\frac{\lambda}{2} + \pi \right) \right\}, \\ \mathbf{B}_{j+u+1,u+1}^{(v)}(\lambda) &= \frac{1}{2} \left[e^{i\lambda \langle \frac{v}{2} \rangle} \tilde{H}^* \left(\frac{\lambda}{2} \right) \mathbf{B}_{j+u,u}^{(\lfloor \frac{v}{2} \rfloor)} \left(\frac{\lambda}{2} \right) + \right. \\ &\quad \left. e^{i(\lambda+2\pi) \langle \frac{v}{2} \rangle} \tilde{H}^* \left(\frac{\lambda}{2} + \pi \right) \mathbf{B}_{j+u,u}^{(\lfloor \frac{v}{2} \rfloor)} \left(\frac{\lambda}{2} + \pi \right) \right], \\ \mathbf{D}_{j+u+1,u+1}^{(v)}(\lambda) &= \frac{1}{2} \left[e^{i\lambda \langle \frac{v}{2} \rangle} \tilde{G}^* \left(e^{i\frac{\lambda}{2}} \right) \mathbf{B}_{j+u,u}^{(\lfloor \frac{v}{2} \rfloor)} \left(\frac{\lambda}{2} \right) \right. \\ &\quad \left. + e^{i(\lambda+2\pi) \langle \frac{v}{2} \rangle} \tilde{G}^* \left(\frac{\lambda}{2} + \pi \right) \mathbf{B}_{j+u,u}^{(\lfloor \frac{v}{2} \rfloor)} \left(\frac{\lambda}{2} + \pi \right) \right], \end{aligned}$$

où z^* désigne le conjugué de z , $[x]$ la partie entière de x et $\langle x \rangle = x - [x]$ la partie fractionnaire de x . $\mathbf{B}_{j,u}^{(v)}$ et $\mathbf{D}_{j,u}^{(v)}$ désignent les $(v+1)$ -ième composantes de $\mathbf{B}_{j,u}$ et $\mathbf{D}_{j,u}$,

$$\mathbf{B}_{j,u}(\lambda) = \left[\mathbf{B}_{j,u}^{(v)}(\lambda) \right]_{v=0,\dots,2^u-1}^T \quad \text{et} \quad \mathbf{D}_{j,u}(\lambda) = \left[\mathbf{D}_{j,u}^{(v)}(\lambda) \right]_{v=0,\dots,2^u-1}^T .$$

La preuve de la Proposition 2.2.1 est donnée en Section 4.5.

Domaine temporel

On peut également en utilisant les covariances définies en (2.70)-(2.72) obtenir les formules récursives suivantes pour le calcul des covariances

Proposition 2.2.2. *Pour tout $k \in \mathbb{Z}$ et $1 \leq i \leq j$.*

$$\begin{aligned} \nu_j[k] &= \sum_{l,l' \in \mathbb{Z}} \tilde{h}[l] \tilde{h}[l'] \nu_{j-1} [2k - l' + l] , \\ \delta_j[k] &= \sum_{l,l' \in \mathbb{Z}} \tilde{g}[l] \tilde{g}[l'] \nu_{j-1} [2k - l' + l] , \\ \gamma_j[k] &= \sum_{l,l' \in \mathbb{Z}} \tilde{g}[l] \tilde{h}[l'] \nu_{j-1} [2k - l + l'] , \\ \gamma_{i,j+1}[k] &= \sum_{l \in \mathbb{Z}} \tilde{h}[l] \gamma_{i,j}[k + 2^{j-i}l] , \\ \delta_{i,j+1}[k] &= \sum_{l \in \mathbb{Z}} \tilde{g}[l] \gamma_{i,j}[k + 2^{j-i}l] . \end{aligned}$$

La preuve de la Proposition 2.2.2 est donnée en Section 4.5.

Les algorithmes donnés par les Propositions 2.2.1 et 2.2.2 sont très intéressants en pratique car ils permettent de calculer récursivement donc rapidement la matrice de covariance des coefficients d'ondelettes d'un processus X pas nécessairement stationnaire mais pouvant être rendu stationnaire après un nombre fini de différentiation.

2.3 Estimation semi-paramétrique du paramètre de mémoire

d

Dans cette section, nous décrivons la cadre semi-paramétrique dans lequel nous nous plaçons pour estimer le paramètre de mémoire d et présentons des estimateurs semi-paramétriques du paramètre de mémoire d en ondelettes.

Nous supposons que le processus X est un processus $M(d)$ de densité spectrale généralisée $f(\lambda) = |\lambda|^{-2d} f^*(\lambda)$ où f^* est une fonction régulière en 0. On souhaite estimer le paramètre de mémoire d inconnu en présence de f^* vue comme paramètre de nuisance. Ce cadre d'estimation s'appelle *estimation semi-paramétrique*.

Nous utilisons une approche locale qui présente l'intérêt de n'avoir à spécifier le comportement de la fonction f^* qu'aux fréquences proches de l'origine, là où l'influence de f^* peut être négligée. Dans la suite, nous présentons deux estimateurs semi-paramétriques du paramètre de mémoire dans le cadre des ondelettes. Il s'agit de l'estimateur local de régression (LRW) et de l'estimateur local de Whittle (LWW).

L'estimateur LRW a été proposée par Abry and Veitch [1998] sous l'hypothèse simple que les coefficients d'ondelettes étaient décorrelés voir aussi Veitch and Abry [1999]. Bardet et al. [2000], Bardet [2002] et Coeurjolly and Istas [2001] fournissent des propriétés asymptotiques pour l'estimateur LRW dans un contexte paramétrique. L'article de Bardet et al. [2000] est le premier à analyser le comportement de l'estimateur LRW dans un contexte semi-paramétrique (ce contexte est toutefois différent de celui considéré ici, les coefficients d'ondelettes étant supposés avoir été obtenus en utilisant la transformée en ondelettes continues et le processus à analyser est supposé être stationnaire gaussien).

L'estimateur LWW a été introduit par Wornell and Oppenheim [1992], et a ensuite été étudié dans Kaplan and Kuo [1993] et McCoy and Walden [1996] (dans un cadre paramétrique, les résultats obtenus étant essentiellement heuristiques). L'analyse dans le cadre semi-paramétrique considéré dans ce travail a été réalisée par Moulines et al. [2008].

2.3.1 L'estimateur local de régression en ondelettes

Soient n , j_0 et j_1 , $j_0 \leq j_1$ des entiers, l'ensemble des coefficients d'ondelettes disponibles à partir de n observations X_1, \dots, X_n ayant des indices entre j_0 et j_1 est

$$\mathcal{I}_n(j_0, j_1) \stackrel{\text{def}}{=} \{(j, k) : j_0 \leq j \leq j_1, 0 \leq k < n_j\}, \quad (2.73)$$

où n_j est donné en (2.32). Considérons deux entiers $J_0 = J_0(n)$ et $J_1 = J_1(n)$ satisfaisant

$$0 \leq J_0 < J_1 \leq J_n \stackrel{\text{def}}{=} \max\{j : n_j \geq 1\}. \quad (2.74)$$

L'indice J_n est l'indice de l'échelle maximale disponible pour n observations; J_0 et J_1 désignent respectivement l'échelle la plus fine et l'échelle la plus grossière utilisées dans l'estimation. Pour un processus $M(d)$, sous certaines conditions de régularité sur la densité spectrale de courte mémoire f^* et pour des fonction d'échelle ϕ et d'ondelette ψ bien choisies, on peut montrer que quand $j \rightarrow \infty$, $\sigma_j^2(d, f^*) = \text{Var}[W_{j,0}] \asymp \sigma^2 2^{2dj}$ et la variance empirique $\hat{\sigma}_j^2$ définie en (2.68) est un estimateur consistant de $\sigma_j^2(d, f^*)$ (voir [Faÿ et al., 2009, Proposition 2]). Un estimateur semi-paramétrique couramment utilisé en pratique du paramètre de mémoire d est l'estimateur local de régression en ondelettes proposé par Abry and Veitch [1998] défini comme l'estimateur des moindres carrés dans le modèle de régression linéaire

$$\log [\hat{\sigma}_j^2] = \log \sigma^2 + dj\{2 \log(2)\} + u_j,$$

où $u_j = \log[\widehat{\sigma}_j^2/\sigma^2 2^{2dj}]$. La solution de ce problème de régression est donnée par :

$$\widehat{d}_n^{\text{LRW}}(J_0, J_1, \mathbf{w}) \stackrel{\text{def}}{=} \sum_{j=J_0}^{J_1} w_{j-J_0} \log(\widehat{\sigma}_j^2), \quad (2.75)$$

où le vecteur des poids $\mathbf{w} \stackrel{\text{def}}{=} [w_0, \dots, w_{J_1-J_0}]^T$ satisfait

$$\sum_{j=0}^{J_1-J_0} w_j = 0 \quad \text{et} \quad 2 \log(2) \sum_{j=0}^{J_1-J_0} j w_j = 1. \quad (2.76)$$

Pour $J_1 - J_0 = \ell \geq 1$, on peut choisir par exemple, \mathbf{w} correspondant au vecteur de régression des moindres carrés pondérés, défini par

$$\mathbf{w} \stackrel{\text{def}}{=} DB(B^T DB)^{-1} \mathbf{b} \quad (2.77)$$

où

$$\mathbf{b} \stackrel{\text{def}}{=} \begin{bmatrix} 0 \\ (2 \log(2))^{-1} \end{bmatrix}, \quad B \stackrel{\text{def}}{=} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 0 & 1 & \dots & \ell \end{bmatrix}^T \quad (2.78)$$

et D une matrice définie positive arbitraire.

2.3.2 Estimateur local de Whittle en ondelettes

La propriété d'échelle $\sigma_j^2(d, f^*) \stackrel{\text{def}}{=} \text{Var}[W_{j,0}] \asymp \sigma^2 2^{2dj}$ et les conditions de faibles dépendances des coefficients d'ondelettes suggèrent de considérer la pseudo log-vraisemblance suivante

$$\widehat{L}_{\mathcal{I}}(\sigma^2, d) = \frac{1}{2\sigma^2} \sum_{(j,k) \in \mathcal{I}} 2^{-2dj} (W_{j,k})^2 + \frac{|\mathcal{I}|}{2} \log(\sigma^2 2^{2\langle \mathcal{I} \rangle} d),$$

où $|\mathcal{I}|$ désigne le cardinal de \mathcal{I} et $\langle \mathcal{I} \rangle$ est l'échelle moyenne, $\langle \mathcal{I} \rangle \stackrel{\text{def}}{=} |\mathcal{I}|^{-1} \sum_{(j,k) \in \mathcal{I}} j$. Définissons $\widehat{\sigma}_{\mathcal{I}}^2(d) \stackrel{\text{def}}{=} \text{Argmin}_{\sigma^2 > 0} \widehat{L}_{\mathcal{I}}(\sigma^2, d) = |\mathcal{I}|^{-1} \sum_{(j,k) \in \mathcal{I}} 2^{-2dj} (W_{j,k})^2$. Le pseudo estimateur du maximum de vraisemblance du paramètre de mémoire est alors égal au maximum de la log-vraisemblance profilée

$$\widehat{d}_n^{\text{LWW}}(J_0, J_1, \sigma^2) \stackrel{\text{def}}{=} \underset{d \in [\Delta_1, \Delta_2]}{\text{Argmin}} \widehat{L}_{\mathcal{I}_n(J_0, J_1)}(\widehat{\sigma}_{\mathcal{I}}^2(d), d) = \underset{d \in [\Delta_1, \Delta_2]}{\text{Argmin}} \widetilde{L}_{\mathcal{I}_n(J_0, J_1)}(d), \quad (2.79)$$

où $[\Delta_1, \Delta_2]$ est un intervalle de valeurs possibles de d et

$$\widetilde{L}_{\mathcal{I}}(d) \stackrel{\text{def}}{=} \log \left(\sum_{(j,k) \in \mathcal{I}} 2^{2d(\langle \mathcal{I} \rangle - j)} (W_{j,k})^2 \right). \quad (2.80)$$

Cet estimateur a été proposé pour analyser les données bruitées dans Wornell and Oppenheim [1992], et a ensuite été examiné par plusieurs auteurs, principalement dans un contexte paramétrique, voir par exemple Kaplan and Kuo [1993] et McCoy and Walden

[1996]. Si \mathcal{I} contient au moins deux échelles différentes, alors $\tilde{L}_{\mathcal{I}}(d) \rightarrow \infty$ quand $d \rightarrow \pm\infty$, et ainsi \hat{d} est fini.

Dans la suite, nous présentons les propriétés asymptotiques développées dans Moulines et al. [2007b] et Moulines et al. [2008] pour ces deux estimateurs. Avant cela, nous définissons le mouvement Brownien fractionnaire généralisé qui joue un rôle essentiel pour établir ces résultats.

2.3.3 Le mouvement Brownien fractionnaire généralisé

Considérons le processus Gaussien $\{B_{(d)}(\theta)\}$ indexé par la famille de fonctions $\theta \in \Theta_{(d)} = \left\{ \theta : \int_{\mathbb{R}} |\xi|^{-2d} |\hat{\theta}(\xi)|^2 d\xi < \infty \right\}$ de moyenne zero et de covariance

$$\text{Cov}(B_{(d)}(\theta_1), B_{(d)}(\theta_2)) = \int_{\mathbb{R}} |\xi|^{-2d} \hat{\theta}_1(\xi) \overline{\hat{\theta}_2(\xi)} d\xi. \quad (2.81)$$

Le processus $\{B^{(d)}(\theta)\}$ est appelé *mouvement Brownien fractionnaire généralisé* (GFBM).

Lorsque $d > 1/2$, la condition $\int |\xi|^{-2d} |\hat{\theta}(\xi)|^2 d\xi < \infty$ requiert que $\hat{\theta}(\xi)$ décroisse assez vite à l'origine et, quand $d < 0$, elle requiert que $\hat{\theta}(\xi)$ décroisse suffisamment vite à l'infini. Tant que $1/2 - \alpha < d < M + 1/2$ où α est défini en (W-2) et M le nombre de moments nuls. La fonction d'ondelette ψ et ses translats dilatés $\psi_{j,k}$ appartiennent tous à $\Theta_{(d)}$. Le coefficient d'ondelette du GFBM à l'échelle j et à la position k est défini comme étant la transformée en ondelette discrète de $B^{(d)}$

$$W_{j,k}^{(d)} \stackrel{\text{def}}{=} B^{(d)}(\psi_{j,k}), \quad j \in \mathbb{Z}, k \in \mathbb{Z} \quad (2.82)$$

et $\mathbf{W}_{j,k}^{(d)}(u) \stackrel{\text{def}}{=} [W_{j-u,2^u k}^{(d)}, \dots, W_{j-u,2^u k+2^u-1}^{(d)}]$. Comme prouvé dans [Moulines et al., 2007b, Relation (35)], pour tout j, k et $k' \in \mathbb{Z}$, et $u \geq 0$, on a

$$\text{Cov}\left(\mathbf{W}_{j,k'}^{(d)}(u), W_{j,k}^{(d)}\right) = 2^{2dj} \int_{-\pi}^{\pi} \mathbf{D}_{\infty,u}(\lambda; d) e^{i\lambda(k-k')} d\lambda, \quad (2.83)$$

où, pour tout $u \geq 0$, $1/2 - \alpha < d < M + 1/2$ et $\lambda \in (-\pi, \pi)$,

$$\begin{aligned} \mathbf{D}_{\infty,u}(\lambda; d) &= \left[\mathbf{D}_{\infty,u}^{(0)}(\lambda; d), \dots, \mathbf{D}_{\infty,u}^{(2^u-1)}(\lambda; d) \right] \\ &= \sum_{l \in \mathbb{Z}} |\lambda + 2l\pi|^{-2d} \mathbf{e}_u(\lambda + 2l\pi) \overline{\hat{\psi}(\lambda + 2l\pi)} \hat{\psi}(2^{-u}(\lambda + 2l\pi)), \end{aligned} \quad (2.84)$$

avec $\mathbf{e}_u(\xi) \stackrel{\text{def}}{=} 2^{-u/2} [1, e^{-i2^{-u}\xi}, \dots, e^{-i(2^u-1)2^{-u}\xi}]^T$. En fait pour tout $d \in \mathbb{R}$, $2^{2jd} \mathbf{D}_{\infty,u}(\cdot; d)$, est la densité inter-spectrale des coefficients d'ondelettes du GFBM. Comme le GFBM est juste une extension du FBM, il reste auto-similaire et la propriété d'auto-similarité entraîne la proposition suivante.

Proposition 2.3.1. Soit $d \in \mathbb{R}$, $a \in \mathbb{R}^*$ fixé et Δ_a un opérateur défini par

$$\begin{aligned}\Delta_a : \Theta_{(d)} &\rightarrow \Theta_{(d)} \\ \theta &\mapsto \Delta_a(\theta)(\cdot),\end{aligned}$$

tel que $\Delta_a(\theta)(\xi) = a^{-1}\theta(a^{-1}\xi)$. Alors on a l'égalité des distributions finies dimensionnelles suivante

$$B^{(d)}(\Delta_a(\cdot)) \stackrel{\mathcal{L}}{=} a^{d-1/2}B^{(d)}(\cdot), \quad (2.85)$$

et pour $j > 0$ fixé et $i \geq 0$, on a la relation d'auto-similarité

$$W_{j+i,k}^{(d)} \stackrel{\mathcal{L}}{=} 2^{dj}W_{i,k}^{(d)}. \quad (2.86)$$

Preuve de la Proposition 2.3.1.

$$\begin{aligned}B^{(d)}(\theta) &= \int B^{(d)}(t)\theta(t)dt = \int B^{(d)}(at)a^{-(d-\frac{1}{2})}\theta(t)dt \\ &= a^{-(d-\frac{1}{2})} \int B^{(d)}(u)a^{-1}\theta(a^{-1}u)du,\end{aligned}$$

où l'égalité doit être comprise au sens d'égalité des distributions finies dimensionnelles. La seconde égalité résulte de la propriété d'auto-similarité du processus $\{B_{(d)}(t)\}_{t \in \mathbb{R}}$ et la dernière du changement de variable $u = at$. Ainsi, on obtient (2.85). Prouvons maintenant (2.86)

$$\begin{aligned}W_{j+i,k}^{(d)} &\stackrel{\text{def}}{=} B^{(d)}(\psi_{j+i,k}) = 2^{j/2}B^{(d)}(\Delta_{2^j}(\psi_{i,k})) = 2^{j/2}2^{j(d-\frac{1}{2})}B^{(d)}(\psi_{i,k}) \\ &\stackrel{\mathcal{L}}{=} 2^{jd}B^{(d)}(\psi_{i,k})\end{aligned}$$

d'où l'égalité (2.86). ■

2.3.4 Approximations asymptotiques.

Comme $d \in \mathbb{R}$, pour obtenir des bornes d'erreurs sur la densité spectrale et la variance des coefficients d'ondelettes du processus $M(d) X$ définis respectivement en (2.65) et (2.67), quelques hypothèses supplémentaires sur la régularité de f^* à la fréquence nulle sont requises. L'hypothèse la plus commune, introduite par Robinson [1995a], est une condition de Hölder sur la densité spectrale de courte mémoire f^* en (2.9).

Définition 2.3.1. Pour tout $0 < \beta \leq 2$, $L > 0$ et $\varepsilon \in (0, \pi]$, $\mathcal{H}(\beta, L, \varepsilon)$ est l'ensemble de toute les fonctions positives et paires g satisfaisant $g(0) > 0$ et pour tout $\lambda \in (-\varepsilon, \varepsilon)$

$$|g(\lambda) - g(0)| \leq Lg(0)|\lambda|^\beta. \quad (2.87)$$

Plus β est grand, plus régulière est la fonction à l'origine. Observons que si f^* est paire comme supposée et infiniment différentiable, alors $f^{*\prime}(0) = 0$ d'où par un

développement de Taylor, la condition (2.87) est obtenue avec $\beta = 2$, i.e $f^* \in \mathcal{H}(2, \gamma, \varepsilon)$. Utilisant [Moulines et al., 2007b, Theorem 1] et [Moulines et al., 2008, Theorem 1], on a le résultat suivant.

Proposition 2.3.2. *Soit X un processus $M(d)$ avec $d \in \mathbb{R}$ et $f^* \in \mathcal{H}(\beta, L, \varepsilon)$ pour tout $0 < \beta \leq 2$, $L > 0$ et $\varepsilon \in (0, \pi]$. Sous les hypothèses (W-1)-(W-4) avec $(1 + \beta)/2 - \alpha < d < M + 1/2$, on a*

$$\left| \sigma_j^2(d, f^*) - f^*(0) K(d) 2^{2jd} \right| \leq C f^*(0) L 2^{(2d-\beta)j}. \quad (2.88)$$

Si de plus $\varepsilon = \pi$ et $d \leq M$, alors, pour tout $\lambda \in (-\pi, \pi)$, $j \geq u \geq 0$,

$$\left| \mathbf{D}_{j,u}(\lambda; d, f^*) - f^*(0) \mathbf{D}_{\infty,u}(\lambda; d) 2^{2jd} \right| \leq C f^*(0) L 2^{(2d-\beta)j} \quad (2.89)$$

où

$$K(d) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} |\xi|^{-2d} |\widehat{\psi}(\xi)|^2 d\xi, \quad (2.90)$$

$\mathbf{D}_{\infty,u}(\lambda; d)$ est défini en (2.84) et $|y|$ désigne la norme Euclidienne du vecteur y .

Ces deux inégalités montrent qu'on peut approcher les quantités $\mathbf{D}_{j,u}(\cdot, d, f^*)$ et $\sigma_j^2(d, f^*)$ qui dépendent toutes les deux de d et de la fonction f^* par des quantités qui ne dépendent que de d et $f^*(0)$. Dans les équations (2.88) et (2.89), la constante C dépend uniquement de d , β et des fonctions ψ et ϕ . Elle peut être rendue indépendante de d sur n'importe quel ensemble inclu dans $((1 + \beta)/2 - \alpha, M + 1/2)$ pour (2.88) et dans $((1 + \beta)/2 - \alpha, M]$ pour (2.89). Cette proposition montre que la covariance des coefficients d'ondelettes d'un processus $M(d)$ ressemble à celle du GFBM B_d aux grandes échelles. Les coefficients d'ondelettes de ce processus ne sont pas décorrelés, comme cela a parfois été un peu hâtivement supposé dans des analyses heuristiques (voir Abry and Veitch [1998] et Veitch and Abry [1999] par exemple). Des décorrelations exactes peuvent se produire mais dans des cas très spécifiques: si $\{\psi_{j,k}\}$ est une base orthonormée de $L^2(\mathbb{R})$ et $d = 0$, voir [Moulines et al., 2007b, Remarque 8]. Comme

$$\text{Cov}(\mathbf{W}_{j,k'}(u), W_{j,k}) = \int_{-\pi}^{\pi} \mathbf{D}_{j,u}(\lambda; f) e^{i\lambda(k-k')} d\lambda,$$

et

$$\text{Cov}(\mathbf{W}_{j,k'}^{(d)}(u), W_{j,k}^{(d)}) = 2^{2dj} \int_{-\pi}^{\pi} \mathbf{D}_{\infty,u}(\lambda; d) e^{i\lambda(k-k')} d\lambda,$$

nous avons les approximations suivantes.

$$\left| \text{Cov}(\mathbf{W}_{j,k'}(u), W_{j,k}) - f^*(0) \text{Cov}(\mathbf{W}_{j,k'}^{(d)}(u), W_{j,k}^{(d)}) \right| \leq C f^*(0) L 2^{(2d-\beta)j}. \quad (2.91)$$

Proposition 2.3.3. *Soient i et j deux indices d'échelle tel que $J_1 \leq i \leq j \leq J_2$, la propriété d'auto-similarité du GFBM implique les relations suivantes :*

$$\text{Cov}(\mathbf{W}_{J_1+j,k'}^{(d)}(u), W_{J_1+j,k}^{(d)}) = 2^{2d(J_1+j-u)} \text{Cov}(\mathbf{W}_{u,k'}^{(d)}(u), W_{u,k}^{(d)}) \quad (2.92)$$

pour k et k' positifs ou nuls. De plus pour J_1 fixé, on a

$$\text{Cov} \left(\mathbf{W}_{J_1+j,k'}^{(d)}(u), W_{J_1+j,k}^{(d)} \right) = 2^{2d(J_1+j-u)} f^*(0) \text{Cov} \left(\mathbf{W}_{u,k'}^{(d)}(u), W_{u,k}^{(d)} \right) + \text{O} \left(2^{(2d-\beta)j} \right). \quad (2.93)$$

Preuve de la Proposition 2.3.3. En utilisant (2.81), (2.82) et $\widehat{\psi}_{j,k}(\lambda) = 2^{j/2} e^{-i2^j \lambda k} \widehat{\psi}(2^j \lambda)$, on a pour tout $v = 0, \dots, 2^u - 1$

$$\begin{aligned} \text{Cov} \left(W_{i,2^{j-i}k'+v}^{(d)}, W_{j,k}^{(d)} \right) &= 2^{i/2} 2^{j/2} \int_{\mathbb{R}} |\lambda|^{-2d} e^{-i\lambda 2^i (2^{j-i}(k'-k)+v)} \widehat{\psi}(2^i \lambda) \overline{\widehat{\psi}}(2^j \lambda) d\lambda \\ &= 2^{2di} 2^{(j-i)/2} \int_{\mathbb{R}} |\lambda|^{-2d} e^{-i\lambda (2^{j-i}(k'-k)+v)} \widehat{\psi}(\lambda) \overline{\widehat{\psi}}(2^{j-i} \lambda) d\lambda \\ &= 2^{2di} \text{Cov} \left(W_{0,2^{j-i}k'+v}^{(d)}, W_{j-i,k}^{(d)} \right). \end{aligned}$$

D'où $\text{Cov} \left(\mathbf{W}_{j,k'}^{(d)}(u), W_{j,k}^{(d)} \right) = 2^{2d(j-u)} \times \text{Cov} \left(\mathbf{W}_{u,k'}^{(d)}(u), W_{u,k}^{(d)} \right)$ et on obtient (2.92) en posant $j = J_1 + j$. L'approximation (2.93) découle alors de (2.92) et (2.91). ■

Dans ce qui suit, nous présentons les théorèmes de la limite centrale pour les estimateurs d_n^{LRW} et d_n^{LWW} établis respectivement dans Moulines et al. [2007a] et dans Moulines et al. [2008]. Le comportement asymptotique des estimateurs $\widehat{d}_n^{\text{LRW}}$ et $\widehat{d}_n^{\text{LWW}}$ définis en (2.75) et (2.79) est obtenu pour des séquences d'indices d'échelles $(J_0(n))$ et $(J_1(n))$ spécifiques. Dans un contexte semi-paramétrique, la séquence d'indice d'échelle la plus fine $(J_0(n))$ gouverne la vitesse de convergence de l'estimateur du paramètre de mémoire. Il existe deux considérations possibles:

(S-1) $J_1(n) - J_0(n)$ est fixé, et égal à $\ell > 0$

(S-2) $J_1(n) \leq J_n$ pour tout n et $J_1(n) - J_0(n) \rightarrow \infty$ quand $n \rightarrow \infty$, où J_n est l'échelle maximale définie en (2.74).

Dans ce mémoire, on ne s'intéressera qu'au cas (S-1). Pour le cas (S-2), voir Moulines et al. [2008].

Théorème 2.3.1. Soit X un processus $M(d)$ Gaussien de densité spectrale généralisée f définie en (2.11) tel que $f^* \in \mathcal{H}(\beta, L, \varepsilon)$ pour tout $L > 0$, $0 < \beta \leq 2$ et $\varepsilon \in (0, \pi]$. Supposons que les hypothèses (W-1)-(W-4) sont satisfaites avec $(1 + \beta)/2 - \alpha < d \leq M$. Alors, si, quand $n \rightarrow \infty$, $J_0(n)$ est tel que

$$n 2^{-(1+2\beta)J_0(n)} \rightarrow 0, \quad (2.94)$$

et $(J_1(n)) = (J_0(n) + \ell)$ une séquence satisfaisant (S-1), alors on a le théorème de la limite centrale suivant:

$$\sqrt{n 2^{-J_0(n)}} \left(\widehat{d}_n^{\text{LRW}}(J_0, \ell, \mathbf{w}) - d \right) \xrightarrow{d} \mathcal{N} \left(0, \mathbf{w}^T \mathbf{V}(d) \mathbf{w} \right), \quad (2.95)$$

où \xrightarrow{d} signifie convergence en loi et $\mathcal{N}(0, \mathbf{w}^T \mathbf{V}(d) \mathbf{w})$ désigne la distribution de la variable aléatoire Gaussienne centrée de variance $\mathbf{w}^T \mathbf{V}(d) \mathbf{w}$, où $\mathbf{V}(d)$ la matrice carrée $(1 + \ell) \times (1 + \ell)$ définie par

$$\mathbf{V}_{i,j}(d) \stackrel{\text{def}}{=} \frac{4\pi 2^{2d|i-j|} 2^{i \wedge j}}{\mathbf{K}(d)^2} \int_{-\pi}^{\pi} |\mathbf{D}_{\infty,|i-j|}(\lambda; d)|^2 d\lambda, \quad 0 \leq i, j \leq \ell, \quad (2.96)$$

avec $\mathbf{D}_{\infty,u}(\cdot; d)$ et $\mathbf{K}(d)$ définis respectivement en (2.84) et (2.90).

Remarque 2.3.1. Le théorème 2.3.1 est établi dans Moulines et al. [2007a] pour le cas où $\varepsilon = \pi$. Le cas $\varepsilon < \pi$ pouvant être obtenu par [Moulines et al., 2008, Corollaire 2]

Théorème 2.3.2. Soit X un processus $M(d)$ Gaussien de densité spectrale généralisée f définie en (2.11) tel que $f^* \in \mathcal{H}(\beta, L, \varepsilon)$ pour tout $L > 0$, $0 < \beta \leq 2$ et $\varepsilon \in (0, \pi]$. Supposons que les hypothèses (W-1)-(W-4) sont satisfaites avec $(1 + \beta)/2 - \alpha < d \leq M$. Soit $(J_0(n))$ une séquence satisfaisant

$$\lim_{n \rightarrow \infty} \left\{ n 2^{-(1+2\beta)J_0(n)} + J_0(n)^2 (n 2^{-J_0(n)})^{-1/4} \right\} = 0 \quad (2.97)$$

et $(J_1(n))$ une séquence satisfaisant (S-1). Alors, quand $n \rightarrow \infty$,

$$(n 2^{-J_0(n)})^{1/2} (\widehat{d}_n^{\text{LWW}}(J_0(n), \ell) - d) \xrightarrow{d} \mathcal{N}(0, \rho^2(d, \ell)), \quad (2.98)$$

où $\ell = J_1(n) - J_0(n)$ et $\rho^2(d, \ell)$ est donnée pour tout $\ell \geq 1$ par,

$$\rho^2(d, \ell) = \frac{\pi}{(2 - 2^{-\ell}) \kappa_\ell (\log(2) \mathbf{K}(d))^2} \times \left\{ \mathbf{I}_0(d) + \frac{2}{\kappa_\ell} \sum_{u=1}^{\ell} \mathbf{I}_u(d) 2^{(2d-1)u} \sum_{i=0}^{\ell-u} \frac{2^{-i}}{2 - 2^{-\ell}} (i - \eta_\ell)(i + u - \eta_\ell) \right\}, \quad (2.99)$$

où $\mathbf{K}(d)$ est défini en (2.90),

$$\eta_\ell \stackrel{\text{def}}{=} \sum_{j=0}^{\ell} j \frac{2^{-j}}{2 - 2^{-\ell}} \quad \text{et} \quad \kappa_\ell \stackrel{\text{def}}{=} \sum_{j=0}^{\ell} (j - \eta_\ell)^2 \frac{2^{-j}}{2 - 2^{-\ell}}, \quad (2.100)$$

et

$$\mathbf{I}_u(d) \stackrel{\text{def}}{=} \int_{-\pi}^{\pi} |\mathbf{D}_{\infty,u}(\lambda; d)|^2 d\lambda = (2\pi)^{-1} \sum_{\tau \in \mathbb{Z}} \text{Cov}^2 \left(W_{0,0}^{(d)}, W_{-u,\tau}^{(d)} \right), \quad (2.101)$$

Les théorèmes 2.3.1 et 2.3.2 ci-dessus ont été étendus aux processus linéaires par Roueff and Taquq [2009].

2.3.5 Estimateurs robustes du paramètre de mémoire d

Malgré toutes ces propriétés asymptotiques intéressantes, l'estimateur $\widehat{d}_n^{\text{LRW}}$ présente quelques limites. Stoev et al. [2006] ont montré que cet estimateur était sensible aux

valeurs aberrantes. En effet, il est bien connu que la variance empirique est très sensible à la présence des valeurs aberrantes dans les observations. Une faible proportion de valeurs atypiques, dans certains cas même une seule valeur peut affecter l'estimateur classique de la variance, le rendant ainsi pratiquement inutile ; voir par exemple Deutsch et al. [1990] Chan [1992], Chan [1995] [Maronna et al., 2006, Chapter 8] et les références qui s'y trouvent. Comme les données aberrantes sont assez courantes en pratique, la définition d'un estimateur de la variance qui est insensible aux observations aberrantes est d'une grande importance pratique (c'est une demande récurrente des chercheurs s'intéressant aux données hydrologiques ou aux trafics en particulier). Stoev et al. [2005] et Stoev et al. [2006] proposent de remplacer à chaque échelle de décomposition les moments empiriques d'ordre 2 (les scalogrammes) par la médiane empirique des carrés des coefficients d'ondelettes. Les auteurs de ces articles ont montré, à l'aide de nombreuses simulations que cette procédure était beaucoup moins sensible à la présence de données aberrantes que l'estimateur basé sur les scalogrammes classiques. Ces auteurs n'ont toutefois pas développés de résultats théoriques pour établir la consistance et la normalité asymptotique de l'estimateur ainsi construit. Dans le contexte similaire de l'estimation du coefficient d'autosimilarité d'un processus gaussien, Coeurjolly [2008b] a proposé des méthodes très similaires pour obtenir des estimateurs semi-paramétriques du paramètre de Hurst à l'aide des méthodes de filtrages discrets pour une classe de processus Gaussien centré à incrément stationnaire et localement auto-similaire. Voir Coeurjolly [2008b] et les références qui s'y trouvent.

Dans cette contribution, nous étudions les propriétés asymptotiques de l'estimateur proposé par Stoev et al. [2006], présentons et étudions les propriétés asymptotiques d'un nouvel estimateur robuste du paramètre de mémoire basé sur le carré de l'estimateur du paramètre d'échelle proposé par Rousseeuw and Croux [1993] et dont les propriétés asymptotiques sont étudiées dans Lévy-Leduc et al. [2009].

Estimateur robuste de Croux et Rousseeuw.

Soit X une variable aléatoire de fonctions de répartition F_X . Un estimateur robuste du paramètre d'échelle est l'estimateur proposé par Rousseeuw and Croux [1993]. Il est basé sur la corrélation intégrale de Grassberger-Procaccia définie par

$$r \mapsto U(r, F_X) = \iint \mathbb{1}_{\{|x-x'| \leq r\}} dF_X(x) dF_X(x'), \quad (2.102)$$

qui mesure la probabilité que deux copies indépendantes X et X' de fonction de répartition F_X tombent à une distance plus petite que r . L'estimateur robuste introduit par [Rousseeuw and Croux, 1993, p. 1277] définit l'échelle $\sigma_{CR}(F_X)$ d'une fonction de répartition F_X comme

étant proportionnelle au premier quartile de $r \mapsto U(r, F_X)$, à savoir,

$$\sigma_{\text{CR}}(F_X) \stackrel{\text{def}}{=} c(F_X) \inf \{r \geq 0, U(r, F_X) \geq 1/4\}, \quad (2.103)$$

où $c(F_X)$ est une constante dépendant uniquement de la forme de F_X . Ce paramètre d'échelle peut être vu comme un analogue à l'estimateur de la différence moyenne de Gini $n^{-1}(n-1)^{-1} \sum_{1 \leq i \neq j \leq n} |X_i - X_j|$, où la moyenne est remplacée par un quantile.

La constant $c(F_X)$ en (2.103) est là pour assurer la consistance. Dans la suite, la fonction de répartition F_X est supposée appartenir à la famille Gaussienne

$$\{\Phi_{\mu, \sigma}(\cdot) = \Phi((\cdot - \mu)/\sigma), \mu \in \mathbb{R}, \sigma \in \mathbb{R}_+^*\}, \quad (2.104)$$

où Φ est la fonction de répartition d'une variable aléatoire Gaussienne standard.

Soit X un processus Gaussien stationnaire. Etant données les observations (X_1, \dots, X_n) , la fonction de répartition des observations peut être estimée en utilisant la fonction de répartition empirique. $r \mapsto F_n(r) = n^{-1} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq r\}}$. En remplaçant F_X par F_n dans (2.103), on obtient l'estimateur robuste du paramètre d'échelle

$$\hat{\sigma}_{n, \text{CR}}[\Phi] = c(\Phi) \{ |X_i - X_j|; 1 \leq i, j \leq n \}_{(k_n)}, \quad (2.105)$$

où $k_n = \lfloor n^2/4 \rfloor$. Ainsi à une constante multiplicative $c(\Phi)$ près, $\hat{\sigma}_{n, \text{CR}}[\Phi]$ est la statistique du k_n -ième ordre de la distance $|X_i - X_j|$ entre toutes les paires d'observations. Comme mentionné dans Rousseeuw and Croux [1993], $\hat{\sigma}_{n, \text{CR}}[\Phi]$ a plusieurs propriétés intéressantes: il a une formule simple et explicite dont l'interprétation est intuitive; l'estimateur n'est pas affecté même si (50%) des observations sont aberrantes; De plus comme prouvée dans Lévy-Leduc et al. [2009] sa fonction d'influence est bornée. Pour une définition de ces quantités, qui sont classiques dans la statistique robuste, voir par exemple (Huber [1981]). L'estimateur Rousseeuw et Croux est aussi intéressant car il peut être implémenté assez efficacement. Sa complexité algorithmique est de $O(n \log n)$.

L'écart absolu moyen.

Bien que de nombreux estimateurs du paramètre d'échelle existent, la médiane empirique reste toujours la plus couramment utilisée. Si (X_1, \dots, X_n) sont des observations d'une variable aléatoire de fonction de répartition F_X , alors on notera sa médiane empirique par $\text{med}_i X_i$. Soit X un processus Gaussien stationnaire. Etant données les observations (X_1, \dots, X_n) , la fonction de répartition des observations peut être estimée en utilisant la fonction de répartition empirique. Un estimateur robuste du paramètre d'échelle est l'écart absolu moyen. donné par

$$\hat{\sigma}_{n, \text{MAD}}(\Phi) = m(\Phi) \left\{ \text{med}_{1 \leq i \leq n} \left| X_i - \text{med}_{1 \leq i' \leq n} X_{i'} \right| \right\}. \quad (2.106)$$

Comme dans le cas de l'estimateur de Rousseeuw et Croux la constante $m(\Phi)$ assure la consistance. Nous montrerons dans le chapitre 5 que cette constante est égale à $1/\Phi^{-1}(3/4) = 1.4826$. Cet estimateur possède également des propriétés intéressantes qui sont essentiellement les mêmes que celles de l'estimateur de Rousseeuw et Croux.

Résultats asymptotiques

Soit X un processus $M(d)$ Gaussien. Alors d'après (2.43), les coefficients d'ondelettes $\{W_{j,k}\}$ sont des processus Gaussien centrés. Le carré de l'écart absolu moyen défini en (2.106) est donné par

$$\hat{\sigma}_{\text{MAD},j}^2 = \left(m(\Phi) \operatorname{med}_{0 \leq i \leq n_j - 1} |W_{j,i}| \right)^2,$$

où les observations (X_1, \dots, X_n) sont remplacées par $(W_{j,0}, \dots, W_{j,n_j-1})$. De même, le carré de l'estimateur de Rousseeuw et Croux défini en (2.105) est donné par

$$\hat{\sigma}_{\text{CR},j}^2 = \left(c(\Phi) \{ |W_{j,i} - W_{j,k}|; 0 \leq i, k \leq n_j - 1 \}_{(k_{n_j})} \right)^2,$$

où $c(\Phi) = 2.21914$ voir (Lévy-Leduc et al. [2009]). En remplaçant dans (2.75) les scalogrammes par $\hat{\sigma}_{\text{CR},j}^2$ et $\hat{\sigma}_{\text{MAD},j}^2$, on obtient deux estimateurs robustes du paramètre de mémoire

$$\hat{d}_{*,n}^{\text{LRW}}(J_0, \mathbf{w}) \stackrel{\text{def}}{=} \sum_{j=J_0}^{J_0+\ell} w_{j-J_0} \log(\hat{\sigma}_{*,j}^2), \quad (2.107)$$

où $*$ désigne CR et MAD;

Théorème 2.3.3. *Soit X un processus $M(d)$ Gaussien de densité spectrale généralisée f définie en (2.11) tel que $f^* \in \mathcal{H}(\beta, L, \varepsilon)$ pour tout $L > 0$, $0 < \beta \leq 2$ et $\varepsilon \in (0, \pi]$. Supposons que les hypothèses (W-1)-(W-4) sont satisfaites avec $(1 + \beta)/2 - \alpha < d \leq M$. Alors, si, quand $n \rightarrow \infty$, $J_0(n)$ est tel que*

$$n2^{-(1+2\beta)J_0(n)} \rightarrow 0, \quad (2.108)$$

et $(J_1(n)) = (J_0(n) + \ell)$ une séquence satisfaisant (S-1), alors on a le théorème de la limite centrale suivant:

$$\sqrt{n2^{-J_0(n)}} \left(\hat{d}_{*,n}(J_0, \mathbf{w}) - d \right) \xrightarrow{d} \mathcal{N}(0, \mathbf{w}^T \mathbf{V}_*(d) \mathbf{w}), \quad (2.109)$$

où $\mathbf{V}_*(d)$ est la matrice $(1 + \ell) \times (1 + \ell)$ définie par

$$\mathbf{V}_{*,i,j}(d) = \sum_{p \geq 2} \frac{4c_p^2(\text{IF}_*)}{p! \text{K}(d)^p} 2^{pd|i-j|+i \wedge j} \sum_{\tau \in \mathbb{Z}} \sum_{r=0}^{2^{|i-j|}-1} \left(\int_{-\pi}^{\pi} \mathbf{D}_{\infty,|i-j|}^{(r)}(\lambda; d) e^{i\lambda\tau} d\lambda \right)^p, \quad 0 \leq i, j \leq \ell. \quad (2.110)$$

Dans (2.110), $\text{K}(d)$ est défini en (2.90), $\mathbf{D}_{\infty,|i-j|}(\cdot; d)$ est définie en (2.84), $c_p(\text{IF}_*) = \mathbb{E}[\text{IF}(X, *, \Phi) H_p(X)]$, où H_p est le p -ième polynôme d'Hermite et $\text{IF}(\cdot, *, \Phi)$ sont les fonctions d'influences définies en (5.17), (5.18) et (5.19).

La preuve du Théorème 2.3.3 est donnée dans le Chapitre 5, Section 5.6. Les performances de nos estimateurs ont été testées via des simulations Monte Carlo, sur les données hydrologiques du Nil et sur les données de trafic Internet. Toutes ces applications suggèrent fortement l'utilisation d'estimateurs robustes pour estimer d . Principalement l'estimateur $\widehat{d}_{CR,n}$ qui a une efficacité asymptotique relative qui oscille entre 0.63 et 0.79 pour une plage de valeurs de d allant au delà de l'intervalle de stationnarité.

2.3.6 Estimateur semi-paramétrique par maximum de vraisemblance local du paramètre de mémoire d

Dans cette sous section, nous utilisons la matrice de covariance des coefficients d'ondelettes définie pour deux échelles i et j par (2.69) pour proposer un estimateur par maximum de vraisemblance local du paramètre de mémoire d dans le domaine des ondelettes, pour un processus $M(d)$ Gaussien.

L'intérêt d'un tel estimateur est de prendre en compte la dépendance des coefficients d'ondelettes à la même échelle et aux différentes échelles.

Plus formellement, soit X_1, \dots, X_n n observations d'un processus $M(d)$ Gaussien X et $W_{j,k}$, $j \geq 1$, $0 \leq k < n_j$ les coefficients d'ondelettes associés. Rappelons que le nombre n_j de coefficients d'ondelettes à l'échelle j pour un échantillon de taille n , est défini par (2.53). Soit J_n l'échelle maximale correspondante, $J_n = \max\{j : n_j \geq 1\}$. De (2.53), on a $n2^{-j} - 2T \leq n_j \leq n2^{-j}$ ainsi $J_n = \log(n)/\log(2) + O(1)$. Il s'en suit que le nombre total de coefficients d'ondelettes satisfait la relation

$$\tilde{n} = \sum_{j=1}^{J_n} n_j \sim n/2 \quad \text{quand } n \rightarrow \infty .$$

Nous réorganisons ainsi les coefficients d'ondelettes $W_{j,k}$ dans un vecteur colonne en indexant (j, k) dans un ordre croissant. Ainsi, on pose $\mathbf{W}^{(n)} = [W_i^{(n)}]^T$ avec $i = 1, \dots, \tilde{n}$ tel que, pour tout $1 \leq j \leq J_n$ et $0 \leq k < n_j$, $W_{j,k} = W_{i(j,k)}^{(n)}$ où $i(j, k) = \sum_{i=1}^{j-1} n_i + k$.

Comme nous sommes dans un contexte Gaussien, en utilisant les hypothèses (H-i) et (H-ii), la moins log vraisemblance de $\mathbf{W}^{(n)}$ est donnée à une constante additive prête par

$$L(\boldsymbol{\beta}) = \frac{1}{2} \left(\log \left(\det(\mathbf{M}_{1, J_n}^{(n)}(\boldsymbol{\beta})) \right) + \left(\mathbf{W}^{(n)} \right)^T \mathbf{M}_{1, J_n}^{(n)-1}(\boldsymbol{\beta}) \mathbf{W}^{(n)} \right), \quad (2.111)$$

où $\mathbf{M}^{(n)}(\boldsymbol{\beta})$ désigne la matrice de covariance de $\mathbf{W}^{(n)}$ suivant le paramètre $\boldsymbol{\beta}$, i.e $\mathbf{M}^{(n)}(\boldsymbol{\beta})$ est la matrice tel que $\mathbf{M}_{i,j}[k, k']$ est définie par (2.69) pour $1 \leq j \leq J_n$ et $0 \leq k < n_j$, réorganisée en utilisant les mêmes indices que $\mathbf{W}^{(n)}$. De plus, la covariance en (2.69) est calculée en utilisant la distribution associée au paramètre $\boldsymbol{\beta}$.

Notre cadre d'estimation étant semi-paramétrique et local, on ne s'intéresse qu'à l'estimation de $\boldsymbol{\beta} = (d, \sigma^2)$ entre deux indices d'échelles J_1 et J_2 tel que $1 \leq J_1 \leq J_2 \leq J_n$.

Ainsi L en (2.111) se récrit

$$L_{J_1, J_2}(\boldsymbol{\beta}) = \frac{1}{2} \left(\log \left(\det(\mathbf{M}_{J_1, J_2}^{(n)}(\boldsymbol{\beta})) \right) + \left(\mathbf{W}_{J_1, J_2}^{(n)} \right)^T \mathbf{M}_{J_1, J_2}^{(n)-1}(\boldsymbol{\beta}) \mathbf{W}_{J_1, J_2}^{(n)} \right), \quad (2.112)$$

où $\mathbf{W}_{J_1, J_2}^{(n)} = \left[W_i^{(n)} \right]^T$ avec $i = 1, \dots, \check{n}$ et $\check{n} = \sum_{j=J_1}^{J_2} n_j$. Pour n'importe quel modèle semi-paramétrique contenant un paramètre d'échelle σ , i.e $\boldsymbol{\beta} = (d, \sigma^2)$ et $\mathbf{M}_{J_1, J_2}^{(n)}(\boldsymbol{\beta}) = \sigma^2 \mathbf{M}_{J_1, J_2}^{(n)}(d)$ (avec un léger abus de notation), la log vraisemblance (2.112) peut être optimisée en σ^2 comme une fonction de d donnant lieu à l'estimateur par maximum de vraisemblance du paramètre d'échelle entre les indices d'échelles J_1 et J_2

$$\hat{\sigma}_{J_1, J_2}^2(d) = \underset{\sigma^2 > 0}{\text{Argmin}} L_{J_1, J_2}(d, \sigma^2) = \frac{1}{\check{n}} \left(\mathbf{W}_{J_1, J_2}^{(n)} \right)^T \mathbf{M}_{J_1, J_2}^{-1}(d) \mathbf{W}_{J_1, J_2}^{(n)}.$$

Ainsi, l'estimateur semi-paramétrique par maximum de vraisemblance de d est définie par le minimum de la moins log vraisemblance profilée

$$L_{J_1, J_2}(d) = \log \left(\left(\mathbf{W}_{J_1, J_2}^{(n)} \right)^T \mathbf{M}_{J_1, J_2}^{-1}(d) \mathbf{W}_{J_1, J_2}^{(n)} \right) + \frac{1}{\check{n}} \log \left(\det \left(\mathbf{M}_{J_1, J_2}(d) \right) \right) \quad (2.113)$$

(encore avec un léger abus de notations).

Notons $\hat{d}_{J_1, J_2}^{\text{LW}}$, l'estimateur semi-paramétrique local obtenu en minimisant la moins log vraisemblance profilée (2.113). Nous comparons dans la suite les propriétés empiriques des estimateurs semi-paramétriques $\hat{d}_{J_1, J_2}^{\text{LW}}$ et $\hat{d}_{J_1, J_2}^{\text{WW}}$ via des simulations Monte carlo. Nous nous sommes focalisés sur des modèles ARFIMA(0, d , 0) et des modèles ARFIMA(1, d , 0) avec un paramètre AR(1) égal à 0.7 qui est une valeur relativement élevée. Nous avons également considéré des modèles DARFIMA qui sont processus ARFIMA présentant une discontinuité dans la densité spectrale à la fréquence λ_0 (voir Chapitre 4 et Andrews and Sun [2004]) pour une définition plus précise de ces processus. Dans toutes les simulations, nous avons fixé l'échelle la plus grossière $J_2 = J_n$ et utilisé les ondelettes de Daubechies avec $M = 4$. Pour obtenir $\hat{d}_{J_1, J_2}^{\text{LW}}$, nous avons ajusté un modèle ARFIMA(0, d , 0) sur les observations et estimer d en minimisant la moins log-vraisemblance profilée (2.113). Après plusieurs simulations, nous avons choisi de présenter les résultats pour $J_1 = 3, 4, 5$ qui correspondent aux échelles fines optimales pour $n = 4096$ observations. En effet choisir, J_1 trop petit peut introduire un biais dans l'estimation et J_1 trop grand une large variance voir (Figure 4.1 et Figure 4.2).

Toutes les deux méthodes semblent bien fonctionner avec des performances presque similaires pour des choix d'indice d'échelle fine J_1 optimaux. Cependant, le nouvel estimateur $\hat{d}_{J_1, J_2}^{\text{LW}}$ que nous proposons présentent des meilleures performances sur certains points.

- Sur un processus ARFIMA(0, d , 0), les racines carrées des erreurs quadratiques moyennes (RMSE) obtenues avec $\hat{d}_{J_1, J_2}^{\text{LW}}$ sont plus faibles que celles obtenues avec

$\hat{d}_{J_1, J_2}^{\text{LWW}}$ et ce pour une large plage de valeur de d voir Figure 2.5. On remarque également que en général, pour les modèles (ARFIMA(1, d , 0) et DARFIMA(1, d , 0)), les RMSE obtenues par la méthode de maximum de vraisemblance sont plus faibles que celles obtenues par la méthode Whittle lorsque $d > 1/2$ (voir Table 4.1 et Table 4.2).

- Quelque soit le processus considéré, (ARFIMA(0, d , 0), ARFIMA(1, d , 0) ou DARFIMA(1, d , 0)), l'écart type empirique obtenu par la méthode par maximum de vraisemblance en ondelettes est plus faible que celui obtenu par la méthode de Whittle (cf Table 4.1, Table 4.2, Figure 4.1 et Figure 4.2 du chapitre 4).
- On remarque également que les écarts types restent beaucoup plus stable pour différentes valeurs de d et J_1 fixé dans la méthode par maximum de vraisemblance comparée à la méthode de Whittle (voir Table 4.1 et Table 4.2). Ceci est intéressant pour calculer les intervalles de confiances : leurs tailles dépendant principalement de l'indice d'échelle J_1 connu et non du paramètre de mémoire d inconnu.

2.4 Test de détection de rupture

Dans cette partie de notre travail, nous proposons une procédure de test pour détecter des changements dans la densité spectrale généralisée d'un $M(d)$ processus. A la différence des méthodes proposées dans la littérature, notre test a la même distribution, que le processus soit à mémoire courte ou à mémoire longue. Notre procédure de test est formulée dans le domaine des ondelettes. Dans ce domaine, un changement dans la densité spectrale généralisée du processus entraîne un changement dans la variance des coefficients d'ondelettes à une ou plusieurs échelles.

2.4.1 Résultats principaux

Soient X_1, \dots, X_n n observations d'une série chronologique, et $W_{j,k}$ pour $(j, k) \in \mathcal{I}_n$ avec \mathcal{I}_n défini en (2.32) les coefficients d'ondelettes associés. Si X est un processus $M(d)$ et le nombre de moments nuls M de l'ondelette satisfait $M > d - 1/2$, alors $\Delta^M X$ est stationnaire et par conséquent, la variance des coefficients d'ondelettes à chaque échelle devrait être constante. Si cette hypothèse n'est pas satisfaite, alors on peut espérer que la variance des coefficients d'ondelettes puisse changer soit graduellement soit brutalement. Ceci motive donc la construction d'un test pour étudier la constance de la variance des coefficients d'ondelettes. Si $\Delta^M X$ est stationnaire et non centré, on peut le centrer en remplaçant M par $M + 1$; d'où sans nuire à la généralité on peut supposer que $\Delta^M X$ est stationnaire et centré, ce qui sera considéré comme notre hypothèse nulle \mathcal{H}_0 . Sous

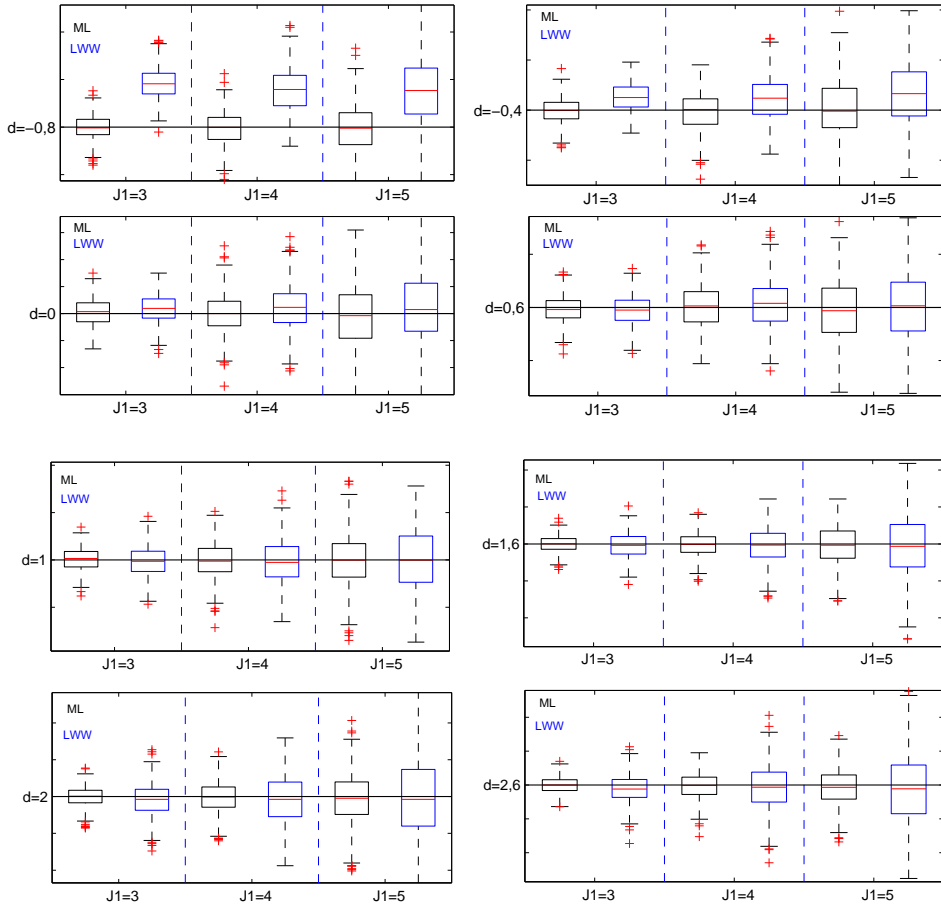


Figure 2.5: Boxplots des estimateurs \hat{d}_{J_1, J_2}^{LW} et \hat{d}_{J_1, J_2}^{LWW} sur des processus ARFIMA(0, d , 0), $d \in \{-0.8, -0.4, 0, 0.2, 0.6, 1, 1.6, 2, 2.6\}$.

\mathcal{H}_0 le processus des coefficients d'ondeletes $\{W_{j,k}, k \in \mathbb{Z}\}$ est centré et stationnaire. La somme cumulée des carrés (CUSUM) utilisée par Inclin and Tiao [1994] pour détecter les changements dans la variance peut alors être adaptée pour obtenir un test d'hypothèse nulle basée sur la somme cumulée des carrés des coefficients d'ondelettes. Cette idée est développée dans le chapitre 3 ci dessous, où par ailleurs les changements dans la variance se produisant simultanément à plusieurs échelles sont considérés résultant ainsi à une procédure de test multi-échelle. Décrivons brièvement la procédure et les résultats principaux qui en découlent. Pour plus de détails voir (Kouamo et al. [2010])

Etant données $J_1 \leq J_2$, définissons le vecteur de scalogrammes

$$Y_{J_1, J_2}[i] = \left[W_{J_2, i}^2, \sum_{u=1}^2 W_{J_2-1, 2(i-1)+u}^2, \dots, \sum_{u=1}^{2^{(J_2-J_1)}} W_{J_1, 2^{(J_2-J_1)}(i-1)+u}^2 \right]^T \quad (2.114)$$

et notons le processus de sommes partielles correspondant par

$$S_{J_1, J_2}(t) = \frac{1}{\sqrt{n_{J_2}}} \sum_{i=1}^{\lfloor n_{J_2} t \rfloor} Y_{J_1, J_2}[i]. \quad (2.115)$$

Notons également Γ_{J_1, J_2} la matrice de covariance des scalogrammes :

$$\Gamma_{J_1, J_2}(j' - J_1 + 1, j - J_1 + 1) = \sum_{\tau \in \mathbb{Z}} \text{Cov} \left(Y_{J_1, J_2}^{j'}[\tau], Y_{J_1, J_2}^j[0] \right), \quad \text{avec } J_1 \leq j' \leq j \leq J_2. \quad (2.116)$$

Alors pour une large classe de processus Gaussiens, sous \mathcal{H}_0 et sous la condition que l'ondelette d'analyse ait au moins M moments nuls, il est montré dans Kouamo et al. [2010] que

$$\Gamma_{J_1, J_2}^{-1/2} (S_{J_1, J_2}(t) - \mathbb{E}[S_{J_1, J_2}(t)]) \xrightarrow{d} B(t) = (B_{J_1}(t), \dots, B_{J_2}(t)), \quad (2.117)$$

où la convergence en distribution est obtenue dans l'espace $D^{J_2 - J_1 + 1}[0, 1]$ des fonctions càdlàg définies sur $[0, 1]$ et à valeurs dans $\mathbb{R}^{J_2 - J_1 + 1}$, $\{B_j(t), t \geq 0\}_{j=J_1, \dots, J_2}$ sont des *mouvements Browniens* indépendants et Γ_{J_1, J_2} désigne la matrice de covariance asymptotique de $S_{J_1, J_2}(1)$. $D^{J_2 - J_1 + 1}[0, 1]$ est équipé de la métrique de Skorokhod.

Comme la matrice Γ_{J_1, J_2} est inconnue, elle doit être estimée. Notons $\widehat{\Gamma}_{J_1, J_2}$ un estimateur consistant de Γ_{J_1, J_2} . Le test est basé sur les statistiques $T_{J_1, J_2} : [0, 1] \rightarrow \mathbb{R}_+$ définies par

$$T_{J_1, J_2}(t) \stackrel{\text{def}}{=} (S_{J_1, J_2}(t) - tS_{J_1, J_2}(1))^T \widehat{\Gamma}_{J_1, J_2}^{-1} (S_{J_1, J_2}(t) - tS_{J_1, J_2}(1)), \quad t \in [0, 1]. \quad (2.118)$$

Comme conséquence de (2.117), $T_{J_1, J_2}(t)$ converge également en distribution dans l'espace de Skorokhod $D([0, 1])$,

$$T_{J_1, J_2}(t) \xrightarrow{d} \sum_{\ell=J_1}^{J_2} [B_\ell^0(t)]^2 \quad (2.119)$$

où $\{B_j^0(t), t \geq 0\}_{j=J_1, \dots, J_2}$ sont des ponts Browniens indépendants. Pour toute fonction continue $F : D[0, 1] \rightarrow \mathbb{R}$, Le théorème de continuité entraîne que

$$F[T_{J_1, J_2}(\cdot)] \xrightarrow{d} F \left[\sum_{\ell=1}^{J_2 - J_1 - 1} [B_\ell^0(\cdot)]^2 \right].$$

On peut appliquer par exemple soit l'intégrale soit le sup pour obtenir une statistique de test de type Cramér-Von Mises ou de type Kolmogorov-Smirnov, respectivement

$$\text{CVM}(J_1, J_2) \stackrel{\text{def}}{=} \int_0^1 T_{J_1, J_2}(t) dt, \quad (2.120)$$

$$\text{KSM}(J_1, J_2) \stackrel{\text{def}}{=} \sup_{0 \leq t \leq 1} T_{J_1, J_2}(t). \quad (2.121)$$

Voir Kouamo et al. [2010] pour les détails de calculs des quantiles.

2.4.2 Estimation de Γ_{J_1, J_2}

Comme les coefficients d'ondelettes sont corrélés, nous proposons dans cette contribution trois estimateurs de la matrice des carrés des coefficients d'ondelettes pour les échelles $\{J_1, \dots, J_2\}$.

- L'estimateur non paramétrique de Bartlett noté $\widehat{\Gamma}_{J_1, J_2}^B$,
- l'estimateur semi-paramétrique noté $\widehat{\Gamma}_{J_1, J_2}^{LLW}$ et basé sur des estimateurs par maximum de vraisemblance semi-paramétriques des covariance des coefficients d'ondelettes entre les échelles J_1 et J_2 ,
- l'estimateur semi-paramétrique asymptotique $\widehat{\Gamma}_{J_2 - J_1}^{(d)}$ basé sur des estimateurs semi-paramétriques des covariances des coefficients d'ondelettes du GFBM. Nous supposons ici que $J_2 - J_1$ est fixe et égal à ℓ .

Estimateur de Bartlett $\widehat{\Gamma}_{J_1, J_2}^B$

L'estimateur de Bartlett est utilisé dans le chapitre 3 pour renormaliser la statistique $T_{J_1, J_2}(t)$ définie en (2.118). Il est donné par

$$\widehat{\Gamma}_{J_1, J_2}^B = \sum_{\tau=-q(n_{J_2})}^{q(n_{J_2})} w_\tau[q(n_{J_2})] \widehat{\gamma}_{J_1, J_2}(\tau), \quad \text{où} \quad (2.122)$$

$$\widehat{\gamma}_{J_1, J_2}(\tau) = \frac{1}{n_{J_2}} \sum_{i, i+\tau=1}^{n_{J_2}} (Y_{J_1, J_2}[i] - \bar{Y}_{J_1, J_2}) (Y_{J_1, J_2}[i + \tau] - \bar{Y}_{J_1, J_2})^T. \quad (2.123)$$

où, pour un entier q , les poids de Bartlett sont définis par

$$w_l(q) = 1 - \frac{|l|}{1 + q}, \quad l \in \{-q, \dots, q\}. \quad (2.124)$$

L'estimateur de Bartlett est un estimateur non paramétrique donc ne suppose aucun modèle sous l'hypothèse nulle. D'un autre coté le choix de la taille de la fenêtre est difficile en pratique et peut conduire à des estimateurs très différents.

Estimateur $\widehat{\Gamma}_{J_1, J_2}^{LLW}$

L'estimateur $\widehat{\Gamma}_{J_1, J_2}^{LLW}$ est considérée dans le chapitre 4 comme une alternative à l'estimateur $\widehat{\Gamma}_{J_1, J_2}^B$. Cette estimateur est basé sur une pré-estimation d'un modèle semi-paramétrique. Une fois le modèle estimé, l'estimateur de Γ_{J_1, J_2} peut être obtenu de deux façons différentes.

- Il peut être basé sur les auto-covariances et les cross-covariances des coefficients d'ondelettes aux différentes échelles,
- ou sur la densité spectrale intra et inter échelle des coefficients d'ondelettes.

Indexant $\Gamma_{J_1, J_2}[i, j]$ par les indices i et j , on a , pour $J_1 \leq i \leq j \leq J_2$,

$$\begin{aligned} \Gamma_{J_1, J_2}[i, j] &= \sum_{\tau \in \mathbb{Z}} \text{Cov} \left(\sum_{l=0}^{2^{J_2-j}-1} W_{j,l}^2, \sum_{u=0}^{2^{J_2-i}-1} W_{i, 2^{J_2-i}\tau+u}^2 \right) \\ &= 2 \sum_{\tau \in \mathbb{Z}} \sum_{l=0}^{2^{J_2-j}-1} \sum_{u=0}^{2^{J_2-i}-1} \text{Cov}^2 (W_{j,l}, W_{i, 2^{J_2-i}\tau+u}) \\ &= 2^{J_2-j+1} \sum_{s \in \mathbb{Z}} \text{Cov}^2 [W_{j,0}, W_{i,s}]. \end{aligned}$$

Cette expression est utile lorsque la covariance du modèle estimé est facile à calculer, ce qui donne un calcul rapide des covariances par la Proposition 2.2.1. Cependant, seule une approximation de la somme infinie ci-dessus peut être calculée par troncature. Le calcul de cette matrice peut également être obtenu dans le domaine spectrale en utilisant la formule de Parseval

$$\begin{aligned} 2^{J_2-j+1} \sum_{s \in \mathbb{Z}} \text{Cov}^2 [W_{j,0}, W_{i,s}] &= 2^{J_2-j+1} \sum_{t \in \mathbb{Z}} \sum_{v=0}^{2^{j-i}-1} \text{Cov}^2 (W_{j,0}, W_{i, t2^{j-i}+v}) \\ &= \frac{2^{J_2-j}}{\pi} \int_{-\pi}^{\pi} \|\mathbf{D}_{j,u}(\lambda)\|^2 d\lambda. \end{aligned}$$

Ici également la formule obtenue ne peut pas être calculée explicitement mais peut être approximée avec un niveau arbitraire de précision en approximant l'intégrale $\int_{-\pi}^{\pi} \|\mathbf{D}_{j,u}(\lambda; f)\|^2 d\lambda$ par une somme de Riemann.

Sinon, une fois qu'un modèle spécifique a été estimé, il semble plus exacte de calculer la matrice non-asymptotique de $S_{J_1, J_2}(1)$ en (2.115), donnée par

$$\begin{aligned} \tilde{\Gamma}_{J_1, J_2}[i, j] &= \frac{1}{n_{J_2}} \sum_{k, k'=0}^{n_{J_2}-1} \text{Cov} \left(\sum_{l=0}^{2^{J_2-j}-1} W_{j, 2^{J_2-j}k+l}^2, \sum_{u=0}^{2^{J_2-i}-1} W_{i, 2^{J_2-i}k'+u}^2 \right) \\ &= \frac{2}{n_{J_2}} \sum_{s, s'=0}^{2^{J_2-j}n_{J_2}-1} \sum_{v=0}^{2^{j-i}-1} \text{Cov}^2 (W_{j,s}, W_{i, s'2^{j-i}+v}), \end{aligned}$$

où on pose $s = 2^{J_2-j}k + l$ et $2^{j-i}s' + v = 2^{J_2-i}k' + u$, en supposant que les coefficients d'ondelettes sont Gaussiens. La dernière expression s'écrit alors

$$\tilde{\Gamma}_{J_1, J_2}[i, j] = 2 \sum_{\tau=-2^{J_2-j}n_{J_2}+1}^{2^{J_2-j}n_{J_2}-1} (2^{J_2-j} - |\tau|/n_{J_2}) \sum_{v=0}^{2^{j-i}-1} \text{Cov}^2 [W_{j,0}, W_{i, 2^{j-i}\tau+v}].$$

Observons que cette matrice de covariance peut être calculer explicitement car elle implique un nombre fini de coefficients de covariances.

Estimateur $\widehat{\Gamma}_\ell^{(d)}$.

Proposition 2.4.1. *Soient i et j deux indices d'échelles tel que $J_1 \leq i \leq j \leq J_2$ et Γ_{J_1, J_2} la matrice de covariance des carrés des coefficients d'ondelettes définie en (2.116). En posant $\ell = J_2 - J_1$ nous avons la convergence composante par composante qui suit*

$$2^{-4dJ_1} \Gamma_{J_1, J_2} \rightarrow (f^*(0))^2 G_\ell(d) \text{ quand } J_1 \rightarrow \infty, \quad (2.125)$$

où $G_\ell(d) = \sum_{\tau \in \mathbb{Z}} \text{Cov} \left(Y_{j,0}^{(d)}, Y_{i,\tau}^{(d)} \right)$, avec $0 \leq i \leq j \leq \ell$ et

$$Y_{\ell,k}^{(d)} = \left(\left(W_{\ell,k}^{(d)} \right)^2, \dots, \sum_{l=0}^{2^{\ell-j}-1} \left(W_{j,2^{\ell-j}k+l}^{(d)} \right)^2, \dots, \sum_{l=0}^{2^\ell-1} \left(W_{0,2^\ell k+l}^{(d)} \right)^2 \right).$$

Preuve de la Proposition 2.4.1.

Considérons pour les indices d'échelles $J_1, \dots, J_1 + \ell$ le processus

$$Y_{J_1, \ell, k} = \left(W_{J_1+\ell, k}^2, \dots, \sum_{l=0}^{2^{\ell-j}-1} W_{J_1+j, 2^{\ell-j}k+l}^2, \dots, \sum_{l=0}^{2^\ell-1} W_{J_1, 2^\ell k+l}^2 \right).$$

Pour $0 \leq i \leq j \leq \ell$, on a

$$\begin{aligned} \Gamma_{J_1, J_2}(i, j) &= \sum_{\tau \in \mathbb{Z}} \text{Cov} \left(\sum_{l=0}^{2^{\ell-j}-1} W_{J_1+j, l}^2, \sum_{u=0}^{2^{\ell-i}-1} W_{J_1+i, 2^{\ell-i}\tau+u}^2 \right) \\ &= \sum_{\tau \in \mathbb{Z}} \sum_{l=0}^{2^{\ell-j}-1} \sum_{u=0}^{2^{\ell-i}-1} \text{Cov}^2 \left(W_{J_1+j, 0}, W_{J_1+i, u+2^{l-i}-2^{j-i}l} \right). \end{aligned}$$

D'après la propriété d'auto-similarité de $\{W_{j,k}^{(d)}\}$, on a également

$$\begin{aligned} &\sum_{\tau \in \mathbb{Z}} \text{Cov} \left(\sum_{l=0}^{2^{\ell-j}-1} \left(W_{J_1+j, l}^{(d)} \right)^2, \sum_{u=0}^{2^{\ell-i}-1} \left(W_{J_1+i, 2^{\ell-i}\tau+u}^{(d)} \right)^2 \right) \\ &= 2^{4d(J_1+i)} \sum_{\tau \in \mathbb{Z}} \sum_{l=0}^{2^{\ell-j}-1} \sum_{u=0}^{2^{\ell-i}-1} \text{Cov}^2 \left(W_{j-i, 0}^{(d)}, W_{0, u+2^{\ell-i}\tau-2^{j-i}l}^{(d)} \right) \\ &= 2^{4dJ_1} G_\ell(d, i, j). \end{aligned}$$

D'où la limite (2.125) provient de (2.93) et des deux calculs ci-dessus. ■

Calcul des covariances des coefficients d'ondelettes du GFBM

Nous présentons ici une méthode de calcul de la covariance des coefficients d'ondelettes du GFBM qui est une approximation de la covariance des coefficients d'ondelettes d'un processus $M(d)$ à une constante multiplicative prête. Nous supposons ici que $d < M/2$

où M est le nombre de moments nuls de l'ondelette. Définissons pour deux échelle i et j avec $i \leq j$

$$\begin{aligned}\delta_{i,j}^{(d)}[k] &\stackrel{\text{def}}{=} \text{Cov} \left(W_{j,0}^{(d)}, W_{i,k}^{(d)} \right) = \int_{\mathbb{R}} |\xi|^{-2d} \widehat{\psi}_{i,k}(\xi) \overline{\widehat{\psi}_{j,0}(\xi)} d\xi \\ &= 2^{i/2} 2^{j/2} \int |\xi|^{-2d} e^{-i2^i \xi k} \widehat{\psi}(2^i \xi) \overline{\widehat{\psi}(2^j \xi)} d\xi \quad \text{et} \\ \gamma_{i,j}^{(d)}[k] &\stackrel{\text{def}}{=} \text{Cov} \left(A_{j,0}^{(d)}, W_{i,k}^{(d)} \right) = \int_{\mathbb{R}} |\xi|^{-2d} \widehat{\psi}_{i,k}(\xi) \overline{\widehat{\phi}_{j,0}(\xi)} d\xi \\ &= 2^{i/2} 2^{j/2} \int |\xi|^{-2d} e^{-i2^i \xi k} \widehat{\psi}(2^i \xi) \overline{\widehat{\phi}(2^j \xi)} d\xi ,\end{aligned}$$

où $A_{j,0}^{(d)}$ est l'approximation de $B^{(d)}$ à l'échelle j et en position 0, $\phi \in L^2(\mathbb{R})$ et $\psi \in L^2(\mathbb{R})$ respectivement les fonctions d'échelle et d'ondelette et $\widehat{\phi}$, $\widehat{\psi}$ les transformées de Fourier associées. Il est facile de voir que que les suites $\{\delta_{i,j}^{(d)}[k]\}$ et $\{\gamma_{i,j}^{(d)}[k]\}$ satisfont la relation d'auto-similarité :

$$\delta_{i,j}^{(d)}[k] = 2^{2di} \delta_{0,j-i}^{(d)}[k] \quad \text{and} \quad \gamma_{i,j}^{(d)}[k] = 2^{2di} \gamma_{0,j-i}^{(d)}[k] . \quad (2.126)$$

La relation d'échelle $\phi(x/2) = \sqrt{2} \sum_{l=-\infty}^{\infty} h[l] \phi(x-l)$ implique que

$$\widehat{\phi}(2\xi) = \frac{1}{\sqrt{2}} \sum_{l=-\infty}^{\infty} h[l] e^{-i\xi l} \widehat{\phi}(\xi) ,$$

où $\{h[l]\}$ est définie en (2.45) et ainsi,

$$\gamma_{0,j+1}^{(d)}[k] = \sum_{l=-\infty}^{\infty} h[l] \gamma_{0,j}^{(d)}[k - 2^j l] . \quad (2.127)$$

De façon similaire, la relation d'échelle pour la fonction d'ondelette $\psi(x/2) = \sqrt{2} \sum_{l=-\infty}^{\infty} g[l] \psi(x-l)$ implique que

$$\widehat{\psi}(2\xi) = \frac{1}{\sqrt{2}} \sum_{l=-\infty}^{\infty} g[l] e^{-i\xi l} \widehat{\psi}(\xi) ,$$

et ainsi,

$$\delta_{0,j+1}^{(d)}[k] = \sum_{l=-\infty}^{\infty} g[l] \delta_{0,j}^{(d)}[k - 2^j l] , \quad (2.128)$$

où $\{g[l]\}$ est définie en (2.48). Pour initialiser ces deux récursions, il est nécessaire d'évaluer les suites $\{\delta_{0,0}^{(d)}[k]\}$ et $\{\gamma_{0,0}^{(d)}[k]\}$.

$$\begin{aligned}\delta_{0,0}^{(d)}[k] &= \int_{\mathbb{R}} |\xi|^{-2d} e^{-i\xi k} \left| \widehat{\psi}(\xi) \right|^2 d\xi \\ \gamma_{0,0}^{(d)}[k] &= \int_{\mathbb{R}} |\xi|^{-2d} e^{-i\xi k} \widehat{\psi}(\xi) \overline{\widehat{\phi}(\xi)} d\xi .\end{aligned}$$

Ces deux intégrales ne peuvent pas être calculer explicitement mais peuvent être approximées avec un niveau arbitraire de précision par des sommes de Riemann en utilisant

la transformée de Fourier rapide. Les fonctions d'ondelettes et d'échelles peuvent être évaluées exactement sur les entiers dyadiques $\{k2^{-J_\phi}\}$ en utilisant l'algorithme pyramidal. Notons $T = 2^{-J_\phi}$ la période d'échantillonnage, $\Phi[k] \stackrel{\text{def}}{=} \{\phi(kT)\}$ l'échantillon des fonctions d'échelles et $\Psi[k] \stackrel{\text{def}}{=} \{\psi(kT)\}$ l'échantillon des fonctions d'ondelettes. Les transformées de Fourier discrètes de ces deux séquences renormalisées par T , évaluées aux fréquences de Fourier $2\pi k/N_F$ (où N_F est le nombre de coefficients de Fourier) sont notées $\{\widehat{\Phi}[k]\}$ et $\{\widehat{\Psi}[k]\}$ respectivement.

$$\widehat{\Phi}[k] = T \sum_{l=0}^{N_F-1} \Phi[k] e^{-i2\pi kl/N_F} \quad \text{et} \quad \widehat{\Psi}[k] = T \sum_{l=0}^{N_F-1} \Psi[k] e^{-i2\pi kl/N_F} .$$

Si on néglige l'erreur de discrétisation,

$$\widehat{\Phi}[k] = \widehat{\phi}\left(\frac{2\pi k}{TN_F}\right) \quad \text{et} \quad \widehat{\Psi}[k] = \widehat{\psi}\left(\frac{2\pi k}{TN_F}\right) , \quad k = 0, \dots, N_F/2 .$$

Notons que dans le but d'accélérer le calcul, toutes ces quantités peuvent être évaluées une fois et pré-enregistrées. Comme

$$\begin{aligned} \int_{-\pi/T}^{\pi/T} |\xi|^{-2d} e^{-i\xi k} \left| \widehat{\psi}(\xi) \right|^2 d\xi &= T^{2d-1} \int_{-\pi}^{\pi} |\xi|^{-2d} e^{-i\xi k/T} \left| \widehat{\psi}(\xi/T) \right|^2 d\xi \\ &\approx \left(\frac{2\pi}{N_F} \right) T^{2d-1} \sum_{l=-N_F/2}^{N_F/2} \left| \frac{2\pi l}{N_F} \right|^{-2d} e^{-i2\pi lk/N_F T} \left| \widehat{\psi}\left(\frac{2\pi l}{N_F}\right) \right|^2 . \end{aligned}$$

On obtient, l'algorithme suivant

i- Initialisation

Evaluer les séquences $\{\delta_{0,0}^{(d)}[k]\}$, et $\{\gamma_{0,0}^{(d)}[k]\}$ for $k = -n_0 + 1, \dots, n_0 - 1$. Pour $i = J_1, \dots, J_2$, posons

$$\text{Cov}\left(W_{i,k}^{(d)}, W_{i,k'}^{(d)}\right) = 2^{2di} \text{Cov}\left(W_{0,k}^{(d)}, W_{0,k'}^{(d)}\right) = 2^{2di} \delta_{0,0}^{(d)}[k' - k] , \quad k, k' = 0, \dots, n_i - 1 .$$

ii- Boucle principale

Pour $u = 0$ à $J_2 - J_1$, calculer les séquences $\gamma_{0,u+1}^{(d)} = \{\gamma_{0,u+1}^{(d)}[k]\}$, $\delta_{0,u+1}^{(d)} = \{\delta_{0,u+1}^{(d)}[k]\}$ suivant les relations

$$\begin{aligned} \gamma_{0,u+1}^{(d)}[k] &= \sum_{l=0}^{N-1} h[l] \gamma_{0,u}^{(d)}[k - 2^u l] , \quad \delta_{0,u+1}^{(d)}[k] = \sum_{l=0}^{N-1} g[l] \gamma_{0,u}^{(d)}[k - 2^u l] , \quad \text{et} \\ \text{Cov}\left(W_{i,k}^{(d)}, W_{i+u+1,k'}^{(d)}\right) &= 2^{2di} \text{Cov}\left(W_{u+1,0}^{(d)}, W_{0,k-2^{u+1}k'}^{(d)}\right) \\ &= 2^{2di} \delta_{0,u+1}^{(d)}[k - 2^{u+1}k'] . \end{aligned}$$

On peut également calculer $\{\delta_{0,u}^{(d)}[k]\}$ directement à chaque étape (au lieu d'utiliser l'algorithme pyramidal). L'avantage de cette procédure réside dans le fait qu'il peut être

appliqué avec $d \leq M$ (au lieu de $d \leq M/2$). Le prix à payer est que cette méthode exige de calculer la transformée de Fourier discrète pour chaque $u = 1, \dots, J_2 - J_1$ et donc prend plus de temps pour être calculée. En utilisant les mêmes notations que ci dessus, la séquence $\{\delta_{0,u}^{(d)}[k]\}$ peut être calculée comme suit

$$\begin{aligned} \delta_{0,u}^{(d)}[k] &\approx \left(\frac{2\pi}{N_F}\right) 2^{u/2} T^{2d-1} \sum_{l=-N_F/2}^{N_F/2} \left|\frac{2\pi l}{N_F}\right|^{-2d} e^{-i2\pi lk/N_F T} \widehat{\psi}\left(\frac{2\pi l}{N_F}\right) \overline{\widehat{\psi}\left(2^u \frac{2\pi l}{N_F}\right)} \\ &= \left(\frac{2\pi}{N_F}\right) 2^{u/2} T^{2d-1} \sum_{l=-N_F/2}^{N_F/2} \left|\frac{2\pi l}{N_F}\right|^{-2d} e^{-i2\pi lk/N_F T} \widehat{\Psi}^{(0)}[l] \overline{\widehat{\Psi}^{(u)}[l]}, \end{aligned}$$

où on note

$$\widehat{\Phi}^{(u)}[k] \stackrel{\text{def}}{=} \widehat{\phi}\left(2^u \frac{2\pi k}{TN_F}\right) \quad \text{et} \quad \widehat{\Psi}^{(u)}[k] \stackrel{\text{def}}{=} \widehat{\psi}\left(2^u \frac{2\pi k}{TN_F}\right).$$

Par l'algorithme pyramidal, les transformées de Fourier discrètes $\{\widehat{\Phi}^{(u)}[k]\}$ et $\{\widehat{\Psi}^{(u)}[k]\}$ peuvent être calculées récursivement comme suit:

$$\widehat{\Phi}^{(u+1)}[k] \stackrel{\text{def}}{=} \frac{1}{\sqrt{2}} \widehat{h}\left(2^u \frac{2\pi k}{TN_F}\right) \widehat{\Phi}^{(u)}[k] \quad \text{et} \quad \widehat{\Psi}^{(u+1)}[k] \stackrel{\text{def}}{=} \frac{1}{\sqrt{2}} \widehat{g}\left(2^u \frac{2\pi k}{TN_F}\right) \widehat{\Psi}^{(u)}[k]$$

Notons finalement que $\{\widehat{h}(2^u 2\pi k/TN_F)\}$ et $\{\widehat{g}(2^u 2\pi k/TN_F)\}$ sont les transformées de Fourier discrètes des séquences $\{h[l]\}$ et $\{g[l]\}$ sur-échantillonnées par un facteur $2^u/T$.

Calcul de la matrice de covariance asymptotique des scalogrammes $\Gamma_\ell^{(d)}$.

Nous présentons maintenant une autre méthode pour estimer Γ_{J_1, J_2} à partir des observations X_1, \dots, X_n . De (2.88), on a

$$\sigma_j^2(d, f^*) \approx A 2^{2d(j-J_1)},$$

avec $A = f^*(0) K(d) 2^{2J_1 d}$. Comme la séquence $\{\widehat{\sigma}_j^2\}_{j=J_1, \dots, J_2}$ est un estimateur consistant de la séquence $\{\sigma_j^2(d, f^*)\}_{j=J_1, \dots, J_2}$, un estimateur semi-paramétrique naturel de $(\log(A), d)$ est

$$\begin{bmatrix} \log \widehat{A}(\ell, \mathbf{w}) \\ \widehat{d}(\ell, \mathbf{w}) \end{bmatrix} = \mathbf{w}^T (\log(\widehat{\sigma}_j^2))_{j=J_1, \dots, J_1+\ell}^T, \quad (2.129)$$

la matrice $(\ell + 1) \times 2$

$$\mathbf{w} \stackrel{\text{def}}{=} \begin{bmatrix} w_{0,1} & \dots & w_{\ell,1} \\ w_{0,2} & \dots & w_{\ell,2} \end{bmatrix}^T,$$

de poids associés à d et A satisfait $\sum_{j=0}^{\ell} w_{j,2} = 0$, $\sum_{j=0}^{\ell} w_{j,1} = 1$, $\sum_{j=0}^{\ell} j w_{j,2} = 0$ et $2 \log(2) \sum_{j=0}^{\ell} j w_j = 1$. Pour $J_1 \geq 1$ et $\ell > 1$, on peut choisir par exemple \mathbf{w} correspondant à la matrice de régression des moindres carrés définie par $\mathbf{w} = DB(B^T DB)^{-1} \mathbf{b}$ où

$$\mathbf{b} \stackrel{\text{def}}{=} \begin{bmatrix} 1 & 0 \\ 0 & (2 \log(2))^{-1} \end{bmatrix}, \quad B \stackrel{\text{def}}{=} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 0 & 1 & \dots & \ell \end{bmatrix}^T$$

et D est une matrice symétrique définie positive arbitraire.

Finalement, on peut alors estimer la matrice de covariance Γ_{J_1, J_2} par $\widehat{\Gamma}_\ell$ où pour tout $0 \leq i \leq j \leq \ell$,

$$\widehat{\Gamma}_\ell^{(\widehat{d})}(i, j) = \frac{2^{4\widehat{d}i+1}\widehat{A}^2}{\mathbf{K}(\widehat{d})^2} \sum_{\tau \in \mathbb{Z}} \sum_{u=0}^{2^{\ell-i}-1} \sum_{l=0}^{2^{\ell-j}-1} \left(\delta_{0, j-i}^{(d)} \right)^2 \left[u + 2^{\ell-i}\tau + 2^{j-i}l \right], \quad (2.130)$$

$$= \frac{2^{4\widehat{d}i+1+\ell-j}\widehat{A}^2}{\mathbf{K}(\widehat{d})^2} \sum_{s \in \mathbb{Z}} \left(\delta_{0, j-i}^{(d)} \right)^2 [s], \quad (2.131)$$

avec \widehat{d} et \widehat{A} définies en (2.129).

Nous avons évalué sur plusieurs $M(d)$ processus, les niveaux des statistiques de test KSM et CVM définis respectivement en (2.121) et en (2.120) en utilisant des simulations Monte Carlo. Les résultats obtenus aux Tableaux 4.3 du chapitre 4 suggèrent fortement l'utilisation de la matrice $\widehat{\Gamma}_{J_1, J_2}^{\text{LW}}$ par rapport aux matrices $\widehat{\Gamma}_{J_1, J_2}^{\text{B}}$ et $\widehat{\Gamma}_{J_1, J_2}^{(d)}$ pour la renormalisation de la statistique $T_{J_1, J_2}(t)$ définie en (2.119). Nous appliquons également nos procédures de test sur les valeurs absolues des log retours de l'indice S&P 500 observés du 03-01-1950 au 27-10-2010 pour détecter plusieurs points de changement voir sous-section 4.4.4 du chapitre 4.

Chapter 3

Change points detection in wavelet domain

The material of this chapter is the article entitled "Testing for homogeneity of variance in the wavelet domain" Kouamo et al. [2010] published in Lecture Notes of Statistics Springer Verlag.

3.1 Introduction

For time series of short duration, stationarity and short-range dependence have usually been regarded to be approximately valid. However, such an assumption becomes questionable in the large data sets currently investigated in geophysics, hydrology or financial econometrics. There has been a long lasting controversy to decide whether the deviations to "short memory stationarity" should be attributed to long-range dependence or are related to the presence of breakpoints in the mean, the variance, the covariance function or other types of more sophisticated structural changes. The links between non-stationarity and long-range dependence (LRD) have been pointed out by many authors in the hydrology literature long ago: Klemes [1974] and Boes and Salas [1978] show that non-stationarity in the mean provides a possible explanations of the so-called Hurst phenomenon. Potter [1976] and later Rao and Yu [1986] suggested that more sophisticated changes may occur, and have proposed a method to detect such changes. The possible confusions between long-memory and some forms of nonstationarity have been discussed in the applied probability literature: Bhattacharya et al. [1983] show that long-range dependence may be confused with the presence of a small monotonic trend. This phenomenon has also been discussed in the econometrics literature. Hidalgo and Robinson [1996] proposed a test of presence of structural change in a long memory environment. Granger and Hyung [1999] showed that linear processes with breaks can mimic the autocovariance

structure of a linear fractionally integrated long-memory process (a stationary process that encounters occasional regime switches will have some properties that are similar to those of a long-memory process). Similar behaviors are considered in Diebold and Inoue [2001] who provided simple and intuitive econometric models showing that long-memory and structural changes are easily confused. Mikosch and Stărică [2004] asserted that what had been seen by many authors as long memory in the volatility of the absolute values or the square of the log-returns might, in fact, be explained by abrupt changes in the parameters of an underlying GARCH-type models. Berkes et al. [2006] proposed a testing procedure for distinguishing between a weakly dependent time series with change-points in the mean and a long-range dependent time series. Hurvich et al. [2005] have proposed a test procedure for detecting long memory in presence of deterministic trends.

The procedure described in this chapter deals with the problem of detecting changes which may occur in the spectral content of a process. We will consider a process X which, before and after the change, is not necessary stationary but whose difference of at least a given order is stationary (see Definition 2.1.8), so that polynomial trends up to that order can be discarded. This class of process include both short-range dependent and long-range dependent processes, but also unit-root and fractional unit-root processes. The main goal of this chapter is to develop a testing procedure for distinguishing between a K -th order stationary process and a non-stationary process. More precisely, given n observations X_1, \dots, X_n of a time series, we consider the following test hypotheses:

$$\begin{cases} \mathcal{H}_0 : X_1, \dots, X_n \text{ admits a generalized spectral density } f \\ \mathcal{H}_1 : X_1, \dots, X_k \text{ and } X_{k+1}, \dots, X_n \text{ are samples admitting} \\ \quad \text{two different generalized spectral densities } f^{(1)} \text{ and } f^{(2)}, \end{cases} \quad (3.1)$$

where k is an unknown change point.

In this paper, we consider the so-called *a posteriori* or *retrospective* method (see [Brodsky and Darkhovsky, 2000, Chapter 3]). The proposed test is formulated in the wavelet domain, where a change in the generalized spectral density results in a change in the variance of the wavelet coefficients. Our test is based on a CUSUM statistic, which is perhaps the most extensively used statistic for detecting and estimating change-points in mean. In our procedure, the CUSUM is applied to the partial sums of the squared wavelet coefficients at a given scale or on a specific range of scales. This procedure extends the test introduced in Inclan and Tiao [1994] to detect changes in the variance of an independent sequence of random variables. To describe the idea, suppose that, under the null hypothesis, the time series is K -th order difference stationary and that, under the alternative, there is one breakpoint where the generalized spectral density of the process changes. We consider the scalogram in the range of scale $J_1, J_1 + 1, \dots, J_2$. Under the null hypothesis, there is no change in the variance of the wavelet coefficients at any given scale

$j \in \{J_1, \dots, J_2\}$. Under the alternative, these variances takes different values before and after the change point. The amplitude of the change depends on the scale, and the change of the generalized spectral density. We consider the $(J_2 - J_1 + 1)$ -dimensional W2-CUSUM statistic $\{T_{J_1, J_2}(t), t \in [0, 1]\}$ defined by (4.32), which is a CUSUM-like statistics applied to the square of the wavelet coefficients. Using $T_{J_1, J_2}(t)$ we can construct an estimator $\hat{\tau}_{J_1, J_2}$ of the change point (no matter if a change-point exists or not), by minimizing an appropriate norm of the W2-CUSUM statistics, $\hat{\tau}_{J_1, J_2} = \text{Argmin}_{t \in [0, 1]} \|T_{J_1, J_2}(t)\|_*$. The statistic $T_{J_1, J_2}(\hat{\tau}_{J_1, J_2})$ converges to a well-know distribution under the null hypothesis (see Theorems 3.2.1 and 3.2.2) but diverges to infinity under the alternative (Theorems 3.4.1 and 3.4.2). A similar idea has been proposed by Whitcher et al. [2002] but these authors did not take into account the dependence of wavelet coefficient, resulting in an erroneous normalization and asymptotic distributions.

The chapter is organized as follows. In Section 3.2, we introduce the relationship between the generalized spectral density and the variance of wavelet coefficients at a given scale (sub-section 3.2.1). Then, our main assumptions are formulated and the asymptotic distribution of the W2-CUSUM statistics is presented first in the single scale (sub-section 3.2.2) and then in the multiple scales (sub-section 3.2.3) cases. In Section 3.3, several possible test procedures are described to detect the presence of changes at a single scale or simultaneously at several scales. In Section 3.4, the pointwise consistency of the test in the presence of a single jump in the generalized spectral density of the process is presented first in the single case (sub-section 3.4.1) and then in the multi-scale case (sub-section 3.4.2). Section 3.5 is consecrate to the study of finite sample performance of the test procedure based on Monte-Carlo experiments.

3.2 Asymptotic distribution of the W2-CUSUM statistic.

We consider $n_1 = 2048$ observations of an $AR(1)$ with parameter $\phi_1 = 0.8$ and $n_2 = 2048$ observations of an $AR(1)$ with $\phi_2 = 0.5$ and the same variance set to 1 for the two processes. That means that the innovation variance $\sigma_1^2 = 0.36$ for the first process and $\sigma_2^2 = 0.75$ for the second. When we compute the wavelet coefficients of the process which resulting to the concatenation of the two preview processes, we obtain a change exactly around the joint point of the two processes as showing Figure (3.1). That is the reason for that at the introduction, we have told that a change in the generalized spectral density of the process X results in a change in the variance of wavelet coefficients among scales. Then, instead of interest us in the change in the generalized spectral density of the process, we focus on the change in the variance structure of wavelet coefficients that we are going now to present.

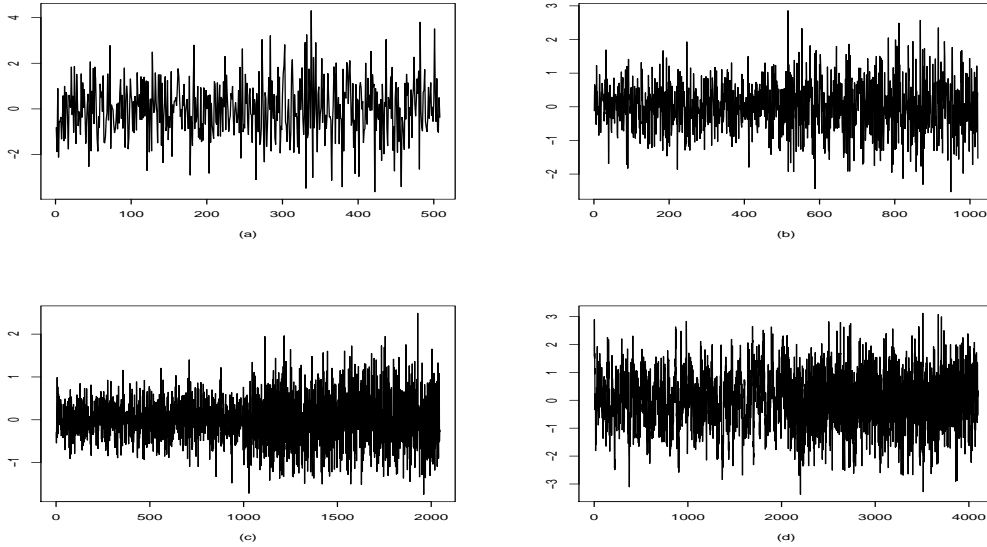


Figure 3.1: (d) corresponds to the original signal obtained by the concatenation of two AR(1) with $\phi = 0.5$ and $\phi = 0.8$ respectively. Figure (a), (b) (c) represent respectively the wavelet coefficients at scale $j = 3, 2, 1$ of the signal (d).

3.2.1 The wavelet spectrum and the scalogram.

Using the wavelet setting introduced in Sub-section 2.2.1, the DWT defined in Sub-section 2.2.2, the wavelet spectrum and the scalogram defined respectively in (2.67) and (2.68) we explain how spectral change-points can be observed in the wavelet domain. The main advantage of using the wavelet domain is to alleviate problems arising when the time series exhibit is long range dependent. Let $X = \{X_t, t \in \mathbb{Z}\}$ be a real-valued process with wavelet coefficients $\{W_{j,k}, k \in \mathbb{Z}\}$ and define

$$\sigma_{j,k}^2 = \text{Var}(W_{j,k}) .$$

If $\Delta^M X$ is stationary, then by Eq (2.43), we have that, for all j , the process of its wavelet coefficients at scale j , $\{W_{j,k}, k \in \mathbb{Z}\}$, is also stationary. Then, the wavelet variance $\sigma_{j,k}^2$ does not depend on k , $\sigma_{j,k}^2 = \sigma_j^2$. It can then be expressed using (2.67).

3.2.2 The single-scale case

To start with simple presentation and statement of results, we first focus in this section on a test procedure aimed at detecting a change in the variance of the wavelet coefficients at a single scale j . Let X_1, \dots, X_n be the n observations of a time series, and denote by $W_{j,k}$ for $(j, k) \in \mathcal{I}_n$ with \mathcal{I}_n defined in (2.32) the associated wavelet coefficients. In view of (2.67), if X_1, \dots, X_n are a n successive observations of a K -th order difference stationary process, then the wavelet variance at each given scale j should be constant.

If the process X is not K -th order stationary, then it can be expected that the wavelet variance will change either gradually or abruptly (if there is a shock in the original time-series). This thus suggests to investigate the consistency of the variance of the wavelet coefficients.

There are many works aimed at detecting the change point in the variance of a sequence of independent random variables; such problem has also been considered, but much less frequently, for sequences of dependent variables. Here, under the null assumption of K -th order difference stationarity, the wavelet coefficients $\{W_{j,k}, k \in \mathbb{Z}\}$ is a covariance stationary sequence whose spectral density is given by (2.66). We will adapt the approach developed in Inclan and Tiao [1994], which uses cumulative sum (CUSUM) of squares to detect change points in the variance.

In order to define the test statistic, we first introduce a change point estimator for the mean of the square of the wavelet coefficients at each scale j .

$$\hat{k}_j = \operatorname{argmax}_{1 \leq k \leq n_j} \left| \sum_{1 \leq i \leq k} W_{j,i}^2 - \frac{k}{n_j} \sum_{1 \leq i \leq n_j} W_{j,i}^2 \right|. \quad (3.2)$$

Using this change point estimator, the W2-CUSUM statistics is defined as

$$T_{n_j} = \frac{1}{n_j^{1/2} s_{j,n_j}} \left| \sum_{1 \leq i \leq \hat{k}_j} W_{j,i}^2 - \frac{\hat{k}_j}{n_j} \sum_{1 \leq i \leq n_j} W_{j,i}^2 \right|, \quad (3.3)$$

where s_{j,n_j}^2 is a suitable estimator of the variance of the sample mean of the $W_{j,i}^2$. Because wavelet coefficients at a given scale are correlated, we use the Bartlett estimator of the variance, which is defined by

$$s_{j,n_j}^2 = \hat{\gamma}_j(0) + 2 \sum_{1 \leq l \leq q(n_j)} w_l(q(n_j)) \hat{\gamma}_j(l), \quad (3.4)$$

where

$$\hat{\gamma}_j(l) \stackrel{\text{def}}{=} \frac{1}{n_j} \sum_{1 \leq i \leq n_j-l} (W_{j,i}^2 - \hat{\sigma}_j^2)(W_{j,i+l}^2 - \hat{\sigma}_j^2), \quad (3.5)$$

are the sample autocovariance of $\{W_{j,i}^2, i = 1, \dots, n_j\}$, $\hat{\sigma}_j^2$ is the scalogram and, for a given integer q ,

$$w_l(q) = 1 - \frac{l}{1+q}, l \in \{0, \dots, q\} \quad (3.6)$$

are the so-called Bartlett weights.

The test differs from statistics proposed in Inclan and Tiao [1994] only in its denominator, which is the square root of a consistent estimator of the partial sum's variance. If $\{X_n\}$ is short-range dependent, the variance of the partial sum of the scalograms is not simply the sum of the variances of the individual square wavelet coefficient, but also

includes the autocovariances of these terms. Therefore, the estimator of the averaged scalogram variance involves not only sums of squared deviations of the scalogram coefficients, but also its weighted autocovariances up to lag $q(n_j)$. The weights $\{w_l(q(n_j))\}$ are those suggested by Newey and West [1987] and always yield a positive sequence of autocovariance, and a positive estimator of the (unnormalized) wavelet spectrum at scale j , at frequency zero using a Bartlett window. We will first established the consistency of the estimator s_{j,n_j}^2 of the variance of the scalogram at scale j and the convergence of the empirical process of the square wavelet coefficients to the Brownian motion. Denote by $D([0, 1])$ is the Skorokhod space of functions which are right continuous at each point of $[0, 1]$ with left limit of $(0, 1]$ (or *cadlag* functions). This space is, in the sequel, equipped with the classical Skorokhod metric.

Theorem 3.2.1. *Suppose that X is a Gaussian process with generalized spectral density f . Let (ϕ, ψ) be a scaling and a wavelet function satisfying (W-1)-(W-4). Let $\{q(n_j)\}$ be a non decreasing sequence of integers satisfying*

$$q(n_j) \rightarrow \infty \quad \text{and} \quad q(n_j)/n_j \rightarrow 0 \quad \text{as} \quad n_j \rightarrow \infty. \quad (3.7)$$

Assume that $\Delta^M X$ is non-deterministic and centered, and that $\lambda^{2M} f(\lambda)$ is two times differentiable in λ with bounded second order derivative. Then for any fixed scale j , as $n \rightarrow \infty$,

$$s_{j,n_j}^2 \xrightarrow{P} \frac{1}{\pi} \int_{-\pi}^{\pi} |\mathbf{D}_{j,0}(\lambda; f)|^2 d\lambda, \quad (3.8)$$

where $\mathbf{D}_{j,0}(\lambda; f)$ is the wavelet coefficients spectral density at scale j see (2.66). Moreover, defining σ_j^2 by (2.67),

$$\frac{1}{n_j^{1/2} s_{j,n_j}} \sum_{i=1}^{\lfloor n_j t \rfloor} (W_{j,i}^2 - \sigma_j^2) \xrightarrow{d} B(t) \quad \text{in} \quad D([0, 1]), \quad \text{as} \quad n \rightarrow \infty \quad (3.9)$$

where $(B(t), t \in [0, 1])$ is the standard Brownian motion.

Remark 3.2.1. *The fact that X is Gaussian can be replaced by the more general assumption that the process X is linear in the strong sense, under appropriate moment conditions on the innovation. The proofs are then more involved, especially to establish the invariance principle which is pivotal in our derivation.*

Remark 3.2.2. *By allowing $q(n_j)$ to increase but at a slower rate than the number of observations, the estimator of the averaged scalogram variance adjusts appropriately for general forms of short-range dependence among the scalogram coefficients. Of course, although the condition (3.7) ensure the consistency of s_{j,n_j}^2 , they provide little guidance in selecting a truncation lag $q(n_j)$. When $q(n_j)$ becomes large relative to the sample size n_j , the finite-sample distribution of the test statistic might be far from its asymptotic limit.*

However $q(n_j)$ cannot be chosen too small since the autocovariances beyond lag $q(n_j)$ may be significant and should be included in the weighted sum. Therefore, the truncation lag must be chosen ideally using some data-driven procedures. Andrews [1991] and Newey and West [1994] provide a data-dependent rule for choosing $q(n_j)$. These contributions suggest that selection of bandwidth according to an asymptotically optimal procedure tends to lead to more accurately sized test statistics than do traditional procedure. The methods suggested by Andrews [1991] for selecting the bandwidth optimally is a plug-in approach. This procedure require the researcher to fit an ARMA model of given order to provide a rough estimator of the spectral density and of its derivatives at zero frequencies (although misspecification of the order affects only optimality but not consistency). The minimax optimality of this method is based on an asymptotic mean-squared error criterion and its behavior in the finite sample case is not precisely known. The procedure outlined in Newey and West [1994] suggests to bypass the modeling step, by using instead a pilot truncated kernel estimates of the spectral density and its derivative. We use these data driven procedures in the Monte Carlo experiments (these procedures have been implemented in the R-package sandwich).

Proof of Theorem 3.2.1. Since X is Gaussian and $\mathbf{\Delta}^M X$ is centered, Eq. (2.43) implies that $\{W_{j,k}, k \in \mathbb{Z}\}$ is a centered Gaussian process, whose distribution is determined by

$$\gamma_j(h) = \text{Cov}(W_{j,0}, W_{j,h}) = \int_{-\pi}^{\pi} \mathbf{D}_{j,0}(\lambda; f) e^{-i\lambda h} d\lambda.$$

From (2.66) and (2.42), we have

$$\mathbf{D}_{j,0}(\lambda; f) = \sum_{l=0}^{2^j-1} f(2^{-j}(\lambda + 2l\pi)) 2^{-j} \left| \tilde{H}_j(2^{-j}(\lambda + 2l\pi)) \right|^2 \left| 1 - e^{-i2^{-j}(\lambda + 2l\pi)} \right|^{2M},$$

where \tilde{H}_j is a trigonometric polynomial. Using that $|1 - e^{-i\xi}|^{2M} = |\xi|^{2M} \left| \frac{1 - e^{-i\xi}}{i\xi} \right|^{2M}$ and that $|\xi|^{2M} f(\xi)$ has a bounded second order derivative, we get that $\mathbf{D}_{j,0}(\lambda; f)$ has also a bounded second order derivative. In particular,

$$\int_{-\pi}^{\pi} |\mathbf{D}_{j,0}(\lambda; f)|^2 d\lambda < \infty \quad \text{and} \quad \sum_{s \in \mathbb{Z}} |\gamma_j(s)| < \infty. \quad (3.10)$$

The proof may be decomposed into 3 steps. We first prove the consistency of the Bartlett estimator of the variance of the squares of wavelet coefficients s_{j,n_j}^2 , that is (3.8). Then we determine the asymptotic normality of the finite-dimensional distributions of the empirical scalogram, suitably centered and normalized. Finally a tightness criterion is proved, to establish the convergence in the Skorokhod space. Combining these three steps completes the proof of (3.9).

Step 1. Observe that, by the Gaussian property, $\text{Cov}(W_{j,0}^2, W_{j,h}^2) = 2\gamma_j^2(h)$. Using Theorem 3-i in Giraitis et al. [2003], the limit (3.8) follows from

$$2 \sum_{h=-\infty}^{+\infty} \gamma_j^2(h) = \frac{1}{\pi} \int_{-\pi}^{\pi} |\mathbf{D}_{j,0}(\lambda; f)|^2 d\lambda < \infty, \quad (3.11)$$

and

$$\sup_{h \in \mathbb{Z}} \sum_{r,s=-\infty}^{+\infty} |\mathcal{K}(h, r, s)| < \infty. \quad (3.12)$$

where

$$\mathcal{K}(h, r, s) = \text{Cum}(W_{j,k}^2, W_{j,k+h}^2, W_{j,k+r}^2, W_{j,k+s}^2). \quad (3.13)$$

Equation (3.11) follows from Parseval's equality and (3.10). Let us now prove (3.12). Using that the wavelet coefficients are Gaussian, we obtain

$$\begin{aligned} \mathcal{K}(h, r, s) = & 12 \{ \gamma_j(h)\gamma_j(r-s)\gamma_j(h-r)\gamma_j(s) \\ & + \gamma_j(h)\gamma_j(r-s)\gamma_j(h-s)\gamma_j(r) + \gamma_j(s-h)\gamma_j(r-h)\gamma_j(r)\gamma_j(s) \}. \end{aligned}$$

The bound of the last term is given by

$$\sup_{h \in \mathbb{Z}} \sum_{r,s=-\infty}^{+\infty} |\gamma_j(s-h)\gamma_j(r-h)\gamma_j(r)\gamma_j(s)| \leq \sup_h \left(\sum_{r=-\infty}^{+\infty} |\gamma_j(r)\gamma_j(r-h)| \right)^2$$

which is finite by the Cauchy-Schwarz inequality, since $\sum_{r \in \mathbb{Z}} \gamma_j^2(r) < \infty$.

Using $|\gamma_j(h)| < \gamma_j(0)$ and the Cauchy-Schwarz inequality, we have

$$\sup_{h \in \mathbb{Z}} \sum_{r,s=-\infty}^{+\infty} |\gamma_j(h)\gamma_j(r-s)\gamma_j(h-r)\gamma_j(s)| \leq \gamma_j(0) \sum_{u \in \mathbb{Z}} \gamma_j^2(u) \sum_{s \in \mathbb{Z}} |\gamma_j(s)|,$$

and the same bound applies to

$$\sup_{h \in \mathbb{Z}} \sum_{r,s=-\infty}^{+\infty} |\gamma_j(h)\gamma_j(r-s)\gamma_j(h-s)\gamma_j(r)|.$$

Hence, we have (3.12) by (3.10), which achieves the proof of Step 1.

Step 2. Let us define

$$S_{n_j}(t) = \frac{1}{\sqrt{n_j}} \sum_{i=1}^{\lfloor n_j t \rfloor} (W_{j,i}^2 - \sigma_j^2), \quad (3.14)$$

where $\sigma_j^2 = \mathbb{E}(W_{j,i}^2)$, and $\lfloor x \rfloor$ is the entire part of x . Step 2 consists in proving that for $0 \leq t_1 \leq \dots \leq t_k \leq 1$, and $\mu_1, \dots, \mu_k \in \mathbb{R}$,

$$\sum_{i=1}^k \mu_i S_{n_j}(t_i) \xrightarrow{d} \mathcal{N} \left(0, \frac{1}{\pi} \int_{-\pi}^{\pi} |\mathbf{D}_{j,0}(\lambda; f)|^2 d\lambda \times \text{Var} \left(\sum_{i=1}^k \mu_i B(t_i) \right) \right). \quad (3.15)$$

Observe that

$$\begin{aligned}
\sum_{i=1}^k \mu_i S_{n_j}(t_i) &= \frac{1}{\sqrt{n_j}} \sum_{i=1}^k \mu_i \sum_{l=1}^{n_j} (W_{j,l}^2 - \sigma_j^2) \mathbb{1}_{\{l \leq \lfloor n_j t_i \rfloor\}} \\
&= \sum_{l=1}^{n_j} W_{j,l}^2 a_{l,n} - E \left(\sum_{l=1}^{n_j} W_{j,l}^2 a_{l,n} \right) \\
&= \xi_{n_j}^T A_{n_j} \xi_{n_j},
\end{aligned}$$

where we set $a_{l,n} = \frac{1}{\sqrt{n_j}} \sum_{i=1}^k \mu_i \mathbb{1}_{\{l \leq \lfloor n_j t_i \rfloor\}}$, $\xi_{n_j} = (W_{j,1}, \dots, W_{j,n_j})^T$ and A_{n_j} is the diagonal matrix with diagonal entries $(a_{1,n_j}, \dots, a_{n_j,n_j})$. Applying [Moulines et al., 2008, Lemma 12], (3.15) is obtained by proving that, as $n_j \rightarrow \infty$,

$$\rho(A_{n_j}) \rho(\Gamma_{n_j}) \rightarrow 0 \quad (3.16)$$

$$\text{Var} \left(\sum_{i=1}^k \mu_i S_{n_j}(t_i) \right) \rightarrow \frac{1}{\pi} \int_{-\pi}^{\pi} |\mathbf{D}_{j,0}(\lambda; f)|^2 d\lambda \times \text{Var} \left(\sum_{i=1}^k \mu_i (B(t_i)) \right), \quad (3.17)$$

where $\rho(A)$ denote the spectral radius of the matrix A , that is, the maximum modulus of its eigenvalues and Γ_{n_j} is the covariance matrix of ξ_{n_j} . The process $(W_{j,i})_{\{i=1, \dots, n_j\}}$ is stationary with spectral density $\mathbf{D}_{j,0}(\cdot; f)$. Thus, by Lemma 2 in Moulines et al. [2007b] its covariance matrix Γ_{n_j} satisfies $\rho(\Gamma_{n_j}) \leq 2\pi \sup_{\lambda} \mathbf{D}_{j,0}(\lambda; f)$. Furthermore, as $n_j \rightarrow \infty$,

$$\rho(A_{n_j}) = \max_{1 \leq l \leq n_j} \frac{1}{\sqrt{n_j}} \left| \sum_{i=1}^k \mu_i \mathbb{1}_{\{l \leq \lfloor n_j t_i \rfloor\}} \right| \leq n_j^{-1/2} \sum_{i=1}^k |\mu_i| \rightarrow 0,$$

and (3.16) holds. We now prove (3.17). Using that $B(t)$ has variance t and independent and stationary increments, and that these properties characterize its covariance function, it is sufficient to show that, for all $t \in [0, 1]$, as $n_j \rightarrow \infty$,

$$\text{Var}(S_{n_j}(t)) \rightarrow t \int_{-\pi}^{\pi} |\mathbf{D}_{j,0}(\lambda; f)|^2 d\lambda, \quad (3.18)$$

and for all $0 \leq r \leq s \leq t \leq 1$, as $n_j \rightarrow \infty$,

$$\text{Cov}(S_{n_j}(t) - S_{n_j}(s), S_{n_j}(r)) \rightarrow 0. \quad (3.19)$$

For any sets $A, B \subseteq [0, 1]$, we set

$$V_{n_j}(\tau, A, B) = \frac{1}{n_j} \sum_{k \geq 1} \mathbb{1}_A((k + \tau)/n_j) \mathbb{1}_B(k/n_j).$$

For all $0 \leq s, t \leq 1$, we have

$$\begin{aligned}
\text{Cov}(S_{n_j}(t), S_{n_j}(s)) &= \frac{1}{n_j} \sum_{i=1}^{\lfloor n_j t \rfloor} \sum_{k=1}^{\lfloor n_j s \rfloor} \text{Cov}(W_{j,i}^2, W_{j,k}^2) \\
&= 2 \sum_{\tau \in \mathbb{Z}} \gamma_j^2(\tau) V_{n_j}(\tau,]0, t],]0, s]).
\end{aligned}$$

The previous display applies to the left-hand side of (3.18) when $s = t$ and for $0 \leq r \leq s \leq t \leq 1$, it yields

$$\text{Cov}(S_{n_j}(t) - S_{n_j}(s), S_{n_j}(r)) = 2 \sum_{\tau \in \mathbb{Z}} \gamma_j^2(\tau) V_{n_j}(\tau,]s, t],]0, r]) .$$

Observe that for all $A, B \subseteq [0, 1]$, $\sup_{\tau} |V_n(j, \tau, A, B)| \leq \frac{k}{n_j} \leq 1$. Hence, by dominated convergence, the limits in (3.18) and (3.19) are obtained by computing the limits of $V_n(j, \tau,]0, t],]0, t])$ and $V_n(j, \tau,]s, t],]0, r])$ respectively. We have for any $\tau \in \mathbb{Z}$, $t > 0$, and n_j large enough,

$$\sum_{k \geq 1} \mathbb{1}_{\{\frac{k+\tau}{n_j} \in]0, t]\}} \mathbb{1}_{\{\frac{k}{n_j} \in]0, t]\}} = \{(n_j t \wedge n_j t - \tau)\}_+ = n_j t - \tau_+ .$$

Hence, as $n_j \rightarrow \infty$, $V_{n_j}(\tau,]0, t],]0, t]) \rightarrow t$ and, by (3.11), (3.18) follows. We have for any $\tau \in \mathbb{Z}$ and $0 < r \leq s \leq t$,

$$\begin{aligned} \sum_{k \geq 1} \mathbb{1}_{\{\frac{k+\tau}{n_j} \in]s, t]\}} \mathbb{1}_{\{\frac{k}{n_j} \in]0, r]\}} &= \{(n_j r \wedge \{n_j t - \tau\}) - (0 \vee \{n_j s - \tau\})\}_+ \\ &= (n_j r - n_j s + \tau)_+ \rightarrow \mathbb{1}_{\{r=s\}} \tau_+ , \end{aligned}$$

where the last equality holds for n_j large enough and the limit as $n_j \rightarrow \infty$. Hence $V_{n_j}(\tau,]s, t],]0, r]) \rightarrow 0$ and (3.19) follows, which achieves Step 2.

Step 3. We now prove the tightness of $\{S_{n_j}(t), t \in [0, 1]\}$ in the Skorokhod metric space. By Theorem 13.5 in Billingsley [1999], it is sufficient to prove that for all $0 \leq r \leq s \leq t$,

$$\mathbb{E}[|S_{n_j}(s) - S_{n_j}(r)|^2 |S_{n_j}(t) - S_{n_j}(s)|^2] \leq C|t - r|^2 ,$$

where $C > 0$ is some constant independent of r, s, t and n_j . We shall prove that, for all $0 \leq r \leq t$,

$$\mathbb{E}[|S_{n_j}(t) - S_{n_j}(r)|^4] \leq C_1 \{n_j^{-1}(\lfloor n_j t \rfloor - \lfloor n_j r \rfloor)\}^2 . \quad (3.20)$$

By the Cauchy-Schwarz inequality, and using that, for $0 \leq r \leq s \leq t$,

$$n_j^{-1}(\lfloor n_j t \rfloor - \lfloor n_j s \rfloor) \times n_j^{-1}(\lfloor n_j s \rfloor - \lfloor n_j r \rfloor) \leq 4(t - r)^2 ,$$

the criterion (3.20) implies the previous criterion. Hence the tightness follows from (3.20), that we now prove. We have, for any $\mathbf{i} = (i_1, \dots, i_4)$,

$$\begin{aligned} \mathbb{E} \left[\prod_{k=1}^4 (W_{j, i_k}^2 - \sigma_j^2) \right] &= \text{Cum}(W_{j, i_1}^2, \dots, W_{j, i_4}^2) + \mathbb{E}(W_{j, i_1}^2, W_{j, i_2}^2) \mathbb{E}[W_{j, i_3}^2, W_{j, i_4}^2] \\ &\quad + \mathbb{E}[W_{j, i_1}^2, W_{j, i_3}^2] \mathbb{E}[W_{j, i_2}^2, W_{j, i_4}^2] + \mathbb{E}[W_{j, i_1}^2, W_{j, i_4}^2] \mathbb{E}[W_{j, i_2}^2, W_{j, i_3}^2] . \end{aligned}$$

It follows that, denoting for $0 \leq r \leq t \leq 1$,

$$\mathbb{E} \left[|S_{n_j}(t) - S_{n_j}(r)|^4 \right] = \frac{1}{n_j^2} \sum_{\mathbf{i} \in A_{r,t}^4} \text{Cum}(W_{j,i_1}^2, \dots, W_{j,i_4}^2) + \frac{3}{n_j^2} \left(\sum_{\mathbf{i} \in A_{r,t}^2} \mathbb{E}[W_{j,i_1} W_{j,i_2}] \right)^2$$

where $A_{r,t} = \{\lfloor n_j r \rfloor + 1, \dots, \lfloor n_j t \rfloor\}$. Observe that

$$0 \leq \frac{1}{n_j} \sum_{\mathbf{i} \in A_{r,t}^2} \mathbb{E}[W_{j,i_1}^2 W_{j,i_2}^2] \leq 2 \sum_{\tau \in \mathbb{Z}} \gamma_j^2(\tau) \times n_j^{-1} (\lfloor n_j t \rfloor - \lfloor n_j r \rfloor).$$

Using that, by (3.13), $\text{Cum}(W_{j,i_1}^2, \dots, W_{j,i_4}^2) = \mathcal{K}(i_2 - i_1, i_3 - i_1, i_4 - i_1)$, we have

$$\begin{aligned} \sum_{\mathbf{i} \in A_{r,t}^4} |\text{Cum}(W_{j,i_1}^2, \dots, W_{j,i_4}^2)| &\leq (\lfloor n_j t \rfloor - \lfloor n_j r \rfloor) \sum_{h,s,l=\lfloor n_j r \rfloor - \lfloor n_j t \rfloor + 1}^{\lfloor n_j t \rfloor - \lfloor n_j r \rfloor - 1} |\mathcal{K}(h, s, l)| \\ &\leq 2(\lfloor n_j t \rfloor - \lfloor n_j r \rfloor)^2 \sup_{h \in \mathbb{Z}} \sum_{r,s=-\infty}^{+\infty} |\mathcal{K}(h, r, s)|. \end{aligned}$$

The last three displays and (3.12) imply (3.20), which proves the tightness.

Finally, observing that the variance (3.11) is positive, unless f vanishes almost everywhere, the convergence (3.9) follows from Slutsky's lemma and the three previous steps. \square

3.2.3 The multiple-scale case

The results above can be extended to test simultaneously changes in wavelet variances occurring simultaneously at multiple time-scales. To construct a multiple scale test, consider the *between-scale* process

$$\{[W_{j,k}, \mathbf{W}_{j,k}(j - j')^T]^T\}_{k \in \mathbb{Z}}, \quad (3.21)$$

defined in (2.63) where $\mathbf{W}_{j,k}(j - j')^T$ is defined in (2.64). It is shown in [Moulines et al., 2007b, Corollary 1] that, when $\Delta^M X$ is covariance stationary, the between scale process $\{[W_{j,k}, \mathbf{W}_{j,k}(j - j')^T]^T\}_{k \in \mathbb{Z}}$ is also covariance stationary. Moreover, for all $0 \leq u \leq j$, the *between scale covariance matrix* is defined as

$$\text{Cov}(W_{j,0}^X, \mathbf{W}_{j,k}(u)) = \int_{-\pi}^{\pi} e^{i\lambda k} \mathbf{D}_{j,u}(\lambda; f) d\lambda, \quad (3.22)$$

where $\mathbf{D}_{j,u}(\lambda; f)$ is the cross-spectral density function of the between-scale process given by (2.65). The case $u = 0$ corresponds to the spectral density of the *within-scale* process $\{W_{j,k}\}_{k \in \mathbb{Z}}$ given in (2.66). Under the null hypothesis that X is K -th order stationary, a *multiple scale* procedure aims at testing that the scalogram in a range satisfies

$$\mathcal{H}_0 : \sigma_{j,1}^2 = \dots = \sigma_{j,n_j}^2, \text{ for all } j \in \{J_1, J_1 + 1, \dots, J_2\} \quad (3.23)$$

where J_1 and J_2 are the *finest* and the *coarsest* scales included in the procedure, respectively. The wavelet coefficients at different scales are not uncorrelated so that both the *within-scale* and the *between scale* covariances need to be taken into account.

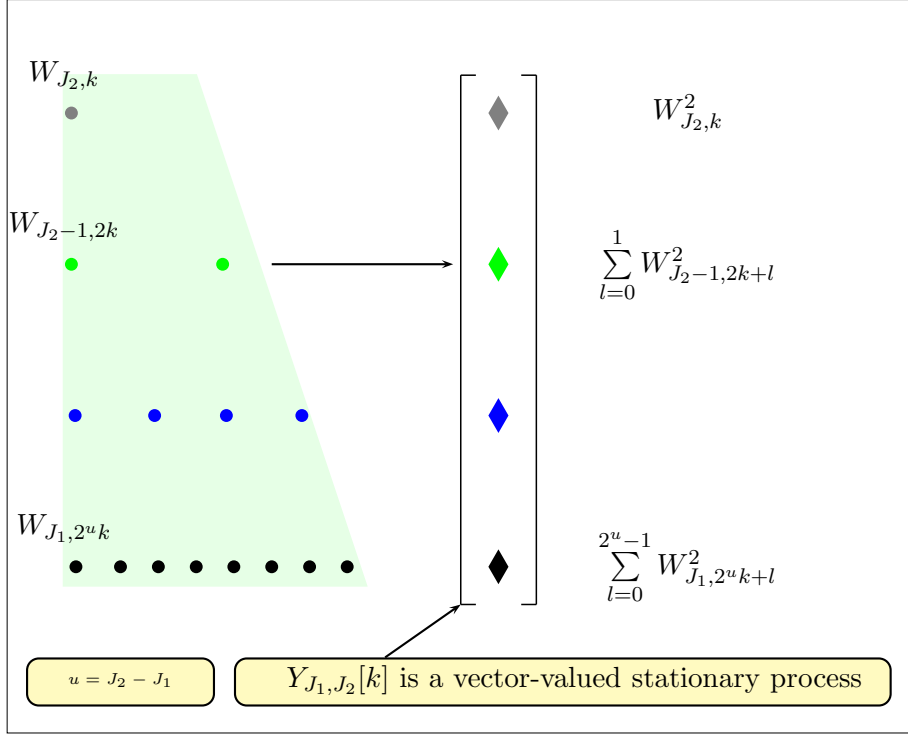


Figure 3.2: Between scale stationary process.

As before, we use a CUSUM statistic in the wavelet domain. However, we now use multiple scale vector statistics. Consider the following process

$$Y_{J_1, J_2}[i] = \left(W_{J_2, i}^2, \sum_{u=0}^1 W_{J_2-1, 2i+u}^2, \dots, \sum_{u=0}^{2^{(J_2-J_1)}} W_{J_1, 2^{(J_2-J_1)}i+u}^2 \right)^T. \quad (3.24)$$

The Bartlett estimator of the covariance matrix of the square wavelet's coefficients for scales $\{J_1, \dots, J_2\}$ is the $(J_2 - J_1 + 1) \times (J_2 - J_1 + 1)$ symmetric definite positive matrix $\hat{\Gamma}_{J_1, J_2}^B$ given by :

$$\hat{\Gamma}_{J_1, J_2}^B = \sum_{\tau=-q(n_{J_2})}^{q(n_{J_2})} w_\tau [q(n_{J_2})] \hat{\Sigma}_{J_1, J_2}^B(\tau), \quad \text{where} \quad (3.25)$$

$$\hat{\Sigma}_{J_1, J_2}^B(\tau) = \frac{1}{n_{J_2}} \sum_{i, i+\tau=1}^{n_{J_2}} (Y_{J_1, J_2}[i] - \bar{Y}_{J_1, J_2}) (Y_{J_1, J_2}[i + \tau] - \bar{Y}_{J_1, J_2})^T, \quad (3.26)$$

$$\text{with } \bar{Y}_{J_1, J_2} = \frac{1}{n_{J_2}} \sum_{i=1}^{n_{J_2}} Y_{J_1, J_2}[i]$$

Finally, let us define the vector of partial sum from scale J_1 to scale J_2 as

$$S_{J_1, J_2}(t) = \frac{1}{\sqrt{n_{J_2}}} \left[\sum_{i=1}^{\lfloor n_j t \rfloor} W_{j,i}^2 \right]_{j=J_1, \dots, J_2}. \quad (3.27)$$

Theorem 3.2.2. *Under the assumptions of Theorem 3.2.1, we have, as $n \rightarrow \infty$,*

$$\widehat{\Gamma}_{J_1, J_2}^B = \Gamma_{J_1, J_2} + O_P\left(\frac{q(n_{J_2})}{n_{J_2}}\right) + O_P(q^{-1}(n_{J_2})), \quad (3.28)$$

where

$$\Gamma_{J_1, J_2}(j' - J_1 + 1, j - J_1 + 1) = \sum_{\tau \in \mathbb{Z}} \text{Cov}\left(Y_{J_1, J_2}^{j'}[\tau], Y_{J_1, J_2}^j[0]\right), \quad \text{with } J_1 \leq j' \leq j \leq J_2 \quad (3.29)$$

and,

$$\left(\widehat{\Gamma}_{J_1, J_2}^B\right)^{-1/2} (S_{J_1, J_2}(t) - \mathbb{E}[S_{J_1, J_2}(t)]) \xrightarrow{d} B(t) = (B_{J_1}(t), \dots, B_{J_2}(t)), \quad (3.30)$$

in $D^{J_2 - J_1 + 1}[0, 1]$, where $\{B_j(t)\}_{j=J_1, \dots, J_2}$ are independent Brownian motions.

Proof of Theorem 3.2.2. The proof has 3 steps as in Theorem 3.2.1. We first prove that the Bartlett estimator $\widehat{\Gamma}_{J_1, J_2}^B$ converges in probability to Γ_{J_1, J_2} with rate given by (3.28). Then, we show the asymptotic normality of the finite-dimensional distribution of the vector $(S_{J_1, J_2}(t))$ and finally, the tightness criterion is proved so that the two last steps yield (4.31).

Step 1. Let j and j' two fixed scales with $J_1 \leq j' \leq j \leq J_2$. For ease of the notation, we denote by $\widehat{\Gamma}^B(j', j)$, $\widehat{\Sigma}_{j', j}^B$ and $\Gamma(j', j)$ the corresponding entries of $\widehat{\Gamma}_{J_1, J_2}^B$, $\widehat{\Sigma}_{J_1, J_2}^B$ and Γ_{J_1, J_2} respectively. We have $\widehat{\Gamma}^B(j', j) = \sum_{\tau=-q(n_{J_2})}^{q(n_{J_2})} w_\tau(q(n_{J_2})) \widehat{\Sigma}_{j', j}^B(\tau)$, where

$$\widehat{\Sigma}_{j', j}^B(\tau) \stackrel{\text{def}}{=} \frac{1}{n_{J_2}} \sum_{i, i+\tau=1}^{n_{J_2}} \left(Y_{J_1, J_2}^{j'}[i] - \bar{Y}_{J_1, J_2}^{j'} \right) \left(Y_{J_1, J_2}^j[i + \tau] - \bar{Y}_{J_1, J_2}^j \right),$$

and $Y_{J_1, J_2}^j[i] = \sum_{u=0}^{2^{J_2-j}-1} W_{j, 2^{J_2-j}i+u}^2$. We decompose the error term in 3 terms.

$$\widehat{\Gamma}^B(j', j) - \Gamma(j', j) = \widehat{\Gamma}^B(j', j) - \widetilde{\Gamma}^B(j', j) + \widetilde{\Gamma}^B(j', j) - \mathbb{E}\left(\widetilde{\Gamma}^B(j', j)\right) + \mathbb{E}\left(\widetilde{\Gamma}^B(j', j)\right) - \Gamma(j', j).$$

where $\widetilde{\Gamma}^B(j', j)$ is defined by $\widetilde{\Gamma}^B(j', j) = \frac{1}{n_{J_2}} \sum_{\tau=-q(n_{J_2})}^{q(n_{J_2})} w_\tau(q(n_{J_2})) \widetilde{\Sigma}_{j', j}^B(\tau)$, with

$$\widetilde{\Sigma}_{j', j}^B(\tau) = \frac{1}{n_{J_2}} \sum_{i, i+\tau=1}^{n_{J_2}} \left(Y_{J_1, J_2}^{j'}[i] - 2^{J_2-j'} \sigma_{j'}^2 \right) \left(Y_{J_1, J_2}^j[i] - 2^{J_2-j} \sigma_j^2 \right).$$

Let $\|\cdot\|_2 = (\mathbb{E}(\cdot)^2)^{1/2}$ and $\|\cdot\|_1 = \mathbb{E}(|\cdot|)$. The bound (3.28) follows from

$$\left\| \widehat{\Gamma}^B(j', j) - \widetilde{\Gamma}^B(j', j) \right\|_1 = O\left(\frac{q(n_{J_2})}{n_{J_2}}\right), \quad \left\| \widetilde{\Gamma}^B(j', j) - \mathbb{E}(\widetilde{\Gamma}^B(j', j)) \right\|_2 = O\left(\frac{q(n_{J_2})}{n_{J_2}}\right) \quad (3.31)$$

and

$$\left| \mathbb{E}(\widetilde{\Gamma}^B(j', j)) - \Gamma(j', j) \right| = O(q^{-1}(n_{J_2})), \quad (3.32)$$

that we now prove. Define $\Sigma_{j', j}(\tau) = \text{Cov}\left(Y_{J_1, J_2}^j[0], Y_{J_1, J_2}^{j'}[\tau]\right)$ then,

$$\begin{aligned} \Sigma_{j', j}(\tau) + \Sigma_{j', j}(-\tau) &= \text{Cov}\left(Y_{J_1, J_2}^j[0] + Y_{J_1, J_2}^{j'}[0], Y_{J_1, J_2}^j[\tau] + Y_{J_1, J_2}^{j'}[\tau]\right) \\ &\quad - \text{Cov}\left(Y_{J_1, J_2}^j[0], Y_{J_1, J_2}^j[\tau]\right) - \text{Cov}\left(Y_{J_1, J_2}^{j'}[0], Y_{J_1, J_2}^{j'}[\tau]\right), \\ \Sigma_{j', j}(0) &= \frac{1}{2} \left\{ \text{Var}\left(Y_{J_1, J_2}^j[0] + Y_{J_1, J_2}^{j'}[0]\right) - \text{Var}\left(Y_{J_1, J_2}^j[0]\right) - \text{Var}\left(Y_{J_1, J_2}^{j'}[0]\right) \right\}, \end{aligned}$$

and thus

$$\begin{aligned} \Gamma(j', j) &= \frac{1}{2} \left\{ \text{Var}\left(Y_{J_1, J_2}^j[0] + Y_{J_1, J_2}^{j'}[0]\right) + 2 \sum_{\tau \geq 1} \text{Cov}\left(Y_{J_1, J_2}^j[0] + Y_{J_1, J_2}^{j'}[0], Y_{J_1, J_2}^j[\tau] + Y_{J_1, J_2}^{j'}[\tau]\right) \right. \\ &\quad \left. - \text{Var}\left(Y_{J_1, J_2}^j[0]\right) - 2 \sum_{\tau \geq 1} \text{Cov}\left(Y_{J_1, J_2}^j[0], Y_{J_1, J_2}^j[\tau]\right) - \text{Var}\left(Y_{J_1, J_2}^{j'}[0]\right) - 2 \sum_{\tau \geq 1} \text{Cov}\left(Y_{J_1, J_2}^{j'}[0], Y_{J_1, J_2}^{j'}[\tau]\right) \right\}. \end{aligned}$$

$\Gamma(j', j)$ is then a linear combination of the covariances of the covariance stationary processes $\{Y_{J_1, J_2}^j[i] + Y_{J_1, J_2}^{j'}[i]\}_{i \in \mathbb{Z}}$, $\{Y_{J_1, J_2}^j[i]\}_{i \in \mathbb{Z}}$ and $\{Y_{J_1, J_2}^{j'}[i]\}_{i \in \mathbb{Z}}$. Using the single case results (3.11) and (3.12), it is sufficient to consider the process $\{Y_{J_1, J_2}^j[i] + Y_{J_1, J_2}^{j'}[i]\}_{i \in \mathbb{Z}}$ in the following. The bounds in (3.31) thus follow from the proof of Theorem 3-i in Giraitis et al. [2003],

$$\sum_{\tau \in \mathbb{Z}} |\Sigma_{j', j}(\tau)| < \infty, \quad (3.33)$$

$$\sup_{h \in \mathbb{Z}} \sum_{r, s \in \mathbb{Z}} \left| \mathcal{K}\left(Y_{J_1, J_2}^j[0] + Y_{J_1, J_2}^{j'}[0], Y_{J_1, J_2}^j[h] + Y_{J_1, J_2}^{j'}[h], Y_{j, r} + Y_{j', r}, Y_{J_1, J_2}^j[s] + Y_{J_1, J_2}^{j'}[s]\right) \right| < \infty. \quad (3.34)$$

Let us prove now (3.33) and (3.34). Using Parseval inequality,

$$\begin{aligned} \sum_{\tau \in \mathbb{Z}} |\Sigma_{j', j}(\tau)| &= \sum_{\tau \in \mathbb{Z}} \sum_{l=0}^{2^{J_2-j'}-1} \sum_{v=0}^{2^{J_2-j}-1} \text{Cov}\left(W_{j, v}^2, W_{j', 2^{J_2-j'}+\tau+l}^2\right) \\ &= 2 \sum_{\tau \in \mathbb{Z}} \sum_{l=0}^{2^{J_2-j'}-1} \sum_{v=0}^{2^{J_2-j}-1} \text{Cov}^2\left(W_{j, v}, W_{j', 2^{J_2-j'}+\tau+l}\right) = 2^{J_2-j+1} \sum_{s \in \mathbb{Z}} \gamma_{j', j}^2(s) \\ &= 2^{J_2-j+1} \sum_{t \in \mathbb{Z}} \sum_{v=0}^{2^{j-j'}-1} \gamma_{j', j}^2(2^{j-j'}t + v) \\ &= 2^{J_2-j+1} \sum_{t \in \mathbb{Z}} \left\| \int_{-\pi}^{\pi} \mathbf{D}_{j, u}(\lambda; f) e^{-i\lambda t} d\lambda \right\|^2 = \frac{2^{J_2-j}}{\pi} \int_{-\pi}^{\pi} \|\mathbf{D}_{j, u}(\lambda; f)\|^2 d\lambda, \end{aligned}$$

where $\mathbf{D}_{j,u}(\lambda; f)$ is the 2^u dimensional vector of spectral density between scale j and j' defined by (2.65) and $\gamma_{j',j}(s) = \text{Cov}(W_{j,0}, W_{j',s})$. Since $|\lambda|^{2M} f(\lambda)$ has a bounded second order derivative, we get that each component of $\mathbf{D}_{j,u}(\lambda; f)$ has also a bounded second order derivative. In particular,

$$\int_{-\pi}^{\pi} \|\mathbf{D}_{j,u}(\lambda; f)\|^2 d\lambda < \infty \quad \text{and} \quad |\gamma_{j',j}(s)| \leq \frac{C}{s^2} \quad \forall s \in \mathbb{Z}, \quad (3.35)$$

where C is a positive constant. The last two displays give (3.33). Let us set $\mathbf{a} = \{a_1, \dots, a_4\}$, with $a_i \leq a_{i+1}$ $\ell = \{l_1, \dots, l_4\}$, $\mathbf{L}_{\mathbf{J}_2, \mathbf{a}_i} = \{0, \dots, 2^{J_2 - a_i} - 1\}$, and $\mathbf{T}_{a_i a_{i'}} = \mathbf{L}_{\mathbf{J}_2, \mathbf{a}_i} \times \mathbf{L}_{\mathbf{J}_2, \mathbf{a}_{i'}}$, $i, i' \in \{1, \dots, 4\}$.

$$\begin{aligned} & \mathcal{K}(Y_{J_1, J_2}^j[0] + Y_{J_1, J_2}^{j'}[0], Y_{J_1, J_2}^j[h] + Y_{J_1, J_2}^{j'}[h], Y_{j,r} + Y_{j',r}, Y_{J_1, J_2}^j[s] + Y_{J_1, J_2}^{j'}[s]) \\ &= \sum_{\mathbf{a} \in \{j, j'\}^4} \mathcal{K}(Y_{a_1, 0}, Y_{a_2, h}, Y_{a_3, r}, Y_{a_4, s}) \\ &= \sum_{\mathbf{a} \in \{j, j'\}^4} \sum_{\ell} \mathcal{K}\left(W_{a_1, l_1}^2, W_{a_2, 2^{J_2 - a_2} h + l_2}^2, W_{a_3, 2^{J_2 - a_3} r + l_3}^2, W_{a_4, 2^{J_2 - a_4} s + l_4}^2\right), \end{aligned}$$

where the second sum in the previous line is performs over $\ell \in \mathbf{L}_{\mathbf{J}_2, \mathbf{a}_1} \times \mathbf{L}_{\mathbf{J}_2, \mathbf{a}_2} \times \mathbf{L}_{\mathbf{J}_2, \mathbf{a}_3} \times \mathbf{L}_{\mathbf{J}_2, \mathbf{a}_4}$. Since the process $\{X_t\}_{t \in \mathbb{Z}}$ is gaussian and the wavelet coefficients $\{W_{j,k}\}_{(j,k) \in \mathcal{I}_n}$ are centered, we have

$$\begin{aligned} & \mathcal{K}\left(W_{a_1, l_1}^2, W_{a_2, 2^{J_2 - a_2} h + l_2}^2, W_{a_3, 2^{J_2 - a_3} r + l_3}^2, W_{a_4, 2^{J_2 - a_4} s + l_4}^2\right) = 12\{\gamma_{a_1, a_2}(h, l_1, l_2)\gamma_{a_3, a_4} \\ & (r-s, l_3, l_4)\gamma_{a_2, a_3}(h-r, l_2, l_3)\gamma_{a_1, a_4}(s, l_1, l_4) + \gamma_{a_1, a_2}(h, l_1, l_2)\gamma_{a_3, a_4}(r-s, l_3, l_4)\gamma_{a_2, a_4}(h-s, l_2, l_4) \\ & \gamma_{a_1, a_3}(r, l_1, l_3) + \gamma_{a_1, a_3}(r, l_1, l_3)\gamma_{a_2, a_3}(h-r, l_2, l_3)\gamma_{a_1, a_4}(s, l_1, l_4)\gamma_{a_2, a_4}(h-s, l_2, l_4)\}, \end{aligned}$$

where, $\gamma_{j',j}(\tau, l, u) = \gamma_{j',j}(2^{J_2 - j'} \tau + l - 2^{j - j'} u)$. To obtain (3.34) we use (3.35), Cauchy-Schwarz Inequality and Parseval inequality.

For all $(l_i, l_{i'}) \in \mathbf{L}_{\mathbf{J}_2, \mathbf{a}_i} \times \mathbf{L}_{\mathbf{J}_2, \mathbf{a}_{i'}}$, we have

$$\begin{aligned}
& \sup_{h \in \mathbb{Z}} \sum_{r, s \in \mathbb{Z}} |\gamma_{a_1, a_3}(r, l_1, l_3) \gamma_{a_2, a_3}(h-r, l_2, l_3) \gamma_{a_1, a_4}(s, l_1, l_4) \gamma_{a_2, a_4}(h-s, l_2, l_4)| \\
& \leq \sup_h \left(\sum_{r \in \mathbb{Z}} \gamma_{a_1, a_3}^2(r, l_1, l_3) \right)^{1/2} \left(\sum_{r \in \mathbb{Z}} \gamma_{a_2, a_3}^2(h-r, l_2, l_3) \right)^{1/2} \\
& \quad \times \left(\sum_{s \in \mathbb{Z}} \gamma_{a_1, a_4}^2(s, l_1, l_4) \right)^{1/2} \left(\sum_{s \in \mathbb{Z}} \gamma_{a_2, a_4}^2(h-s, l_2, l_4) \right)^{1/2} \\
& \leq 2^{J_2 - a_3} \sup_h \left(\sum_{r' \in \mathbb{Z}} \sum_{l'_3=0}^{2^{a_3 - a_1} - 1} \gamma_{a_1, a_3}^2(2^{a_3 - a_1} r' + l'_3) \right)^{1/2} \left(\sum_{r' \in \mathbb{Z}} \sum_{l'_3=0}^{2^{a_3 - a_2} - 1} \gamma_{a_2, a_3}^2(2^{a_3 - a_2} r' + l'_3) \right)^{1/2} \\
& \leq 2^{J_2 - a_4} \sup_h \left(\sum_{s' \in \mathbb{Z}} \sum_{l'_4=0}^{2^{a_4 - a_1} - 1} \gamma_{a_1, a_4}^2(2^{a_4 - a_1} s' + l'_4) \right)^{1/2} \left(\sum_{s' \in \mathbb{Z}} \sum_{l'_4=0}^{2^{a_4 - a_2} - 1} \gamma_{a_2, a_4}^2(2^{a_4 - a_2} s' + l'_4) \right)^{1/2} \\
& \leq \frac{C}{2^\pi} \sup_h \left(\int_{-\pi}^{\pi} \|\mathbf{D}_{a_3, a_3 - a_1}(\lambda; f)\|^2 d\lambda \right)^{1/2} \left(\int_{-\pi}^{\pi} \|\mathbf{D}_{a_3, a_3 - a_2}(\lambda; f)\|^2 d\lambda \right)^{1/2} \\
& \quad \times \frac{C'}{2^\pi} \left(\int_{-\pi}^{\pi} \|\mathbf{D}_{a_4, a_4 - a_1}(\lambda; f)\|^2 d\lambda \right)^{1/2} \left(\int_{-\pi}^{\pi} \|\mathbf{D}_{a_4, a_4 - a_2}(\lambda; f)\|^2 d\lambda \right)^{1/2} < \infty,
\end{aligned}$$

where C and C' are positive constants and $\mathbf{D}_{a_i, a_i - a_{i'}}(\lambda; f)$ is the $2^{a_i - a_{i'}}$ -dimensional vector of the cross spectral density between scale a_i and scale $a_{i'}$ where we assume that $i' \leq i$.

$$\begin{aligned}
& \sup_{h \in \mathbb{Z}} \sum_{r, s \in \mathbb{Z}} |\gamma_{a_1, a_2}(h, l_1, l_2) \gamma_{a_3, a_4}(r-s, l_3, l_4) \gamma_{a_2, a_3}(h-r, l_2, l_3) \gamma_{a_1, a_4}(s, l_1, l_4)| \\
& \leq \sup_h |\gamma_{a_1, a_2}(h, l_1, l_2)| \sum_{s \in \mathbb{Z}} |\gamma_{a_1, a_4}(s, l_1, l_4)| \left(\sum_{r \in \mathbb{Z}} \gamma_{a_2, a_3}^2(h-r, l_2, l_3) \right)^{1/2} \left(\sum_{r' \in \mathbb{Z}} \gamma_{a_3, a_4}^2(r', l_3, l_4) \right)^{1/2} \\
& \leq \sup_h |\gamma_{a_1, a_2}(h, l_1, l_2)| \sum_{s \in \mathbb{Z}} |\gamma_{a_1, a_4}(s, l_1, l_4)| \times \\
& \quad 2^{J_2 - \frac{a_3 + a_4}{2}} \left(\sum_{r' \in \mathbb{Z}} \sum_{l'_3=0}^{2^{a_3 - a_2} - 1} \gamma_{a_2, a_3}^2(2^{a_3 - a_2} r' + l'_3) \right)^{1/2} \left(\sum_{r' \in \mathbb{Z}} \sum_{l'_4=0}^{2^{a_4 - a_3} - 1} \gamma_{a_3, a_4}^2(2^{a_4 - a_3} r' + l'_4) \right)^{1/2} \\
& \leq \frac{C}{2^\pi} \sup_h |\gamma_{a_1, a_2}(h, l_1, l_2)| \sum_{s \in \mathbb{Z}} |\gamma_{a_1, a_4}(s, l_1, l_4)| \left(\int_{-\pi}^{\pi} \|\mathbf{D}_{a_3, a_3 - a_2}(\lambda; f)\|^2 d\lambda \right)^{1/2} \\
& \quad \times \left(\int_{-\pi}^{\pi} \|\mathbf{D}_{a_4, a_4 - a_3}(\lambda; f)\|^2 d\lambda \right)^{1/2} < \infty,
\end{aligned}$$

with C a positive and the same bound apply to

$$\sup_{h \in \mathbb{Z}} \sum_{r, s \in \mathbb{Z}} |\gamma_{a_1, a_2}(h, l_1, l_2) \gamma_{a_3, a_4}(r-s, l_3, l_4) \gamma_{a_2, a_4}(h-s, l_2, l_4) \gamma_{a_1, a_3}(r, l_1, l_3)|.$$

Then,

$$\sup_{h \in \mathbb{Z}} \sum_{r, s \in \mathbb{Z}} \sum_{\ell} \left| \mathcal{K} \left(W_{a_1, l_1}^2, W_{a_2, 2^{J_2 - a_2} h + l_2}^2, W_{a_3, 2^{J_2 - a_3} r + l_3}^2, W_{a_4, 2^{J_2 - a_4} s + l_4}^2 \right) \right| < \infty,$$

and (3.34) follows.

Let us prove now (3.32)

$$\begin{aligned}\mathbb{E}(\widetilde{\Gamma}^{\text{B}}(j', j)) &= \sum_{|\tau| \leq q(n_{J_2})} \mathbb{E} \left(\omega_\tau(q(n_{J_2})) \widetilde{\Sigma}_{j', j}^{\text{B}}(\tau) \right) = \frac{1}{n_{J_2}} \sum_{|\tau| \leq q(n_{J_2})} \omega_\tau(q(n_{J_2})) \sum_{i, i+\tau=1}^{n_{J_2}} \Sigma_{j', j}(\tau) \\ &= \sum_{|\tau| \leq q(n_{J_2})} \omega_\tau(q(n_{J_2})) \left(1 - \frac{|\tau|}{n_{J_2}}\right) \Sigma_{j', j}(\tau),\end{aligned}$$

$$\begin{aligned}\left| \mathbb{E}(\widetilde{\Gamma}^{\text{B}}(j', j)) - \Gamma(j', j) \right| &= 2 \sum_{|\tau| \leq q(n_{J_2})} \left(1 - \frac{|\tau|}{q(n_{J_2})}\right) \left(1 - \frac{|\tau|}{n_{J_2}}\right) \gamma_{j', j}^2(\tau) - 2 \sum_{\tau \in \mathbb{Z}} \gamma_{j', j}^2(\tau) \\ &= 2 \sum_{|\tau| \leq q(n_{J_2})} \left(-\frac{|\tau|}{q(n_{J_2})} - \frac{|\tau|}{n_{J_2}} + \frac{|\tau|^2}{q(n_{J_2})n_{J_2}} \right) \gamma_{j', j}^2(\tau) - 2 \sum_{|\tau| > q(n_{J_2})} \gamma_{j', j}^2(\tau) \\ &= O(q^{-1}(n_{J_2})), \text{ from (3.35)}\end{aligned}$$

which achieves the proof of Step 1

Step 2 Let us define

$$S_{n,j}(t) = \frac{1}{\sqrt{n_{J_2}}} \sum_{i=1}^{\lfloor n_j t \rfloor} (W_{j,i}^2 - \sigma_j^2), \text{ for all } j \in \{J_1, \dots, J_2\}.$$

Step 2 consists in proving that for $0 = t_0 \leq t_1 \leq \dots \leq t_k = 1$ and $\lambda_{i,j} \in \mathbb{R}$,

$$\sum_{j=J_1}^{J_2} \sum_{i=1}^k \{ \lambda_{j,i} (S_{n,j}(t_i)) \} \xrightarrow{d} \sum_{j=J_1}^{J_2} \sum_{i=1}^k \{ \lambda_{j,i} B_j(t_i) \} \text{ as } n \rightarrow \infty, \quad (3.36)$$

where $(B_j(t))_{j=J_1, \dots, J_2}$ are brownian motions satisfying $\text{Cov}(B_{J_1}, \dots, B_{J_2}) = \Gamma_{J_1, J_2}$. Observe that

$$\begin{aligned}\sum_{j=J_1}^{J_2} \sum_{i=1}^k \{ \lambda_{j,i} (S_{n,j}(t_i)) \} &= \sum_{j=J_1}^{J_2} \sum_{i=1}^k \lambda_{j,i} \frac{1}{\sqrt{n_{J_2}}} \sum_{l=1}^{n_j} (W_{j,l}^2 - E(W_{j,l}^2)) \mathbb{1}_{\{l \leq \lfloor n_j t_i \rfloor\}} \\ &= \sum_{j=J_1}^{J_2} \left\{ \sum_{l=1}^{n_j} W_{j,l}^2 \sum_{i=1}^k \lambda_{j,i} \frac{1}{\sqrt{n_{J_2}}} \mathbb{1}_{\{l \leq \lfloor n_j t_i \rfloor\}} - E \left(\sum_{l=1}^{n_j} W_{j,l}^2 \sum_{i=1}^k \lambda_{j,i} \frac{1}{\sqrt{n_{J_2}}} \mathbb{1}_{\{l \leq \lfloor n_j t_i \rfloor\}} \right) \right\}.\end{aligned}$$

We consider the gaussian vector $\mathbf{W}_n = [\widetilde{W}_{J_1}^T, \dots, \widetilde{W}_J^T, \dots, \widetilde{W}_{J_2}^T]^T$, where $\widetilde{W}_j^T = [W_{j,1}, \dots, W_{j,n_j}]^T$

and $a_{n,j,l} = \sum_{i=1}^k \lambda_{j,i} \frac{1}{\sqrt{n_{J_2}}} \mathbb{1}_{\{l \leq \lfloor n_j t_i \rfloor\}}$. Then

$$\sum_{j=J_1}^{J_2} \sum_{i=1}^k \{ \lambda_{j,i} (S_{n,j}(t_i)) \} = \mathbf{W}_n^T \mathbf{A}_n \mathbf{W}_n - E(\mathbf{W}_n^T \mathbf{A}_n \mathbf{W}_n)$$

where \mathbf{A}_n is the diagonal block matrix with diagonal entries $(A_{n,J_1}, \dots, A_{n,J_1}, \dots, A_{n,J_2})$ where each $A_{n,j}$ is a diagonal matrix with diagonal entries $(a_{n,j,1}, \dots, a_{n,j,l}, \dots, a_{n,j,n_j})$.

Applying Lemma 4 in Moulines et al. [2007a], (3.36) is obtain by proving that as $n_j \rightarrow \infty$,

$$\rho(\mathbf{A}_n)\rho(\Gamma_n) \rightarrow 0, \text{ and} \quad (3.37)$$

$$\text{Var} \left(\sum_{j=J_1}^{J_2} \sum_{i=1}^k \{ \lambda_{j,i}(S_{n,j}(t_i)) \} \right) \rightarrow \text{Var} \left(\sum_{j=J_1}^{J_2} \sum_{i=1}^k \{ \lambda_{j,i}B_j(t_i) \} \right) < \infty, \quad (3.38)$$

where Γ_n the covariance matrix of \mathbf{W}_n . Observe that

$$\rho(A_n) = \max_{J_1 \leq j \leq J_2} \left\{ \max_{1 \leq l \leq n_{J_1}} |a_{n,j,l}| = O(n_{J_2}^{-1/2}) \right\}.$$

Let $\Gamma_{n,J_2}, \dots, \Gamma_{n,J_1}$ denote the covariance matrix of $\widetilde{W}_{J_2}^T, \dots, \widetilde{W}_j^T, \dots, \widetilde{W}_{J_1}^T$ respectively. Then, $\Gamma_{n,j}$ is a $n_j \times n_j$ Toeplitz matrix associated to a bounded spectral density and thus $\rho(\Gamma_{n,j}) < \infty$ for $j = J_1, \dots, J_2$. From Lemma 6 in Moulines et al. [2007a], we get $\rho(\Gamma_n) < \sum_{j=J_1}^{J_2} \rho(\Gamma_{n,j}) < \infty$ and (3.37) holds. We now prove (3.38). Since each $(B_j(t))_{t \in [0,1]}$ have uncorrelated arguments, it is sufficient to prove that for $0 \leq r \leq s \leq t \leq 1$ and two scales j, j' with $j' < j$,

$$\text{Cov} (S_{n,j'}(t) - S_{n,j'}(s), S_{n,j}(r)) \rightarrow 0 \text{ as } n \rightarrow \infty, \quad (3.39)$$

$$\text{Cov} (S_{n,j'}(r), S_{n,j}(t) - S_{n,j}(s)) \rightarrow 0 \text{ as } n \rightarrow \infty, \quad (3.40)$$

$$\text{and, } \text{Cov} (S_{n,j}(t), S_{n,j'}(t)) \rightarrow \Gamma_{J_1, J_2}(j', j)t \text{ as } n \rightarrow \infty. \quad (3.41)$$

For any set $A, B \subseteq [0, 1]$, we set

$$V_n(J_2, j', j, \tau, l, u, A, B) = \frac{1}{n_{J_2}} \sum_{k \geq 1} \mathbb{1}_A \left(\frac{2^{J_2-j'}\tau + l}{n_{j'}} \right) \mathbb{1}_B \left(\frac{2^{J_2-j}k + u}{n_j} \right).$$

For all $0 \leq s, t \leq 1$, we have

$$\begin{aligned} \text{Cov} (S_{n,j'}(t), S_{n,j}(s)) &= \frac{1}{n_{J_2}} \sum_{i=1}^{\lfloor n_{j'}t \rfloor} \sum_{k=1}^{\lfloor n_js \rfloor} \text{Cov}(W_{j',i}^2, W_{j,k}^2) \\ &= \sum_{l=0}^{2^{J_2-j'}-1} \sum_{u=0}^{2^{J_2-j}-1} 2 \sum_{\tau \in \mathbb{Z}} \gamma_{j',j}^2(\tau, l, u) \frac{1}{n_{J_2}} \sum_{k \geq 1} \mathbb{1}_{\left\{ \frac{2^{J_2-j'}(k+\tau)+l}{n_{j'}} \in]0,t] \right\}} \mathbb{1}_{\left\{ \frac{2^{J_2-j}k+u}{n_j} \in]0,s] \right\}} \\ &= \sum_{l=0}^{2^{J_2-j'}-1} \sum_{u=0}^{2^{J_2-j}-1} 2 \sum_{\tau \in \mathbb{Z}} \gamma_{j',j}^2(\tau, l, u) V_n(J_2, j', j, \tau, l, u,]0, t],]0, s]) \\ \text{Cov} (S_{n,j'}(t) - S_{n,j'}(s), S_{n,j}(r)) &= \frac{1}{n_{J_2}} \sum_{i=\lfloor n_{j'}s \rfloor + 1}^{\lfloor n_{j'}t \rfloor} \sum_{k=1}^{\lfloor n_js \rfloor} \text{Cov}(W_{j',i}^2, W_{j,k}^2) \\ &= \sum_{l=0}^{2^{J_2-j'}-1} \sum_{u=0}^{2^{J_2-j}-1} 2 \sum_{\tau \in \mathbb{Z}} \gamma_{j',j}^2(\tau, l, u) V_n(J_2, j', j, \tau, l, u,]s, t],]0, r]). \end{aligned}$$

$\sup_{l,u,\tau} |V_n(J_2, j', j, \tau, l, u, A, B)| \leq \frac{k}{n_{J_2}} \leq 1$. Hence, by dominated convergence, the limit in (3.39), (3.40) and (3.41) are obtained by computing the limit of $V_n(J_2, j', j, \tau, l, u,]s, t],]0, r])$, $V_n(J_2, j', j, \tau, l, u,]0, r],]s, t])$ and $V_n(J_2, j', j, \tau, l, u,]0, t],]0, t])$. We have for any $\tau \in \mathbb{Z}$, $l = 0, \dots, 2^{J_2-j'} - 1$, $u = 0, \dots, 2^{J_2-j} - 1$ and $0 \leq r \leq s \leq t$.

$$\begin{aligned} n_{J_2} V_n(J_2, j', j, \tau, l, u,]s, t],]0, r]) &= \left\{ \left((2^{j'-J_2}(n_{j'}t - l) - \tau) \wedge (2^{j'-J_2}(n_{j'}r - u)) \right) \right. \\ &\quad \left. - \left((2^{j'-J_2}(n_{j'}s - l) - \tau) \vee (-2^{j'-J_2}u) \right) \right\}_+ \\ &= \left\{ 2^{j'-J_2}(n_{j'}r - u) - \left(2^{j'-J_2}(n_{j'}s - l) - \tau \right) \right\}_+ \rightarrow \tau \mathbb{1}_{\{(r,u)=(s,l)\}}, \end{aligned}$$

where the last equality holds for n large enough. Hence $V_n(J_2, j', j, \tau, l, u,]s, t],]0, r]) \rightarrow 0$ as $n \rightarrow \infty$ and (3.39) follows. Similarly, we have

$$\begin{aligned} n_{J_2} V_n(J_2, j', j, \tau, l, u,]0, r],]s, t]) &= \left\{ \left((2^{j'-J_2}(n_{j'}r - l) - \tau) \wedge (2^{j'-J_2}(n_{j'}t - u)) \right) \right. \\ &\quad \left. - \left((2^{j'-J_2}(n_{j'}s - u)) \vee (-2^{j'-J_2}l - \tau) \right) \right\}_+ \\ &= \left\{ 2^{j'-J_2}(n_{j'}r - l) - \tau - (2^{j'-J_2}(n_{j'}s - u)) \right\}_+ \rightarrow -\tau \mathbb{1}_{\{(r,l)=(s,u)\}}, \end{aligned}$$

where the last equality holds for n large enough. Hence $V_n(J_2, j', j, \tau, l, u,]0, r],]s, t]) \rightarrow 0$ as $n \rightarrow \infty$ and (3.40) follows.

$$\begin{aligned} \text{Cov}(S_{n,j}(t), S_{n,j'}(t)) &= \frac{1}{n_{J_2}} \sum_{i=1}^{[n_{j'}t]} \sum_{k=1}^{[n_{j'}t]} \text{Cov}(W_{j,i}^2, W_{j',k}^2) \\ &= \sum_{l=0}^{2^{J_2-j'}-1} \sum_{u=0}^{2^{J_2-j}-1} 2 \sum_{\tau \in \mathbb{Z}} \gamma_{j',j}^2(\tau, l, u) V_n(J_2, j, j', \tau, l, u,]0, t],]0, t]) \end{aligned}$$

We have for $t > 0$ and n large enough,

$$\begin{aligned} n_{J_2} V_n(J_2, j', j, \tau, l, u,]0, t],]0, t]) &= \left\{ \left((2^{j'-J_2}(n_{j'}t - l) - \tau) \wedge (2^{j'-J_2}(n_{j'}t - u)) \right) \right. \\ &\quad \left. - \left((-2^{j'-J_2}u) \vee (-2^{j'-J_2}l - \tau) \right) \right\}_+ \\ &= \left\{ 2^{j'-J_2}(n_{j'}t - u) + 2^{j'-J_2}l + \tau \right\}_+ \end{aligned}$$

Hence

$$V_n(J_2, j', j, \tau, l, u,]0, t],]0, t]) = \frac{1}{n_{J_2}} \left\{ 2^{j'-J_2}(n_{j'}t - u) + 2^{j'-J_2}l + \tau \right\}_+ \rightarrow t \text{ as } n \rightarrow \infty.$$

and (3.41) follows with $\Gamma_{J_1, J_2}(j', j) = 2 \sum_{l=0}^{2^{J_2-j'}-1} \sum_{u=0}^{2^{J_2-j}-1} \sum_{\tau \in \mathbb{Z}} \gamma_{j',j}^2(\tau, l, u)$, which achieves the proof of step 2.

Step 3. We have proved in Step 3 of Theorem 3.2.1 that each sequence of partial sum $\{S_{n,j}(t), t \in [0, 1]\}$ is tight for any scale $j \in \{J_1, \dots, J_2\}$ then $S_{J_1, J_2}(t)$ is also tight.

Finally, the convergence (4.31) follows from Slutsky's lemma and the three previous steps, which achieves the proof of Theorem 3.2.2. \square

3.3 Test statistics

Under the assumption of Theorem 3.2.1, the statistics

$$T_{J_1, J_2}(t) \stackrel{\text{def}}{=} (S_{J_1, J_2}(t) - tS_{J_1, J_2}(1))^T \widehat{\Gamma}_{J_1, J_2}^{-1} (S_{J_1, J_2}(t) - tS_{J_1, J_2}(1)) \quad (3.42)$$

converges in weakly in the Skorokhod space $D([0, 1])$

$$T_{J_1, J_2}(t) \xrightarrow{d} \sum_{\ell=1}^{J_2 - J_1 - 1} [B_\ell^0(t)]^2 \quad (3.43)$$

where $t \mapsto (B_1^0(t), \dots, B_{J_2 - J_1 - 1}^0(t))$ is a vector of $J_2 - J_1 + 1$ independent Brownian bridges

For any continuous function $F : D[0, 1] \rightarrow \mathbb{R}$, the continuous mapping Theorem implies that

$$F[T_{J_1, J_2}(\cdot)] \xrightarrow{d} F \left[\sum_{\ell=1}^{J_2 - J_1 - 1} [B_\ell^0(\cdot)]^2 \right].$$

We may for example apply either integral or max functionals, or weighted versions of these. A classical example of integral function is the so-called Cramér-Von Mises functional given by

$$\text{CVM}(J_1, J_2) \stackrel{\text{def}}{=} \int_0^1 T_{J_1, J_2}(t) dt, \quad (3.44)$$

which converges to $C(J_2 - J_1 + 1)$ where for any integer d ,

$$C(d) \stackrel{\text{def}}{=} \int_0^1 \sum_{\ell=1}^d [B_\ell^0(t)]^2 dt. \quad (3.45)$$

The test rejects the null hypothesis when $\text{CVM}_{J_1, J_2} \geq c(J_2 - J_1 + 1, \alpha)$, where $c(d, \alpha)$ is the $1 - \alpha$ th quantile of the distribution of $C(d)$. The distribution of the random variable $C(d)$ has been derived by Kiefer [1959] (see also Carmona et al. [1999] for more recent references). It holds that, for $x > 0$,

$$\mathbb{P}(C(d) \leq x) = \frac{2^{(d+1)/2}}{\pi^{1/2} x^{d/4}} \sum_{j=0}^{\infty} \frac{\Gamma(j + d/2)}{j! \Gamma(d/2)} e^{-(j+d/4)^2/x} \text{Cyl}_{(d-2)/2} \left(\frac{2j + d/2}{x^{1/2}} \right)$$

where Γ denotes the gamma function and Cyl are the parabolic cylinder functions. The quantile of this distribution are given in table 3.1 for different values of $d = J_2 - J_1 + 1$. It is also possible to use the max. functional leading to an analogue of the Kolmogorov-

Nominal S.	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$	$d = 6$
0.95	0.4605	0.7488	1.0014	1.2397	1.4691	1.6848
0.99	0.7401	1.0721	1.3521	1.6267	1.8667	2.1259

Table 3.1: Quantiles of the distribution $C(d)$ (see (3.45)) for different values of d

d	1	2	3	4	5	6
0.95	1.358	1.58379	1.7472	1.88226	2.00	2.10597
0.99	1.627624	1.842726	2.001	2.132572	2.24798	2.35209

Table 3.2: Quantiles of the distribution $D(d)$ (see (3.47)) for different values of d .

Smirnov statistics,

$$\text{KSM}(J_1, J_2) \stackrel{\text{def}}{=} \sup_{0 \leq t \leq 1} T_{J_1, J_2}(t) \quad (3.46)$$

which converges to $D(J_2 - J_1 + 1)$ where for any integer d ,

$$D(d) \stackrel{\text{def}}{=} \sup_{0 \leq t \leq 1} \sum_{\ell=1}^d [B_\ell^0(t)]^2. \quad (3.47)$$

The test reject the null hypothesis when $\text{KSM}_{J_1, J_2} \geq \delta(J_2 - J_1 + 1, \alpha)$, where $\delta(d, \alpha)$ is the $(1 - \alpha)$ -quantile of $D(d)$. The distribution of $D(d)$ has again be derived by Kiefer [1959] (see also Pitman and Yor [1999] for more recent references). It holds that, for $x > 0$,

$$\mathbb{P}(D(d) \leq x) = \frac{2^{1+(2-d)/2}}{\Gamma(d/2)a^d} \sum_{n=1}^{\infty} \frac{j_{\nu, n}^{2\nu}}{J_{\nu+1}^2(j_{\nu, n})} \exp\left(-\frac{j_{\nu, n}^2}{2x^2}\right),$$

where $0 < j_{\nu, 1} < j_{\nu, 2} < \dots$ is the sequence of positive zeros of J_ν , the Bessel function of index $\nu = (d - 2)/2$. The quantiles of this distribution are given in Table 3.2.

3.4 Power of the W2-CUSUM statistics

3.4.1 Power of the test in single scale case

In this section we investigate the power of the test. A minimal requirement is to establish that the test procedure is pointwise consistent in a presence of a breakpoint, *i.e.* that under a fixed alternative, the probability of detection converges to one as the sample size goes to infinity. We must therefore first define such alternative. For simplicity, we will consider an alternative where the process exhibit a single breakpoint, though it is likely that the test does have power against more general class of alternatives.

The alternative that we consider in this section is defined as follows. Let f_1 and f_2 be two given generalized spectral densities and suppose that, at a given scale j , $\int_{-\pi}^{\pi} |H_j(\lambda)|^2 f_i(\lambda) d\lambda < \infty$, $i = 1, 2$, and

$$\int_{-\pi}^{\pi} |H_j(\lambda)|^2 (f_1(\lambda) - f_2(\lambda)) d\lambda \neq 0. \quad (3.48)$$

Define by $(X_{l,i})_{l \in \mathbb{Z}}$, $i = 1, 2$, be two Gaussian processes, defined on the same probability space, with generalized spectral density f_1 . We do not specify the dependence structure between these two processes, which can be arbitrary. Let $\kappa \in]0, 1[$ be a breakpoint. We consider a sequence of Gaussian processes $(X_k^n)_{k \in \mathbb{Z}}$, such that

$$X_k^{(n)} = X_{k,i} \text{ for } k \leq \lfloor n\kappa \rfloor \text{ and } X_k^{(n)} = X_{k,2} \text{ for } k \geq \lfloor n\kappa \rfloor + 1. \quad (3.49)$$

Theorem 3.4.1. *Consider $\{X_k^n\}_{k \in \mathbb{Z}}$ be a sequence of processes specified by (3.48) and (3.49). Assume that $q(n_j)$ is non decreasing and :*

$$q(n_j) \rightarrow \infty \text{ and } \frac{q(n_j)}{n_j} \rightarrow 0 \text{ as } n_j \rightarrow \infty. \quad (3.50)$$

Then the statistic T_{n_j} defined by (3.3) satisfies

$$\frac{\sqrt{n_j}}{\sqrt{2q(n_j)}} \sqrt{\kappa(1-\kappa)}(1 + o_p(1)) \leq T_{n_j} \xrightarrow{P} \infty. \quad (3.51)$$

Proof. Let $k_j = \lfloor n_j \kappa \rfloor$ the change point in the wavelet spectrum at scale j . We write q for $q(n_j)$ and suppress the dependence in n in this proof to alleviate the notation. By definition $T_{n_j} = \sup_{0 \leq t \leq 1} |S_{n_j}(t) - tS_{n_j}(1)|$, where the process $t \mapsto S_{n_j}(t)$ is defined in (3.14). Therefore, $T_{n_j} \geq S_{n_j}(\kappa) - \kappa S_{n_j}(1)$. The proof consists in establishing that $S_{n_j}(\kappa) - \kappa S_{n_j}(1) = \frac{\sqrt{n_j}}{\sqrt{2q(n_j)}} \sqrt{\kappa(1-\kappa)}(1 + o_p(1))$. We first decompose this difference as follows

$$\begin{aligned} S_{n_j}(\kappa) - \kappa S_{n_j}(1) &= \frac{1}{\sqrt{n_j}} \left| \sum_{i=1}^{\lfloor n_j \kappa \rfloor} W_{j,i}^2 - \kappa \sum_{i=1}^{n_j} W_{j,i}^2 \right| \\ &= B_{n_j} + f_{n_j} \end{aligned}$$

where B_{n_j} is a fluctuation term

$$B_{n_j} = \frac{1}{\sqrt{n_j}} \left| \sum_{i=1}^{k_j} (W_{j,i}^2 - \sigma_{j,i}^2) - \kappa \sum_{i=1}^{n_j} (W_{j,i}^2 - \sigma_{j,i}^2) \right| \quad (3.52)$$

and f_{n_j} is a bias term

$$f_{n_j} = \frac{1}{\sqrt{n_j}} \left| \sum_{i=1}^{k_j} \sigma_{j,i}^2 - \kappa \sum_{i=1}^{n_j} \sigma_{j,i}^2 \right|. \quad (3.53)$$

Since support of $h_{j,l}$ is included in $[-T(2^j + 1), 0]$ where $h_{j,l}$ is defined in (2.37), there exists a constant $a > 0$ such that

$$W_{j,i} = W_{j,i;1} = \sum_{l \leq k} h_{j,2^j i - l} X_{l,1}, \text{ for } i < k_j, \quad (3.54)$$

$$W_{j,i} = W_{j,i;2} = \sum_{l > k} h_{j,2^j i - l} X_{l,2} \text{ for } i > k_j + a, \quad (3.55)$$

$$W_{j,i} = \sum_l h_{j,2^j i - l} X_l, \text{ for } k_j \leq i < k_j + a. \quad (3.56)$$

Since the process $\{X_{l,1}\}_{l \in \mathbb{Z}}$ and $\{X_{l,2}\}_{l \in \mathbb{Z}}$ are both K -th order covariance stationary, the two processes $\{W_{j,i;1}\}_{i \in \mathbb{Z}}$ and $\{W_{j,i;2}\}_{i \in \mathbb{Z}}$ are also covariance stationary. The wavelet coefficients $W_{j,i}$ for $i \in \{k_j, \dots, k_j + a\}$ are computed using observations from the two processes X_1 and X_2 . Let us show that there exists a constant $C > 0$ such that, for all integers l and τ ,

$$\text{Var} \left(\sum_{i=l}^{l+\tau} W_{j,i}^2 \right) \leq C\tau. \quad (3.57)$$

Using (3.11), we have, for $\epsilon = 1, 2$,

$$\text{Var} \left(\sum_{i=l}^{l+\tau} W_{j,i;\epsilon}^2 \right) \leq \frac{\tau}{\pi} \int_{-\pi}^{\pi} |\mathbf{D}_{j,0;\epsilon}(\lambda)|^2 d\lambda$$

where, $\mathbf{D}_{j,0;1}(\lambda)$ and $\mathbf{D}_{j,0;2}(\lambda)$ denote the spectral density of the stationary processes $\{W_{j,i;1}\}_{i \in \mathbb{Z}}$ and $\{W_{j,i;2}\}_{i \in \mathbb{Z}}$ respectively. Using Minkovski inequality, we have for $l \leq k_j \leq k_j + a < l + \tau$ that $\left(\text{Var} \sum_{i=l}^{l+\tau} W_{j,i}^2 \right)^{1/2}$ is at most

$$\begin{aligned} & \left(\text{Var} \sum_{i=l}^{k_j} W_{j,i}^2 \right)^{1/2} + \sum_{i=k_j+1}^{k_j+a} (\text{Var} W_{j,i}^2)^{1/2} + \left(\text{Var} \sum_{i=k_j+a+1}^{l+\tau} W_{j,i}^2 \right)^{1/2} \\ & \leq \left(\text{Var} \sum_{i=l}^{k_j} W_{j,i;1}^2 \right)^{1/2} + a \sup_i (\text{Var} W_{j,i}^2)^{1/2} + \left(\text{Var} \sum_{i=k_j+a+1}^{l+\tau} W_{j,i;2}^2 \right)^{1/2}. \end{aligned}$$

Observe that $\text{Var}(W_{j,i}^2) \leq 2(\sum_l |h_{j,l}|)^2 (\sigma_{j,1}^2 \vee \sigma_{j,2}^2)^2 < \infty$ for $k_j \leq i < k_j + a$, where

$$\sigma_{j;1}^2 = \mathbb{E} [W_{j,i;1}^2], \text{ and } \sigma_{j;2}^2 = \mathbb{E} [W_{j,i;2}^2] \quad (3.58)$$

The three last displays imply (3.57) and thus that B_{n_j} is bounded in probability. Moreover, since f_{n_j} reads

$$\begin{aligned} & \frac{1}{\sqrt{n_j}} \left| \sum_{i=1}^{\lfloor n_j \kappa \rfloor} \sigma_{j;1}^2 - \kappa \sum_{i=1}^{\lfloor n_j \kappa \rfloor} \sigma_{j;1}^2 - \kappa \sum_{i=\lfloor n_j \kappa \rfloor + 1}^{\lfloor n_j \kappa \rfloor + a} \sigma_{j,i}^2 - \kappa \sum_{i=\lfloor n_j \kappa \rfloor + a + 1}^{n_j} \sigma_{j;2}^2 \right| \\ & = \sqrt{n_j} \kappa (1 - \kappa) |\sigma_{j;1}^2 - \sigma_{j;2}^2| + O(n_j^{-1/2}), \end{aligned}$$

we get

$$S_{n_j}(\kappa) - \kappa S_{n_j}(1) = \sqrt{n_j} \kappa (1 - \kappa) (\sigma_{j;1}^2 - \sigma_{j;2}^2) + O_P(1). \quad (3.59)$$

We now study the denominator s_{j,n_j}^2 in (3.3). Denote by

$$\bar{\sigma}_j^2 = \frac{1}{n_j} \sum_{i=1}^{n_j} \sigma_{j,i}^2$$

the expectation of the scalogram (which now differs from the wavelet spectrum). Let us consider for $\tau \in \{0, \dots, q(n_j)\}$ $\widehat{\gamma}_j(\tau)$ the empirical covariance of the wavelet coefficients defined in (3.5).

$$\begin{aligned} \widehat{\gamma}_j(\tau) = & \frac{1}{n_j} \sum_{i=1}^{n_j-\tau} (W_{j,i}^2 - \bar{\sigma}_j^2)(W_{j,i+\tau}^2 - \bar{\sigma}_j^2) - \left(1 + \frac{\tau}{n_j}\right) (\bar{\sigma}_j^2 - \widehat{\sigma}_j^2)^2 \\ & + \frac{1}{n_j} (\widehat{\sigma}_j^2 - \bar{\sigma}_j^2) \left\{ \sum_{i=n_j-\tau+1}^{n_j} (W_{j,i}^2 - \bar{\sigma}_j^2) + \sum_{i=1}^{\tau} (W_{j,i}^2 - \bar{\sigma}_j^2) \right\}. \end{aligned}$$

Using Minkowski inequality and (3.57), there exists a constant C such that for all $1 \leq l \leq l + \tau \leq n_j$,

$$\begin{aligned} \left\| \sum_{i=l}^{l+\tau} (W_{j,i}^2 - \overline{\sigma_{j,n_j}^2}) \right\|_2 & \leq \left\| \sum_{i=l}^{l+\tau} (W_{j,i}^2 - \sigma_{j,i}^2) \right\|_2 + \left\| \sum_{i=l}^{l+\tau} (\sigma_{j,i}^2 - \overline{\sigma_{j,n_j}^2}) \right\|_2 \\ & \leq C(\tau^{1/2} + \tau), \end{aligned}$$

and similarly

$$\|\widehat{\sigma}_j^2 - \bar{\sigma}_j^2\|_2 \leq \frac{C}{\sqrt{n_j}}.$$

By combining these two latter bounds, the Cauchy-Schwarz inequality implies that

$$\left\| \frac{1}{n_j} (\widehat{\sigma}_j^2 - \bar{\sigma}_j^2) \sum_{i=l}^{l+\tau} (W_{j,i}^2 - \bar{\sigma}_j^2) \right\|_1 \leq \frac{C(\tau^{1/2} + \tau)}{n_j^{3/2}}.$$

Recall that $s_{j,n_j}^2 = \sum_{\tau=-q}^q w_\tau(q) \widehat{\gamma}_j(\tau)$ where $w_\tau(q)$ are the so-called Bartlett weights defined in (3.6). We now use the bounds above to identify the limit of s_{j,n_j}^2 as the sample size goes to infinity. The two previous identities imply that

$$\sum_{\tau=0}^q w_\tau(q) \left(1 + \frac{\tau}{n_j}\right) \|\bar{\sigma}_j^2 - \widehat{\sigma}_j^2\|_2 \leq C \frac{q^2}{n_j^{3/2}}$$

and

$$\sum_{\tau=0}^q w_\tau(q) \left\| \frac{1}{n_j} (\widehat{\sigma}_j^2 - \bar{\sigma}_j^2) \sum_{i=l}^{l+\tau} (W_{j,i}^2 - \bar{\sigma}_j^2) \right\|_1 \leq C \frac{q^2}{n_j^{3/2}},$$

Therefore, we obtain

$$s_{j,n_j}^2 = \sum_{\tau=-q}^q w_\tau(q) \widetilde{\gamma}_j(\tau) + O_P\left(\frac{q^2}{n_j^{3/2}}\right), \quad (3.60)$$

where $\widetilde{\gamma}_j(\tau)$ is defined by

$$\widetilde{\gamma}_j(\tau) = \frac{1}{n_j} \sum_{i=1}^{n_j-\tau} (W_{j,i}^2 - \bar{\sigma}_j^2)(W_{j,i+\tau}^2 - \bar{\sigma}_j^2). \quad (3.61)$$

Observe that since $q = o(n_j)$, $k_j = \lfloor n_j \kappa \rfloor$ and $0 \leq \tau \leq q$, then for any given integer a and n large enough $0 \leq \tau \leq k_j \leq k_j + a \leq n_j - \tau$ thus in (3.61) we may write $\sum_{i=1}^{n_j - \tau} = \sum_{i=1}^{k_j - \tau} + \sum_{i=k_j - \tau + 1}^{k_j + a} + \sum_{i=k_j + a + 1}^{n_j - \tau}$. Using $\sigma_{j;1}^2$ and $\sigma_{j;2}^2$ in (3.61) and straightforward bounds that essentially follow from (3.57), we get $s_{j,n_j}^2 = \bar{s}_{j,n_j}^2 + o_P\left(\frac{q^2}{n_j}\right)$, where

$$\begin{aligned} \bar{s}_{j,n_j}^2 = \sum_{\tau=-q}^q w_\tau(q) & \left(\frac{k}{n_j} \tilde{\gamma}_{j;1}(\tau) + \frac{n_j - k_j - a}{n_j} \tilde{\gamma}_{j;2}(\tau) \right. \\ & \left. + \frac{k_j - |\tau|}{n_j} (\sigma_{j;1}^2 - \bar{\sigma}_j^2)^2 + \frac{n_j - k_j - a - |\tau|}{n_j} (\sigma_{j;2}^2 - \bar{\sigma}_j^2)^2 \right) \end{aligned}$$

with

$$\begin{aligned} \tilde{\gamma}_{j;1}(\tau) &= \frac{1}{k_j} \sum_{i=1}^{k_j - \tau} (W_{j,i}^2 - \sigma_{j;1}^2) (W_{j,i+\tau}^2 - \sigma_{j;1}^2), \\ \tilde{\gamma}_{j;2}(\tau) &= \frac{1}{n_j - k_j - a} \sum_{i=k_j + a + 1}^{n_j - \tau} (W_{j,i}^2 - \sigma_{j;2}^2) (W_{j,i+\tau}^2 - \sigma_{j;2}^2). \end{aligned}$$

Using that $\bar{\sigma}_j^2 \rightarrow \kappa \sigma_{j;1}^2 + (1 - \kappa) \sigma_{j;2}^2$ as $n_j \rightarrow \infty$, and that, for $\epsilon = 1, 2$,

$$s_{j,n_j;\epsilon}^2 \stackrel{\text{def}}{=} \sum_{\tau=-q}^q w_\tau(q) \tilde{\gamma}_{j;\epsilon}(\tau) \xrightarrow{P} \frac{1}{\pi} \int_{-\pi}^{\pi} |\mathbf{D}_{j,0;\epsilon}(\lambda)|^2 d\lambda,$$

we obtain

$$\begin{aligned} s_{j,n_j}^2 &= \frac{1}{\pi} \int_{-\pi}^{\pi} \left\{ \kappa |\mathbf{D}_{j,0;1}(\lambda)|^2 + (1 - \kappa) |\mathbf{D}_{j,0;2}(\lambda)|^2 \right\} d\lambda \\ &= 2q\kappa(1 - \kappa) (\sigma_{j;1}^2 - \sigma_{j;2}^2)^2 + o_p(1) + O_P\left(\frac{q^2}{n_j}\right). \end{aligned} \quad (3.62)$$

Using (3.59), the last display and that $o_p(1) + O_P\left(\frac{q^2}{n_j}\right) = o_p(q)$, we finally obtain

$$\begin{aligned} S_{n_j}(\kappa) - \kappa S_{n_j}(1) &= \frac{\sqrt{n_j} \kappa (1 - \kappa) \left| \sigma_{j;1}^2 - \sigma_{j;2}^2 \right| + O_P(1)}{\sqrt{2q(\kappa(1 - \kappa)) \left| \sigma_{j;1}^2 - \sigma_{j;2}^2 \right|} + o_p(\sqrt{q})} \\ &= \frac{\sqrt{n_j}}{\sqrt{2q}} \sqrt{\kappa(1 - \kappa)} (1 + o_p(1)), \end{aligned}$$

which concludes the proof of Theorem 3.4.1. \square

3.4.2 Power of the test in multiple scales case

The results obtained in the previous Section in the single scale case easily extend to the test procedure designed to handle the multiple scales case. The alternative is specified exactly in the same way than in the single scale case but instead of considering the square

of the wavelet coefficients at a given scale, we now study the behavior of the between-scale process. Consider the following process for $\epsilon = 1, 2$,

$$Y_{J_1, J_2, i; \epsilon} = \left(W_{J_2, i; \epsilon}^2 \sum_{u=1}^2 W_{J_2-1, 2(i-1)+u; \epsilon}^2, \dots, \sum_{u=1}^{2^{(J_2-J_1)}} W_{J_1, 2^{(J_2-J_1)}(i-1)+u; \epsilon}^2 \right)^T,$$

where J_1 and J_2 are respectively the finest and the coarsest scale considered in the test, $W_{j, i; \epsilon}$ are defined in (3.54) and (3.55) and $\Gamma_{J_1, J_2; \epsilon}$ the $(J_2 - J_1 + 1) \times (J_2 - J_1 + 1)$ symmetric non negative matrix such that

$$\Gamma_{J_1, J_2; \epsilon}(j, j') = \sum_{h \in \mathbb{Z}} \text{Cov}(Y_{j, 0; \epsilon}, Y_{j', h; \epsilon}) = \int_{-\pi}^{\pi} \|\mathbf{D}_{j, u; \epsilon}(\lambda; f)\|^2 d\lambda, \quad (3.63)$$

with $1 \leq j, j' \leq J_2 - J_1 + 1$ for $\epsilon = 1, 2$.

Theorem 3.4.2. *Consider $\{X_k^n\}_{k \in \mathbb{Z}}$ be a sequence of processes specified by (3.48) and (3.49). Assume in addition that for at least one $j \in \{J_1, \dots, J_2\}$ and that at least one of the two matrices $\Gamma_{J_1, J_2; \epsilon}$ $\epsilon = 1, 2$ defined in (3.63) is positive definite. Assume in addition that Finally, assume that the number of lags $q(n_{J_2})$ in the Barlett estimate of the covariance matrix (3.25) is non decreasing and:*

$$q(n_{j_2}) \rightarrow \infty \quad \text{and} \quad \frac{q^2(n_{J_2})}{n_{J_2}} \rightarrow 0, \quad \text{as } n_{J_2} \rightarrow \infty, \quad (3.64)$$

Then, the W_2 -CUSUM test statistics T_{J_1, J_2} defined by (4.32) satisfies

$$\frac{n_{J_2}}{2q(n_{J_2})} \kappa(1 - \kappa)(1 + o_p(1)) \leq T_{J_1, J_2} \xrightarrow{P} \infty \quad \text{as } n_{J_2} \rightarrow \infty$$

Proof. As in the single scale case we drop the dependence in n_{J_2} in the expression of q in this proof section. Let $k_j = \lfloor n_j \kappa \rfloor$ the change point in the wavelet spectrum at scale j . Then using (4.30), we have that $T_{J_1, J_2} \geq S_{J_1, J_2}(\kappa) - \kappa S_{J_1, J_2}(1)$ where

$$S_{J_1, J_2}(\kappa) - \kappa S_{J_1, J_2}(1) = \frac{1}{\sqrt{n_{J_2}}} [n_j(B_{n_j} + f_{n_j})]_{j=J_1, \dots, J_2},$$

where B_{n_j} and f_{n_j} are defined respectively by (3.52) and (3.53). Hence as in (3.59), we have

$$S_{J_1, J_2}(\kappa) - \kappa S_{J_1, J_2}(1) = \sqrt{n_{J_2}} \kappa(1 - \kappa) \Delta + O_P(1),$$

where $\Delta = \left[\sigma_{J_1, J_2; 1}^2 - \sigma_{J_1, J_2; 2}^2 \right]^T$ and

$$\sigma_{J_1, J_2; \epsilon}^2 = (\sigma_{J_2; \epsilon}^2, \dots, 2^{J_2 - J_1} \sigma_{J_1; \epsilon}^2)^T.$$

We now study the asymptotic behavior of $\widehat{\Gamma}_{J_1, J_2}$. Using similar arguments as those leading to (3.62) in the proof of Theorem 3.4.1, we have

$$\begin{aligned} \widehat{\Gamma}_{J_1, J_2} &= 2q\kappa(1 - \kappa)\Delta\Delta^T + \kappa\Gamma_{J_1, J_2; 1} + (1 - \kappa)\Gamma_{J_1, J_2; 2} \\ &\quad + O_P\left(\frac{q}{n_{J_2}}\right) + O_P(q^{-1}) + O_P\left(\frac{q^2}{n_{J_2}}\right). \end{aligned}$$

For Γ a positive definite matrix, consider the matrix $\mathbf{M}(\Gamma) = \Gamma + 2q\kappa(1 - \kappa)\Delta\Delta^T$. Using the matrix inversion lemma, the inverse of $\mathbf{M}(\Gamma)$ may be expressed as

$$\mathbf{M}^{-1}(\Gamma) = \left(\Gamma^{-1} - \frac{2q\kappa(1 - \kappa)\Gamma^{-1}\Delta\Delta^T\Gamma^{-1}}{1 + 2q\kappa(1 - \kappa)\Delta^T\Gamma^{-1}\Delta} \right),$$

which implies that

$$\Delta^T\mathbf{M}^{-1}(\Gamma)\Delta = \frac{\Delta^T\Gamma^{-1}\Delta}{1 + 2q\kappa(1 - \kappa)\Delta^T\Gamma^{-1}\Delta}.$$

Applying these two last relations to $\Gamma_0 = \kappa\Gamma_{J_1, J_2}^{(1)} + (1 - \kappa)\Gamma_{J_1, J_2}^{(2)}$ which is symmetric and definite positive (since, under the stated assumptions at least one of the two matrix $\Gamma_{J_1, J_2; \epsilon}$, $\epsilon = 1, 2$ is positive) we have

$$\begin{aligned} T_{J_1, J_2} &\geq \kappa^2(1 - \kappa)^2 n_{J_2} \Delta^T \mathbf{M}^{-1} \left(\Gamma_0 + O_P \left(\frac{q^2}{n_{J_2}} \right) + O_P(q^{-1}) \right) \Delta + O_P(1) \\ &= n_{J_2} \kappa^2 (1 - \kappa)^2 \frac{\Delta^T \Gamma_0^{-1} \Delta + O_P \left(\frac{q^2}{n_{J_2}} \right) + O_P(q^{-1})}{2q\kappa(1 - \kappa)\Delta^T \Gamma_0^{-1} \Delta (1 + o_p(1))} + O_P(1) \\ &= \frac{n_{J_2}}{2q} \kappa(1 - \kappa) (1 + o_p(1)) . \end{aligned}$$

Thus $T_{J_1, J_2} \xrightarrow{P} \infty$ as $n_{J_2} \rightarrow \infty$, which completes the proof of Theorem 3.4.2. \square

Remark 3.4.1. *The term corresponding to the "bias" term $\kappa\Gamma_{J_1, J_2; 1} + (1 - \kappa)\Gamma_{J_1, J_2; 2}$ in the single case is $\frac{1}{\pi} \int_{-\pi}^{\pi} \{ \kappa |\mathbf{D}_{j, 0; 1}(\lambda)|^2 + (1 - \kappa) |\mathbf{D}_{j, 0; 2}(\lambda)|^2 \} d\lambda = O(1)$, which can be neglected since the main term in s_{j, n_j}^2 is of order $q \rightarrow \infty$. In multiple scale case, the main term in $\widehat{\Gamma}_{J_1, J_2}$ is still of order q but is no longer invertible (the rank of the leading term is equal to 1). A closer look is thus necessary and the term $\kappa\Gamma_{J_1, J_2; 1} + (1 - \kappa)\Gamma_{J_1, J_2; 2}$ has to be taken into account. This also explains why we need the more stringent condition (3.64) on the bandwidth size in the multiple scales case.*

3.5 Some examples

In this section, we report the results of a limited Monte-Carlo experiment to assess the finite sample property of the test procedure. Recall that the test rejects the null if either $\text{CVM}(J_1, J_2)$ or $\text{KSM}(J_1, J_2)$, defined in (4.34) and (4.35) exceeds the $(1 - \alpha)$ -th quantile of the distributions $C(J_2 - J_1 + 1)$ and $D(J_2 - J_1 + 1)$, specified in (3.45) and (3.47). The quantiles are reported in Tables (3.1) and (3.2), and have been obtained by truncating the series expansion of the cumulative distribution function. To study the influence on the test procedure of the strength of the dependency, we consider different classes of Gaussian processes, including white noise, autoregressive moving average (ARMA) processes as well as fractionally integrated ARMA (ARFIMA(p, d, q)) processes which are known to be long range dependent. In all the simulations we set the lowest scale to $J_1 = 1$ and vary the

White noise						
n		512	1024	2048	4096	8192
$J = 3$	KSM	0.02	0.01	0.03	0.02	0.02
$J = 3$	CVM	0.05	0.045	0.033	0.02	0.02
$J = 4$	KSM	0.047	0.04	0.04	0.02	0.02
$J = 4$	CVM	0.041	0.02	0.016	0.016	0.01
$J = 5$	KSM	0.09	0.031	0.02	0.025	0.02
$J = 5$	CVM	0.086	0.024	0.012	0.012	0.02

Table 3.3: Empirical level of KSM – CVM for a white noise.

MA(1)[$\theta = 0.9$]						
n		512	1024	2048	4096	8192
$J = 3$	KSM	0.028	0.012	0.012	0.012	0.02
$J = 3$	CVM	0.029	0.02	0.016	0.016	0.01
$J = 4$	KSM	0.055	0.032	0.05	0.025	0.02
$J = 4$	CVM	0.05	0.05	0.03	0.02	0.02
$J = 5$	KSM	0.17	0.068	0.02	0.02	0.02
$J = 5$	CVM	0.13	0.052	0.026	0.021	0.02

Table 3.4: Empirical level of KSM – CVM for a $MA(q)$ process.

coarsest scale $J_2 = J$. We used a wide range of values of sample size n , of the number of scales J and of the parameters of the ARMA and FARIMA processes but, to conserve space, we present the results only for $n = 512, 1024, 2048, 4096, 8192$, $J = 3, 4, 5$ and four different models: an AR(1) process with parameter 0.9, a MA(1) process with parameter 0.9, and two ARFIMA(1,d,1) processes with memory parameter $d = 0.3$ and $d = 0.4$, and the same AR and MA coefficients, set to 0.9 and 0.1. In our simulations, we have used the Newey-West estimate of the bandwidth $q(n_j)$ for the covariance estimator (as implemented in the R-package *sandwich*).

Asymptotic level of KSM and CVM.

We investigate the finite-sample behavior of the test statistics $CVM(J_1, J_2)$ and $KSM(J_1, J_2)$ by computing the number of times that the null hypothesis is rejected in 1000 independent replications of each of these processes under \mathcal{H}_0 , when the asymptotic level is set to 0.05.

We notice that in general the empirical levels for the CVM are globally more accurate than the ones for the KSM test, the difference being more significant when the strength of the dependence is increased, or when the number of scales that are tested simultaneously get larger. The tests are slightly too conservative in the white noise and the MA case (tables 3.3 and 3.4); in the AR(1) case and in the ARFIMA cases, the test rejects the null much too often when the number of scales is large compared to the sample size (the difficult problem being in that case to estimate the covariance matrix of the test). For

AR(1)[$\phi = 0.9$]						
n		512	1024	2048	4096	8192
$J = 3$	KSM	0.1	0.073	0.078	0.064	0.04
$J = 3$	CVM	0.05	0.05	0.043	0.032	0.03
$J = 4$	KSM	0.36	0.234	0.21	0.12	0.073
$J = 4$	CVM	0.2	0.12	0.12	0.084	0.068
$J = 5$	KSM	0.547	0.314	0.254	0.22	0.11
$J = 5$	CVM	0.378	0.221	0.162	0.14	0.093

Table 3.5: Empirical level of KSM – CVM for an $AR(1)$ process.

ARFIMA(1,0.3,1)[$\phi = 0.9, \theta = 0.1$]						
n		512	1024	2048	4096	8192
$J = 3$	KSM	0.068	0.047	0.04	0.031	0.032
$J = 3$	CVM	0.05	0.038	0.03	0.02	0.02
$J = 4$	KSM	0.45	0.42	0.31	0.2	0.11
$J = 4$	CVM	0.39	0.32	0.20	0.16	0.091
$J = 5$	KSM	0.87	0.742	0.649	0.59	0.32
$J = 5$	CVM	0.41	0.352	0.292	0.16	0.17

Table 3.6: Empirical level of KSM – CVM for an $ARFIMA(1, 0.3, 1)$ process.

$J = 4$, the number of samples required to meet the target rejection rate can be as large as $n = 4096$ for the CVM test and $n = 8192$ for the KSM test. The situation is even worse in the ARFIMA case (table 3.6). When the number of scales is equal to 4 or 5, the test rejects the null hypothesis much too often.

Power of KSM and CVM.

We assess the power of test statistic by computing the test statistics in presence of a change in the spectral density. To do so, we consider an observation obtained by concatenation of n_1 observations from a first process and n_2 observations from a second process, independent from the first one and having a different spectral density. The length of the resulting observations is $n = n_1 + n_2$. In all cases, we set $n_1 = n_2 = n/2$, and we present the results for $n_1 = 512, 1024, 2048, 4096$ and scales $J = 4, 5$. We consider the following situations: the two processes are white Gaussian noise with two different variances, two AR processes with different values of the autoregressive coefficient, two MA processes with different values of the moving average coefficient and two ARFIMA with same moving average and same autoregressive coefficients but different values of the memory parameter d . The scenario considered is a bit artificial but is introduced here to assess the ability of the test to detect abrupt changes in the spectral content. For 1000 simulations, we report the number of times \mathcal{H}_1 was accepted, leading the following results.

The power of our two statistics gives us satisfying results for the considered processes,

white-noise		[$\sigma_1^2 = 1, \sigma_2^2 = 0.7$]			
$n_1 = n_2$		512	1024	2048	4096
$J = 4$	KSM	0.39	0.78	0.89	0.95
$J = 4$	CVM	0.32	0.79	0.85	0.9
$J = 5$	KSM	0.42	0.79	0.91	0.97
$J = 5$	CVM	0.40	0.78	0.9	0.9

Table 3.7: Power of KSM – CVM on two white noise processes.

MA(1)+MA(1)		[$\theta_1 = 0.9, \theta_2 = 0.5$]			
$n_1 = n_2$		512	1024	2048	4096
$J = 4$	KSM	0.39	0.69	0.86	0.91
$J = 4$	CVM	0.31	0.6	0.76	0.93
$J = 5$	KSM	0.57	0.74	0.84	0.94
$J = 5$	CVM	0.46	0.69	0.79	0.96

Table 3.8: Power of KSM – CVM on a concatenation of two different MA processes.

AR(1)+AR(1)		[$\phi_1 = 0.9, \phi_2 = 0.5$]			
$n_1 = n_2$		512	1024	2048	4096
$J = 4$	KSM	0.59	0.72	0.81	0.87
$J = 4$	CVM	0.53	0.68	0.79	0.9
$J = 5$	KSM	0.75	0.81	0.94	0.92
$J = 5$	CVM	0.7	0.75	0.89	0.91

Table 3.9: Power of KSM – CVM on a concatenation of two different AR processes.

ARFIMA(1,0.3,1)	+	ARFIMA(1,0.4,1)	[$\phi = 0.9, \theta = 0.1$]		
$n_1 = n_2$		512	1024	2048	4096
$J = 4$	KSM	0.86	0.84	0.8	0.81
$J = 4$	CVM	0.81	0.76	0.78	0.76
$J = 5$	KSM	0.94	0.94	0.9	0.92
$J = 5$	CVM	0.93	0.92	0.96	0.91

Table 3.10: Power of KSM – CVM two ARFIMA(1,d,1) with same AR and MA part but two different values of memory parameter d .

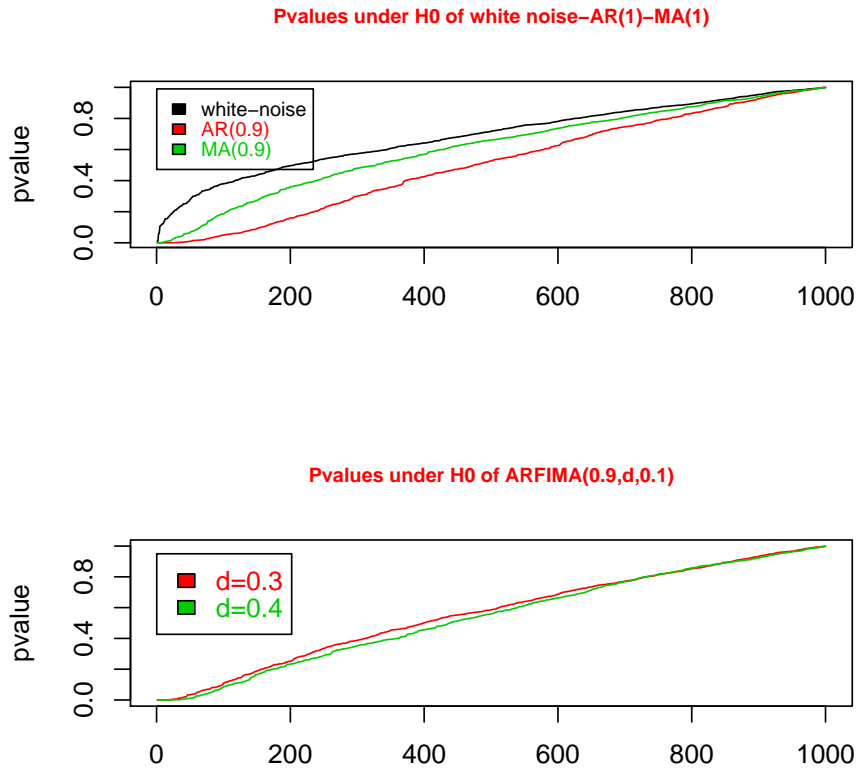


Figure 3.3: Pvalue under \mathcal{H}_0 of the distribution $D(J)$ $n = 1024$ for white noise and MA(1) processes and $n = 4096$ for AR(1) and ARFIMA(1,d,1) processes; the coarsest scale is $J = 4$ for white noise, MA and AR processes and $J = 3$ for the ARFIMA process. The finest scale is $J_1 = 1$.

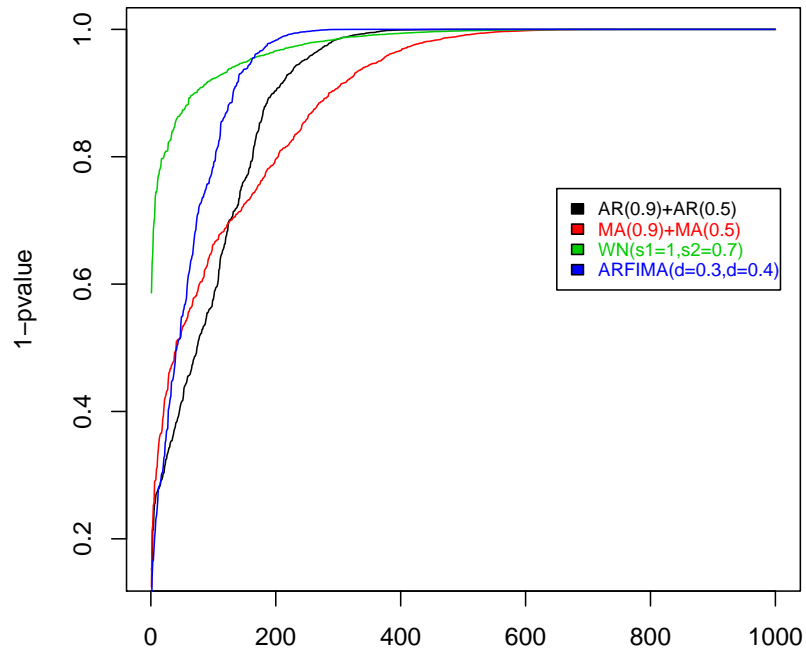


Figure 3.4: Empirical power of KSM($d = 4$) for white noise, AR, MA and ARFIMA processes.

n_1	512	512	512	1024	4096	8192
n_2	512	2048	8192	1024	4096	8192
mean _{CVM}	478	822	1853	965	3945	8009
med _{CVM}	517	692	1453	1007	4039	8119
IC _{CVM}	[283,661]	[380,1369]	[523,3534]	[637,1350]	[3095,4614]	[7962,8825]

Table 3.11: Estimation of the change point and confidence interval at 95% in the generalized spectral density of a process which is obtain by concatenation of two AR(1) processes.

n_1	512	512	512	1024	4096	8192
n_2	512	2048	8192	1024	4096	8192
mean _{CVM}	531	1162	3172	1037	4129	8037
med _{CVM}	517	1115	3215	1035	4155	8159
IC _{CVM}	[227,835]	[375,1483]	[817,6300]	[527,1569]	[2985,5830]	[6162,9976]

Table 3.12: Estimation of the change point and confidence interval at 95% in the generalized spectral density of a process which is obtain by concatenation of two ARFIMA(1,d,1) processes.

especially if the sample size tends to infinity.

Estimation of the change point in the original process.

We know that for each scale j , the number n_j of wavelet coefficients is $n_j = 2^{-j}(n - T + 1) - T + 1$. If we denote by k_j the change point in the wavelet coefficients at scale j and k the change point in the original signal, then $k = 2^j(k_j + T - 1) + T - 1$. In this paragraph, we estimate the change point in the generalized spectral density of a process when it exists and give its 95% confidence interval. For that, we proceed as before. We consider an observation obtained by concatenation of n_1 observations from a first process and n_2 observations from a second process, independent from the first one and having a different spectral density. The length of the resulting observations is $n = n_1 + n_2$. we estimate the change point in the process and we present the result for $n_1 = 512, 1024, 4096, 8192$, $n_2 = 512, 2048, 8192$, $J = 3$, the statistic CVM, two AR processes with different values of the autoregressive coefficient and two ARFIMA with same moving average and same autoregressive coefficients but different values of the memory parameter d . For 10000 simulations, the bootstrap confidence intervals obtained are set in the tables below. we also give the empirical mean and the median of the estimated change point.

We remark that the change point belongs always to the considered confidence interval excepted for $n_1 = 512$, $n_2 = 8192$ where the confidence interval is [523, 3534] and the change point $k = 512$ doesn't belong it. One can noticed that when the size of the sample increases and $n_1 = n_2$, the interval becomes more accurate. However, as expected, this interval becomes less accurate when the change appears either at the beginning or at the end of the observations.

Chapter 4

Inference of a generalized long memory process in the wavelet domain

The material of this chapter is the article entitled "Inference of a generalized long memory process in the wavelet domain" which is submitted to *IEEE Transactions on Signal Processing*.

4.1 Introduction

Long-memory processes have been successfully used to model the persistent correlation that is empirically observed for time series in several fields such as hydrology, finance, astronomy, and Internet traffic, see for example Hurst [1951], Beran [1994], Doukhan et al. [2003]. These processes have unbounded spectral densities and are often characterized by auto-correlations that decay at a polynomial rate with time. This unusually slow decreasing is interpreted as *long-memory*. The power law exponent is related to the **long memory parameter**, henceforth denoted by d . A central example is obtained by fractionally integrating a white noise. In this case, d corresponds to the integration order. More generally, a long memory process can be defined as a short memory process (for which $d = 0$) that has been fractionally integrated and d corresponds to the integration order. It is possible to extend this definition to any real values of d , yielding *negative long memory* if $d < 0$ and non-stationary processes of $d \geq 1/2$. For instance, unit root processes correspond to $d = 1$. Several methods have been proposed to estimate the memory parameter d of a process in parametric or semi-parametric framework. The most well known method, introduced by Geweke and Porter-Hudak [1983] in a parametric context, consists in estimating d by performing a log-regression on the periodogram. This

methodology has been adopted developed in the semi-parametric framework by Robinson [1995a]. Another estimator of the memory parameter is the Fourier-based Whittle estimator which is a pseudo-maximum likelihood method developed by Fox and Taqqu [1987] in a parametric setting and extend in a semi-parametric setting by Künsch [1987], Robinson [1995b]. These methods are based on the periodogram and thus on a Fourier analysis. Similar methods have also been proposed based on a wavelet analysis to estimate the memory parameter, see for instance Abry and Veitch [1998], Percival and Walden [2000], Moulines et al. [2007b]. Statistical performances of Wavelet and Fourier methods for long memory estimation have been compared in the recent review paper Faÿ et al. [2009]. Roughly speaking, the main inconvenient of wavelet methods lies in the fact that the second order properties of wavelet coefficients cannot be approximated as being uncorrelated. For instance, if such an approximation is made, the obtained asymptotic variance of the wavelet estimators are not correct. This is in sharp contrast with Fourier estimators, for which such an approximation yields correct asymptotic variances. On the other hand, estimation based on wavelet analysis offers several practical advantages. First, wavelet representations are widely spread for their sparsity properties and it is thus of interest to develop methods that directly apply in the wavelet domain for pre- or post-processing purposes. Second, wavelet estimators are easy to localize in time. As a consequence, they are better adapted for tracking changes of the spectral content (possibly in a long memory context), either looking for abruptly changing spectral densities as in Kouamo et al. [2010] or smoothly evolving ones as in Roueff and von Sachs [2009]. It is thus of interest to investigate further how to improve wavelet methods for long memory estimation. The first step in this direction is to develop a finer analysis of the second order characteristics of wavelet coefficients. The main goal of this paper is to provide low complexity algorithms to compute these characteristics in a very general fashion that applies in the particular case of long memory processes with arbitrary long memory parameter d . Two applications are then investigated: an estimator of d based on the likelihood maximization in the wavelet domain and an estimator of the scalogram covariance matrix used in spectral change detection in the wavelet domain. In these two applications, the obtained methods exhibit significantly better statistical performances.

Chapter is organized as follows. In Section 4.2, we introduce wavelet and wavelet transform for times series using Multiresolution analysis. In Section 4.3, we derive iterative formula for computing the exact second order characteristics of the wavelet coefficients of a very general class of processes that includes $M(d)$ processes for arbitrary values of d , provided that the wavelet enjoy sufficiently many vanishing moments. Section 4.4 is dedicated to several applications of this approach. In particular we propose a maximum likelihood estimation approach in the wavelet domain and an alternative estimator to the Bartlett estimator of the scalogram covariance matrix. The former is used as an

alternative to the local Whittle wavelet estimator for the semi-parametric estimation of d . The latter is used in the context of detection of changes in the spectral content of a (possibly long range dependent) time series. Numerical experiments on synthetic or real data are used to illustrate these approaches. Finally Appendix 4.5 contains postponed proofs.

4.2 Wavelet setting.

4.2.1 Multiresolution analysis

We consider a pair $\{\phi, \psi\}$ of *father* and *mother* wavelets defining a multiresolution analysis (MRA), see Cohen [2003]. This is the case for the celebrated Daubechies wavelets. MRA wavelets are widely used in practice, since they allow a fast computation of the discrete wavelet transform (DWT) using the *pyramidal algorithm*; see for instance Mallat [1998]. We adopt the engineering convention under which large values of the scale index j correspond to coarse scales, that is, the analyzing wavelets are defined as $\{\psi_{j,k}, j \in \mathbb{Z}, k \in \mathbb{Z}\}$ with $\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k)$, $j \in \mathbb{Z}, k \in \mathbb{Z}$. The DWT of X can be interpreted as follows. Using the father wavelet (or *scaling function*) ϕ and the interpolated function \mathbf{X} defined in (2.30), the wavelet coefficients are defined as

$$W_{j,k} \stackrel{\text{def}}{=} \int \mathbf{X}(t) \psi_{j,k}(t) dt, \quad j \geq 1, k \in \mathbb{Z}. \quad (4.1)$$

Based on the quadrature mirror filters h and g defined respectively in (2.45) and (2.48), one can compute the wavelet coefficients as follows

$$A_{0,k} = X_k, \quad k \in \mathbb{Z}, \quad (4.2)$$

$$A_{j,k} = \sum_{n=-\infty}^{\infty} h[2k - n] A_{j-1,n} = [\downarrow^2 (h \star A_{j-1,\cdot})]_k, \quad k \in \mathbb{Z}, j \geq 1, \quad (4.3)$$

$$W_{j,k} = \sum_{n=-\infty}^{\infty} g[2k - n] A_{j-1,n} = [\downarrow^2 (g \star A_{j-1,\cdot})]_k, \quad k \in \mathbb{Z}, j \geq 1, \quad (4.4)$$

where $\{h[k]\}_{k \in \mathbb{Z}}$ and $\{g[k]\}_{k \in \mathbb{Z}}$ denote the impulse response of h and g , \star denotes the convolution operator and \downarrow^2 the down-sampling operator by factor 2. We assume that ψ is compactly supported and denote by M the number of vanishing moments ($M \geq 1$). This corresponds to the following assumption on the filters h and g .

Assumption 1. *The filters h and g have support $\{-T+1, \dots, 0\}$ and g has M vanishing moments, that is,*

$$\sum_{t=-T+1}^0 g[t] t^\ell = 0 \quad \text{for all } \ell = 0, 1, \dots, M-1. \quad (4.5)$$

This assumption is supposed to hold throughout the chapter. It follows from the finite support assumption that a finite set of wavelet coefficients can be computed exactly from a finite sample X_0, \dots, X_{n-1} . In this case the pyramidal algorithm is applied with $k = 0, \dots, n_0 - 1$ in (4.2) and $k = 0, \dots, n_j - 1$ in (4.3) and (4.4), where

$$n_0 = n \quad \text{and} \quad n_j = \{[(n_{j-1} - T)/2] + 1\}_+, \quad j \geq 1, \quad (4.6)$$

and $\lfloor x \rfloor$ denotes the largest integer at most equal to x and x_+ the non-negative part of x , $\max(x, 0)$.

4.2.2 pyramidal algorithm based on approximation coefficients increments

The DWT (4.2–4.4) with input $\{X_k, 0 \leq k < n\}$, provides $\{W_{j,k}, j \geq 1, 0 \leq k < n_j\}$ as output. We now wish to provide the same output with $\{[\Delta^M X]\}_k, M \leq k < n$.

To this end we define the approximation coefficients increments at successive orders $\ell = 0, 1, \dots$ by $A_{j,k}^{(\ell)} = [\Delta^\ell A_{j,\cdot}]_k$, that is $A_{j,k}^{(0)} = A_{j,k}$ and for $\ell \geq 0$, $A_{j,k}^{(\ell+1)} = A_{j,k}^{(\ell)} - A_{j,k-1}^{(\ell)}$. Because X may not be stationary but becomes stationary after differencing sufficiently many times, it is a key step to derive a pyramidal algorithm in which $A_{j,k}$ in (4.2) and (4.3) is replaced by $A_{j,k}^{(\ell)}$ with ℓ the large enough. By assumption 1 we may take $\ell = M$. This new scheme is stated in Lemma 2.2.1.

4.3 Second order properties in the wavelet domain

4.3.1 Assumptions and notation

Let X be an $M(d)$ process with memory parameter d and with generalized spectral density f . By Definition 2.1.10 there exists K such that $\Delta^K X$ is second order stationary. By possibly increasing K to remove possibly remaining polynomial trends, we further assume that $\Delta^K X$ is centered. In fact, we will not use in this section that X is an $M(d)$ process but only the two following assumptions.

(H-i) The K^{th} order increment process $\Delta^K X$ is centered and covariance stationary.

(H-ii) Assumption 1 holds with $M \geq K$.

Let us now introduce some notation valid all along the chapter. For any scales i and j , we denote by $\mathbf{M}_{i,j} \stackrel{\text{def}}{=} (\mathbf{M}_{i,j}[k, k'], 0 \leq k \leq n_i - 1, 0 \leq k' \leq n_j - 1)$ the covariance matrix of the wavelet coefficients at scales i and j , respectively,

$$\mathbf{M}_{i,j}[k, k'] \stackrel{\text{def}}{=} \text{Cov}(W_{i,k}, W_{j,k'}). \quad (4.7)$$

Following Moulines et al. [2007b], we pool blocks of 2^u wavelet coefficients at scale $j - u$ within column vectors denoted by

$$\mathbf{W}_{j,k}(u) \stackrel{\text{def}}{=} [W_{j-u,2^u k}, W_{j-u,2^u k+1}, \dots, W_{j-u,2^u k+2^u-1}]^T, \quad (4.8)$$

and define the *between-scale* process at scale j , time k and scale difference u , as

$$\{[\mathbf{W}_{j,k}^T(u), W_{j,k}]^T\}_{k \in \mathbb{Z}}.$$

Observe that $\mathbf{W}_{j,k}(u)$ is a 2^u -dimensional vector of wavelet coefficients at scale $i = j - u$ and involves all possible translations of the position index $2^u k$ by $v = 0, 1, \dots, 2^u - 1$. The index u in (4.8) denotes the scale difference $j - i \geq 0$ between the finest scale i and the coarsest scale j . By convention $\mathbf{W}_{j,k}(0)$ ($u = 0$) coincides with the scalar $W_{j,k}$. By [Moulines et al., 2007b, Corollary 1], $\{[\mathbf{W}_{j,k}^T(u), W_{j,k}]^T\}_{k \in \mathbb{Z}}$ is then covariance stationary.

In contrast, the process of wavelet coefficient at the finer scale $j - u$ and the approximation coefficient at scale j $\{[\mathbf{W}_{j,k}^T(u), A_{j,k}]^T\}_{k \in \mathbb{Z}}$ may not be stationary. Nevertheless, it follows from (H-i) and (H-ii) above that $\{A_{0,k}^{(M)}, k \in \mathbb{Z}\}$ is centered and covariance stationary, hence we conclude from Lemma 2.2.1 that $\{[\mathbf{W}_{j,k}^T(u), A_{j,k}^{(M)}]^T\}_{k \in \mathbb{Z}}$ is centered and covariance stationary with second order properties described below using iterative formula. Before that we need some additional notation.

Let us define the covariance functions $\delta_{i,j}$, $\gamma_{i,j}$ and $\nu_{i,j}$ by

$$\begin{aligned} \gamma_{i,j}[k] &\stackrel{\text{def}}{=} \text{Cov}(W_{i,k}, A_{j,0}^{(M)}), & \delta_{i,j}[k] &\stackrel{\text{def}}{=} \text{Cov}(W_{i,k}, W_{j,0}), & 1 \leq i \leq j, \\ \nu_i[k] &\stackrel{\text{def}}{=} \text{Cov}(A_{i,k}^{(M)}, A_{i,0}^{(M)}), & & & 0 \leq i. \end{aligned}$$

When $i = j$, we denote $\gamma_{i,j} = \gamma_i$ and $\delta_{i,j} = \delta_i$. Using these definitions and the stationary structure in the wavelet coefficients, we obtain that

$$\mathbf{M}_{i,j}[k, k'] = \delta_{i,j}[k - 2^{j-i}k'], \quad 1 \leq i \leq j.$$

Similarly we introduce some notation for spectral densities. We denote by $\mathbf{D}_{j,u}$ the cross-spectral density function of the between-scale process $\{[\mathbf{W}_{j,k}^T(u), W_{j,k}]^T\}_{k \in \mathbb{Z}}$ and by $\mathbf{B}_{j,u}$ the cross-spectral density function of the process $\{[\mathbf{W}_{j,k}^T(u), A_{j,k}^{(M)}]^T\}_{k \in \mathbb{Z}}$. Hence, we have, for all $1 \leq i = j - u \leq j$,

$$\begin{aligned} [\delta_{i,j}[2^u k], \dots, \delta_{i,j}[2^u(k+1) - 1]]^T &= \text{Cov}(\mathbf{W}_{j,k}(u), W_{j,0}) = \int_{-\pi}^{\pi} \mathbf{D}_{j,u}(\lambda) e^{i\lambda k} d\lambda \quad (4.9) \\ [\gamma_{i,j}[2^u k], \dots, \gamma_{i,j}[2^u(k+1) - 1]]^T &= \text{Cov}(\mathbf{W}_{j,k}(u), A_{j,0}^{(M)}) = \int_{-\pi}^{\pi} \mathbf{B}_{j,u}(\lambda) e^{i\lambda k} d\lambda. \end{aligned} \quad (4.10)$$

By convention $\mathbf{D}_{j,0}$ ($u = 0$) denotes the spectral density of $\{W_{j,k}, k \in \mathbb{Z}\}$. Finally we let \mathbf{A}_j denote the spectral density function of the process $\{A_{j,k}^{(M)}\}_{k \in \mathbb{Z}}$. Hence, for all $j \geq 0$,

$$\nu_j[k] = \text{Cov}(A_{j,k}^{(M)}, A_{j,0}^{(M)}) = \int_{-\pi}^{\pi} \mathbf{A}_j(\lambda) e^{i\lambda k} d\lambda. \quad (4.11)$$

4.3.2 Iterative formula in spectral domain

Observe that \mathbf{A}_0 and ν_0 are the spectral density and autocovariance function of $\Delta^M X$, respectively. In particular, if f denotes the generalized spectral density of X ,

$$\mathbf{A}_0(\lambda) = |1 - e^{-i\lambda}|^{2M} f(\lambda) \text{ and } \nu_0(\tau) = \int_{-\pi}^{\pi} \mathbf{A}_0(\lambda) e^{i\lambda\tau} d\lambda$$

are easily derived for second order properties of X . We now wish to obtain the second order properties of the wavelet coefficients. This can be done either in the spectral domain, that is, by computing $\mathbf{D}_{j,u}$ for all $j \geq 1$ and $0 \leq u \leq j$, or in the time domain, that is by computing $\delta_{i,j}$ for all $1 \leq i \leq j$. Other spectral densities \mathbf{A}_j and $\mathbf{B}_{j,u}$, or covariances $\gamma_{i,j}$ and $\nu_{i,j}$ are key extra quantities to achieve this goal. The following results indeed provide simple iterative to compute all these quantities simply starting from \mathbf{A}_0 or ν_0 .

The first result provides iterative formula for the spectral densities \mathbf{A}_j , $\mathbf{D}_{j,u}(\cdot)$, and $\mathbf{B}_{j,u}(\cdot)$, $u \in \mathbb{Z}$, based on the filters \tilde{h} and \tilde{g} defined in (2.56) and (2.55).

Proposition 4.3.1. *The following formula hold for all $j \geq 1$, $0 \leq u \leq j - 1$ and $0 \leq v < 2^{u+1}$,*

$$\mathbf{A}_j(\lambda) = \frac{1}{2} \left\{ \mathbf{A}_{j-1} \left(\frac{\lambda}{2} \right) \left| \tilde{H} \left(\frac{\lambda}{2} \right) \right|^2 + \mathbf{A}_{j-1} \left(\frac{\lambda}{2} + \pi \right) \left| \tilde{H} \left(\frac{\lambda}{2} + \pi \right) \right|^2 \right\}, \quad (4.12)$$

$$\mathbf{D}_{j,0}(\lambda) = \frac{1}{2} \left\{ \mathbf{A}_{j-1} \left(\frac{\lambda}{2} \right) \left| \tilde{G} \left(\frac{\lambda}{2} \right) \right|^2 + \mathbf{A}_{j-1} \left(\frac{\lambda}{2} + \pi \right) \left| \tilde{G} \left(\frac{\lambda}{2} + \pi \right) \right|^2 \right\}, \quad (4.13)$$

$$\begin{aligned} \mathbf{B}_{j,0}(\lambda) = \frac{1}{2} \left\{ \mathbf{A}_{j-1} \left(\frac{\lambda}{2} \right) \tilde{G} \left(\frac{\lambda}{2} \right) \tilde{H}^* \left(\frac{\lambda}{2} \right) \right. \\ \left. + \mathbf{A}_{j-1} \left(\frac{\lambda}{2} + \pi \right) \tilde{G} \left(\frac{\lambda}{2} + \pi \right) \tilde{H}^* \left(\frac{\lambda}{2} + \pi \right) \right\}, \quad (4.14) \end{aligned}$$

$$\begin{aligned} \mathbf{B}_{j+u+1,u+1}^{(v)}(\lambda) = \frac{1}{2} \left[e^{i\lambda \langle \frac{v}{2} \rangle} \tilde{H}^* \left(\frac{\lambda}{2} \right) \mathbf{B}_{j+u,u}^{(\lfloor \frac{v}{2} \rfloor)} \left(\frac{\lambda}{2} \right) + \right. \\ \left. e^{i(\lambda+2\pi) \langle \frac{v}{2} \rangle} \tilde{H}^* \left(\frac{\lambda}{2} + \pi \right) \mathbf{B}_{j+u,u}^{(\lfloor \frac{v}{2} \rfloor)} \left(\frac{\lambda}{2} + \pi \right) \right], \quad (4.15) \end{aligned}$$

$$\begin{aligned} \mathbf{D}_{j+u+1,u+1}^{(v)}(\lambda) = \frac{1}{2} \left[e^{i\lambda \langle \frac{v}{2} \rangle} \tilde{G}^* \left(e^{i\frac{\lambda}{2}} \right) \mathbf{B}_{j+u,u}^{(\lfloor \frac{v}{2} \rfloor)} \left(\frac{\lambda}{2} \right) \right. \\ \left. + e^{i(\lambda+2\pi) \langle \frac{v}{2} \rangle} \tilde{G}^* \left(e^{i\frac{\lambda}{2} + \pi} \right) \mathbf{B}_{j+u,u}^{(\lfloor \frac{v}{2} \rfloor)} \left(\frac{\lambda}{2} + \pi \right) \right], \quad (4.16) \end{aligned}$$

where z^* denote the conjugate of z , $[x]$ denotes the integer part of x and $\langle x \rangle = x - [x]$ its fractional part, and $\mathbf{B}_{j,u}^{(v)}$ and $\mathbf{D}_{j,u}^{(v)}$ denote the $(v+1)$ -th entry of $\mathbf{B}_{j,u}$ and $\mathbf{D}_{j,u}$,

$$\mathbf{B}_{j,u}(\lambda) = \left[\mathbf{B}_{j,u}^{(v)}(\lambda) \right]_{v=0,\dots,2^u-1}^T \text{ and } \mathbf{D}_{j,u}(\lambda) = \left[\mathbf{D}_{j,u}^{(v)}(\lambda) \right]_{v=0,\dots,2^u-1}^T. \quad (4.17)$$

The proof of Proposition 4.3.1 is postponed in Section 4.5

Remark 4.3.1. All the quantities in Proposition 4.3.1 can be successively computed from \mathbf{A}_0 as follows. The spectral densities \mathbf{A}_j , $j \geq 1$, can be computed from \mathbf{A}_0 using the iterative formula (4.12). The cross-spectral densities $\mathbf{D}_{j,0}$ and $\mathbf{B}_{j,0}$, $j \geq 1$, can be computed from \mathbf{A}_j , $j \geq 0$, using (4.13) and (4.14). Finally, $\mathbf{B}_{j,u}$ and $\mathbf{D}_{j,u}$, $1 \leq j$, $0 \leq u \leq j - 1$ can be computed from $\mathbf{B}_{j,0}$, $j \geq 1$ using the iterative formula (4.15) and (4.16).

4.3.3 Iterative formula in temporal domain

The second result provides iterative formula for the covariance functions ν_i , $\delta_{i,j}$ and $\gamma_{i,j}$, $1 \leq i \leq j$, based on the filters \tilde{h} and \tilde{g} defined in (2.56) and (2.55).

Proposition 4.3.2. The following formula hold for all $k \in \mathbb{Z}$ and $1 \leq i \leq j$.

$$\nu_j[k] = \sum_{l,l' \in \mathbb{Z}} \tilde{h}[l]\tilde{h}[l']\nu_{j-1}[2k - l' + l] , \quad (4.18)$$

$$\delta_j[k] = \sum_{l,l' \in \mathbb{Z}} \tilde{g}[l]\tilde{g}[l']\nu_{j-1}[2k - l' + l] , \quad (4.19)$$

$$\gamma_j[k] = \sum_{l,l' \in \mathbb{Z}} \tilde{g}[l]\tilde{h}[l']\nu_{j-1}[2k - l + l'] , \quad (4.20)$$

$$\gamma_{i,j+1}[k] = \sum_{l \in \mathbb{Z}} \tilde{h}[l]\gamma_{i,j}[k + 2^{j-i}l] , \quad (4.21)$$

$$\delta_{i,j+1}[k] = \sum_{l \in \mathbb{Z}} \tilde{g}[l]\gamma_{i,j}[k + 2^{j-i}l] . \quad (4.22)$$

The proof of Proposition 4.3.2 in postponed in Section 4.5

Remark 4.3.2. All the quantities in Proposition 4.3.2 can be successively computed from ν_0 as follows. The autocovariance functions ν_j , $j \geq 1$, can be computed from ν_0 using the iterative formula (4.18). The covariance functions δ_j and γ_j can be computed from ν_j , $j \geq 0$, using (4.19) and (4.20). Finally $\gamma_{i,j}$ and $\delta_{i,j}$, $1 \leq i < j$ can be computed from γ_j , $j \geq 1$ using the iterative formula (4.21) and (4.22). Note also that the double summation and simple summation signs appearing in (4.18-4.22) only involve $(T - M)^2$ and $T - M$ non-vanishing terms, since \tilde{h} and \tilde{g} have lengths both equal to $T - M$.

4.4 Application

In this section we consider two applications for the fast computation algorithm of the covariance matrix of wavelet coefficients among scales. We use our algorithm to exhibit a semi parametric maximum likelihood estimation in the wavelet domain, where we allow values of the memory parameter d to be outside of the stationary range $(-\frac{1}{2}, \frac{1}{2})$. Then, we use our algorithm to compute a change point test statistic in the wavelet domain for long memory processes see Kouamo et al. [2010]. Finally, we apply this new procedure of test to detect multiple changepoints in the S&P 500 stock market absolute log returns.

4.4.1 Likelihood estimation

Let X_1, \dots, X_n be n observations of an $M(d)$ Gaussian process X and $W_{j,k}$, $j \geq 1$, $0 \leq k < n_j$ be the corresponding wavelet coefficients. Recall that n_j denotes the number of available wavelet coefficients at scale j from a sample of length n , and is defined by (4.6). We let J_n denote the corresponding maximal available scale index, $J_n = \max\{j : n_j \geq 1\}$. From (2.53), we see that $n2^{-j} - 2T \leq n_j \leq n2^{-j}$ and thus $J_n = \log(n)/\log(2) + O(1)$. It follows that the total number of available wavelet coefficients satisfies, as $n \rightarrow \infty$,

$$\tilde{n} = \sum_{j=1}^{J_n} n_j \sim n/2.$$

We rearrange the wavelet coefficients $W_{j,k}$ in a column vector by indexing (j, k) in a lexical order. Hence we set $\mathbf{W}^{(n)} = [W_i^{(n)}]^T$ with $i = 1, \dots, \tilde{n}$ such that, for all $1 \leq j \leq J_n$ and $0 \leq k < n_j$, $W_{j,k} = W_{i(j,k)}^{(n)}$ where $i(j, k) = \sum_{i=1}^{j-1} n_i + k$.

Since we are in a Gaussian context, and using Assumptions (H-i) and (H-ii), the negated log likelihood of $\mathbf{W}^{(n)}$ is given up to a constant additive term by

$$\mathcal{L}(\boldsymbol{\beta}) = \frac{1}{2} \left(\log \left(\det(\mathbf{M}^{(n)}(\boldsymbol{\beta})) \right) + \left(\mathbf{W}^{(n)} \right)^T \left(\mathbf{M}^{(n)}(\boldsymbol{\beta}) \right)^{-1} \mathbf{W}^{(n)} \right), \quad (4.23)$$

where $\mathbf{M}^{(n)}(\boldsymbol{\beta})$ denotes the covariance matrix of $\mathbf{W}^{(n)}$ under parameter $\boldsymbol{\beta}$, that is, $\mathbf{M}^{(n)}(\boldsymbol{\beta})$ have entries $\mathbf{M}_{i,j}[k, k']$ defined by (4.7) for $1 \leq j \leq J_n$ and $0 \leq k < n_j$, rearranged using the same indexing as $\mathbf{W}^{(n)}$. Moreover the covariance in (4.7) is computed using the distribution associated to the parameter $\boldsymbol{\beta}$. We call \mathcal{L} in (4.23) the negated log likelihood *in the wavelet domain*. For any parametric model containing a scale parameter, that is $\boldsymbol{\beta} = (\boldsymbol{\eta}, \sigma^2)$ and $\mathbf{M}^{(n)}(\boldsymbol{\beta}) = \sigma^2 \mathbf{M}^{(n)}(\boldsymbol{\eta})$ (with a slight abuse of notation), the log-likelihood (4.23) can be optimized in σ^2 as a function $\boldsymbol{\eta}$, giving rise to the scale maximal likelihood estimator

$$\hat{\sigma}^2(\boldsymbol{\eta}) = \underset{\sigma^2 > 0}{\text{Argmin}} \mathcal{L}(\boldsymbol{\eta}, \sigma^2) = \frac{1}{\tilde{n}} \left(\mathbf{W}^{(n)} \right)^T \left(\mathbf{M}^{(n)}(\boldsymbol{\eta}) \right)^{-1} \mathbf{W}^{(n)}. \quad (4.24)$$

In turn the maximal likelihood estimator of $\boldsymbol{\eta}$ is defined as the minimizer of the profile negated log likelihood

$$\mathcal{L}(\boldsymbol{\eta}) = \log \left(\left(\mathbf{W}^{(n)} \right)^T \left(\mathbf{M}^{(n)}(\boldsymbol{\eta}) \right)^{-1} \mathbf{W}^{(n)} \right) + \frac{1}{\tilde{n}} \log \left(\det \left(\mathbf{M}^{(n)}(\boldsymbol{\eta}) \right) \right)$$

(again with a slight abuse of notation).

A celebrated example of parametric long memory model is the ARFIMA(p, d, q) process. In the context of generalized Gaussian processes, it can be defined as the collection of Gaussian processes admitting the following generalized spectral density,

$$f(\lambda; \boldsymbol{\beta}) = |1 - e^{-i\lambda}|^{-2d} \frac{\sigma^2}{2\pi} \left| \frac{1 + \sum_{k=1}^q \theta_k e^{-i\lambda k}}{1 + \sum_{k=1}^p \phi_k e^{-i\lambda k}} \right|^2. \quad (4.25)$$

For identifiability requirements, $\phi(B) = I + \phi_1 B + \dots + \phi_p B^p$ and $\theta(B) = I + \theta_1 B + \dots + \theta_q B^q$ are assumed to have no common roots and to be causally invertible. This defines an open subset of \mathbb{R}^{p+q} for the parameter $(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$. The parameters d and σ^2 are allowed to be in \mathbb{R} and $(0, \infty)$, respectively.

The spectral density structure of an ARFIMA model is easier to handle than its covariance structure; thus we use spectral density defined in (4.25) to compute $\partial \nu_0[k] / \partial \beta_r$ according to

$$\frac{\partial \nu_0[k]}{\partial \beta_r} = \int_{-\pi}^{\pi} \frac{\partial f(\lambda; \boldsymbol{\beta}^{(-M)})}{\partial \beta_r} e^{i\lambda k} d\lambda = \int_{-\pi}^{\pi} \frac{\partial \log f(\lambda; \boldsymbol{\beta}^{(-M)})}{\partial \beta_r} e^{i\lambda k} f(\lambda; \boldsymbol{\beta}^{(-M)}) d\lambda. \quad (4.26)$$

where $\boldsymbol{\beta}^{(-M)}$ denotes the parameter $(d - M, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2)$ when $\boldsymbol{\beta} = (d, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2)$.

The integral (4.26) is not known in closed form, except for $p = q = 0$, see Gradshteyn and Ryzhik [2000], but, for M large enough so as to make $f(\lambda; \boldsymbol{\beta}^{(-M)})$ and its derivative with respect to $\boldsymbol{\beta}$ a smooth function of λ , can be approximated by Riemann sums with arbitrary precision.

4.4.2 Local likelihood wavelet (LLW) estimator of d

As shown in Moulines et al. [2007b], the wavelet coefficients of any $M(d)$ process have second order properties that only depend on d at large scales, up to a multiplicative constant. In particular, as $j \rightarrow \infty$, $\text{Var}(W_{j,0}) \sim \sigma^2 2^{2dj}$. The local Whittle Wavelet (LWW) estimator of d introduced in Moulines et al. [2008] is based on the approximation of nearly the wavelet coefficients by independent Gaussian coefficients with exact variance $\sigma^2 2^{2dj}$ over a given set of time scale indices \mathcal{I} , which yields the following *pseudo* negative log-likelihood

$$\widehat{\mathcal{L}}_{\mathcal{I}}(\sigma^2, d) = \frac{1}{2\sigma^2} \sum_{(j,k) \in \mathcal{I}} 2^{-2dj} (W_{j,k})^2 + \frac{|\mathcal{I}|}{2} \log(\sigma^2 2^{2\langle \mathcal{I} \rangle d}),$$

where $|\mathcal{I}|$ denotes the cardinal of \mathcal{I} and $\langle \mathcal{I} \rangle$ is the average scale, $\langle \mathcal{I} \rangle \stackrel{\text{def}}{=} |\mathcal{I}|^{-1} \sum_{(j,k) \in \mathcal{I}} j$. Define $\widehat{\sigma}_{\mathcal{I}}^2(d) \stackrel{\text{def}}{=} \text{Argmin}_{\sigma^2 > 0} \widehat{\mathcal{L}}_{\mathcal{I}}(\sigma^2, d) = |\mathcal{I}|^{-1} \sum_{(j,k) \in \mathcal{I}} 2^{-2dj} (W_{j,k})^2$. The pseudo maximum likelihood estimator of the memory parameter is then equal to the minimum of the negated profile log-likelihood,

$$\widehat{d}^{\text{LWW}}(\mathcal{I}) \stackrel{\text{def}}{=} \text{Argmin}_{d \in [\Delta_1, \Delta_2]} \widehat{\mathcal{L}}_{\mathcal{I}}(\widehat{\sigma}_{\mathcal{I}}^2(d), d) = \text{Argmin}_{d \in [\Delta_1, \Delta_2]} \widetilde{\mathcal{L}}_{\mathcal{I}}(d), \quad (4.27)$$

where $[\Delta_1, \Delta_2]$ is an interval of admissible values for d (that only depends on the wavelet) and

$$\widetilde{\mathcal{L}}_{\mathcal{I}}(d) \stackrel{\text{def}}{=} \log \left(\sum_{(j,k) \in \mathcal{I}} 2^{2d(\langle \mathcal{I} \rangle - j)} (W_{j,k})^2 \right). \quad (4.28)$$

Since the equivalence $\text{Var}[W_{j,0}] \sim \sigma^2 2^{2dj}$ holds at large scales ($j \rightarrow \infty$), the set of scale indices \mathcal{I} is chosen of the form

$$\mathcal{I}_n(\ell) = \{(j, k) : \ell \leq j \leq J_n, 0 \leq k < n_j\} .$$

(recall that n_j is the number of available wavelet coefficients at scale j defined in (2.53) and J_n the largest observed scale index). Henceforth we will simply denote the corresponding estimator as $\widehat{d}^{\text{LWW}}(\ell)$. The asymptotic properties (consistency and central limit theorem as $\ell, n \rightarrow \infty$) of the LWW estimators have been studied in Moulines et al. [2008] in the Gaussian case and in Roueff and Taqqu [2009] in the linear case, see also Faÿ et al. [2009] for comparison with other wavelet estimators and Fourier estimator of the memory parameter d .

However, as already mentioned, it is shown in Moulines et al. [2007b] that the wavelet coefficients of an $M(d)$ process cannot be approximated by independent coefficients, even at large scales. Instead, they have second order properties that are equivalent to those of the wavelet coefficients of a generalized continuous time fractional Brownian motion, up to a multiplicative constant, see Moulines et al. [2007b]. In particular, although the independence of wavelet coefficients is not verified at large scales, the second order properties keep depending only on the unknown parameter d and on a multiplicative constant. (In fact this constant only depends on the wavelet and $f^*(0)$). As a consequence, at large scales, the wavelet coefficients of any $M(d)$ process has a distribution well approximated by those of an ARFIMA(0, d , 0). Hence we propose to define a *local likelihood wavelet* (LLW) estimator by maximizing the likelihood associated to the wavelet coefficients of an ARFIMA(0, d , 0) with indices in $\mathcal{I}_n(\ell)$ for some lower scale ℓ . That is, we define $\widehat{d}^{\text{LLW}}(\ell)$ as the minimizer of

$$\mathcal{L}_\ell(d) = \log \left(\left(\mathbf{W}^{(\ell, n)} \right)^T \left(\mathbf{M}^{(\ell, n)}(d) \right)^{-1} \mathbf{W}^{(\ell, n)} \right) + \frac{1}{\tilde{n}(\ell)} \log \left(\det \left(\mathbf{M}^{(\ell, n)}(d) \right) \right) ,$$

where $\mathbf{W}^{(\ell, n)}$ contains all the available wavelet coefficients with scale indices between ℓ and J_n , $\tilde{n}(\ell)$ denotes the number of such coefficients and $\mathbf{M}^{(\ell, n)}(d)$ is the *exact* covariance matrix of these wavelet coefficients for the ARFIMA(0, d , 0) process. Note that this matrix can be computed using the iterative algorithm derived in Proposition 4.3.2.

Since the approximation of the second order properties is finer for the likelihood used for defining $\widehat{d}^{\text{LLW}}(\ell)$ than for the pseudo likelihood used for defining $\widehat{d}^{\text{LWW}}(\ell)$, one expects a better performance for the former estimator than for the second one. This is in fact only partially true. This better approximation yields an estimator based on a more complete information of the *asymptotic model* that appears at large scales, since not only the variance of the wavelet coefficients but also their cross-correlations are used. This should clearly yield a smaller variance of the estimator. On the other hand the bias introduced by the fact that the true model may not be an ARFIMA(0, d , 0) should not be improved :

both the variance and the cross-correlations are well approximated only for ℓ large. Indeed by approximating not only the variance but also the cross-correlations by an asymptotic model, it may happen that a larger bias is introduced. In other words, the likelihood could be more *model dependent* than the pseudo likelihood. These two phenomena are observed in our Monte-Carlo simulations, that we now present. We study the finite sample properties of the two estimators $\hat{d}^{\text{LWW}}(\ell)$ and $\hat{d}^{\text{LLW}}(\ell)$ for two different models.

1. Gaussian ARFIMA(1, d ,0) model with d in $\{-0.8, -0.4, 0, 0.2, 0.6, 1, 1.6, 2, 2.6\}$ and the AR coefficient equal to 0.7.
2. Gaussian DARFIMA model, as defined in Andrews and Sun [2004]. The spectral density of the DARFIMA(1, d ,0) process is equal to that of an ARFIMA(1, d ,0) process on the interval $[-\lambda_0, \lambda_0]$ and is zero for $(\lambda_0, \pi]$. It is obtained by low-pass filtering of an ARFIMA(1, d ,0) trajectory by a truncated *sinc* function in the time domain. We chose $\lambda_0 = \pi/2$ and the same parameters as previously for the ARFIMA(1, d ,0) part.

In all the simulations, we use Daubechies wavelets with 4 vanishing moments. In this study, we consider sample of length $n = 2^{12}$, which, with the chosen wavelet, gives $J_n = 8$.

		$n = 4096, J_n = 8$								
		ARFIMA(1, d , 0)					$\sigma^2 = 1$			
d		-0.8	-0.4	0	0.2	0.6	1	1.6	2	2.6
$\ell = 3$	Bias ^{LLW}	-0.057	-0.04	-0.03	-0.03	-0.02	-0.02	0.01	-0.01	-0.02
	S.E ^{LLW}	(0.026)	(0.029)	(0.026)	(0.026)	(0.027)	(0.029)	(0.025)	(0.027)	(0.024)
	RMSE ^{LLW}	0.063	0.051	0.043	0.0540	0.043	0.033	0.036	0.032	0.029
	Bias ^{LWW}	-0.029	-0.02	-0.02	-0.026	-0.02	-0.02	-0.01	-0.02	-0.03
	S.E ^{LWW}	(0.042)	(0.031)	(0.027)	(0.036)	(0.046)	(0.039)	(0.032)	(0.040)	(0.046)
	RMSE ^{LWW}	0.032	0.041	0.037	0.036	0.049	0.037	0.045	0.044	0.037
$\ell = 4$	Bias ^{LLW}	-0.006	-0.009	-0.01	-0.02	-0.01	-0.007	-0.009	-0.009	-0.004
	S.E ^{LLW}	(0.045)	(0.047)	(0.045)	(0.045)	(0.047)	(0.043)	(0.044)	(0.041)	(0.038)
	RMSE ^{LLW}	0.046	0.065	0.047	0.046	0.050	0.041	0.036	0.042	0.037
	Bias ^{LWW}	0.03	0.01	-0.004	-0.005	-0.01	-0.008	-0.009	-0.01	-0.08
	S.E ^{LWW}	(0.059)	(0.054)	(0.051)	(0.047)	(0.053)	(0.054)	(0.068)	(0.11)	(0.071)
	RMSE ^{LWW}	0.063	0.061	0.049	0.046	0.054	0.055	0.056	0.066	0.069
$\ell = 5$	Bias ^{LLW}	0.004	0.008	-0.006	-0.01	-0.009	-0.006	-0.01	-0.016	-0.01
	S.E ^{LLW}	(0.070)	(0.071)	(0.072)	(0.075)	(0.074)	(0.073)	(0.076)	(0.064)	(0.037)
	RMSE ^{LLW}	0.069	0.091	0.08	0.077	0.074	0.073	0.072	0.066	0.091
	Bias ^{LWW}	0.051	-0.009	-0.01	-0.007	-0.009	-0.006	-0.01	-0.03	-0.009
	S.E ^{LWW}	(0.091)	(0.097)	(0.09)	(0.082)	(0.088)	(0.092)	(0.11)	(0.10)	(0.07)
	RMSE ^{LWW}	0.1	0.091	0.092	0.083	0.087	0.095	0.106	0.12	0.085

Table 4.1: Bias, standard deviation and root mean-square error for $\hat{d}^{\text{LLW}}(L)$ and $\hat{d}^{\text{LWW}}(L)$ over 1000 replications for a time series generated from the Gaussian ARFIMA(1, d ,0) model. The lowest RMSE among the the method and all scales considered appears in boldface.

In view of Table 4.1, Table 4.2 and Figure 4.1, the two methods appear to work well with similar performances at the optimal finest scale index ℓ . An important property of \hat{d}^{LLW} is that its standard deviations appear to remain more stable than for the estimator \hat{d}^{LWW} as the unknown memory parameter d evolves for a fixed finest scale ℓ . This is of interest for computing confidence intervals: their size mainly depend on (the known) ℓ ,

		$n = 4096, J_n = 8$			DARFIMA(1, d , 0)		$\sigma^2 = 1$			
d		-0.8	-0.4	0	0.2	0.6	1	1.6	2	2.6
$\ell = 3$	Bias ^{LLW}	-0.02	-0.02	-0.009	-0.028	-0.023	-0.020	-0.021	-0.02	-0.02
	S.E ^{LLW}	(0.030)	(0.03)	(0.031)	(0.029)	(0.027)	(0.027)	(0.030)	(0.028)	(0.026)
	RMSE ^{LLW}	0.036	0.039	0.08	0.041	0.035	0.032	0.036	0.037	0.032
	Bias ^{LWW}	0.009	-0.01	0.004	-0.022	-0.02	-0.02	-0.02	-0.022	-0.01
	S.E ^{LWW}	(0.032)	(0.035)	(0.051)	(0.033)	(0.031)	(0.034)	(0.038)	(0.040)	(0.041)
	RMSE ^{LWW}	0.034	0.035	0.049	0.039	0.036	0.039	0.043	0.046	0.043
$\ell = 4$	Bias ^{LLW}	-0.01	-0.01	-0.01	-0.002	-0.01	-0.002	-0.006	-0.006	0.004
	S.E ^{LLW}	(0.04)	(0.042)	(0.043)	(0.047)	(0.047)	(0.037)	(0.041)	(0.041)	(0.038)
	RMSE ^{LLW}	0.055	0.044	0.034	0.047	0.049	0.037	0.041	0.041	0.036
	Bias ^{LWW}	0.02	0.007	-0.006	0.004	-0.007	-0.006	-0.007	-0.009	-0.008
	S.E ^{LWW}	(0.056)	(0.050)	(0.049)	(0.053)	(0.054)	(0.050)	(0.056)	(0.061)	(0.067)
	RMSE ^{LWW}	0.069	0.051	0.032	0.054	0.055	0.051	0.06	0.062	0.069
$\ell = 5$	Bias ^{LLW}	-0.008	-0.004	0.008	0.001	-0.002	-0.01	-0.01	-0.02	-0.01
	S.E ^{LLW}	(0.067)	(0.072)	(0.072)	(0.077)	(0.075)	(0.07)	(0.069)	(0.074)	(0.061)
	RMSE ^{LLW}	0.12	0.072	0.079	0.077	0.08	0.067	0.07	0.077	0.058
	Bias ^{LWW}	0.03	0.02	0.005	0.007	-0.002	-0.03	-0.02	-0.03	-0.03
	S.E ^{LWW}	(0.09)	(0.088)	(0.083)	(0.088)	(0.089)	(0.084)	(0.099)	(0.11)	(0.12)
	RMSE ^{LWW}	0.44	0.09	0.07	0.089	0.088	0.085	0.10	0.12	0.11

Table 4.2: Bias, standard deviation and root mean-square error for $\hat{d}^{LLW}(L)$ and $\hat{d}^{LWW}(L)$ over 1000 replications for a time series generated from the Gaussian DARFIMA(1, d , 0) model. The lowest RMSE among the two methods and all scales considered appears in boldface.

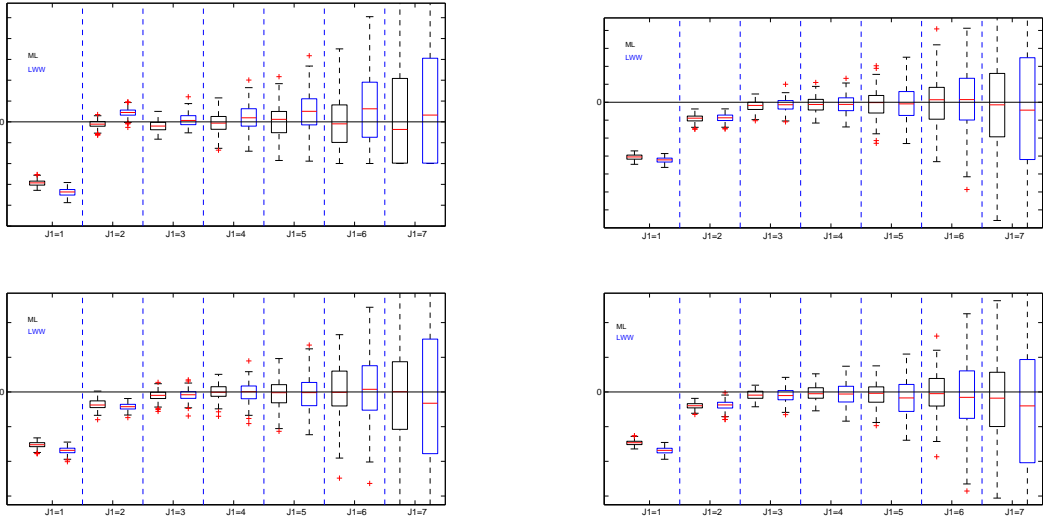


Figure 4.1: Monte Carlo simulation to compare $\hat{d}^{LLW}(\ell)$ and $\hat{d}^{LWW}(\ell)$ in a semi parametric setting. (Top left) plot is an ARFIMA(1,-0.8,0), (top right) is an ARFIMA(1,0.6,0), (bottom left) is an ARFIMA(1,1,0) and the (bottom right) is an ARFIMA(1,2,0).

not on (the unknown) d . Finally, in most cases, the RMSE is smaller for \hat{d}^{LLW} especially when $d > 1/2$ and $\ell \geq 4$. However, as we explained above, \hat{d}^{LWW} may enjoy a lower absolute bias, resulting in a smaller RMSE here for many small values of d when $\ell = 3$, in spite of a larger variance. As could be expected, this occurs for the smallest considered ℓ of the tables, since it corresponds to a larger bias and a smaller variance, that is, when the bias is more influent on the RMSE than the variance.

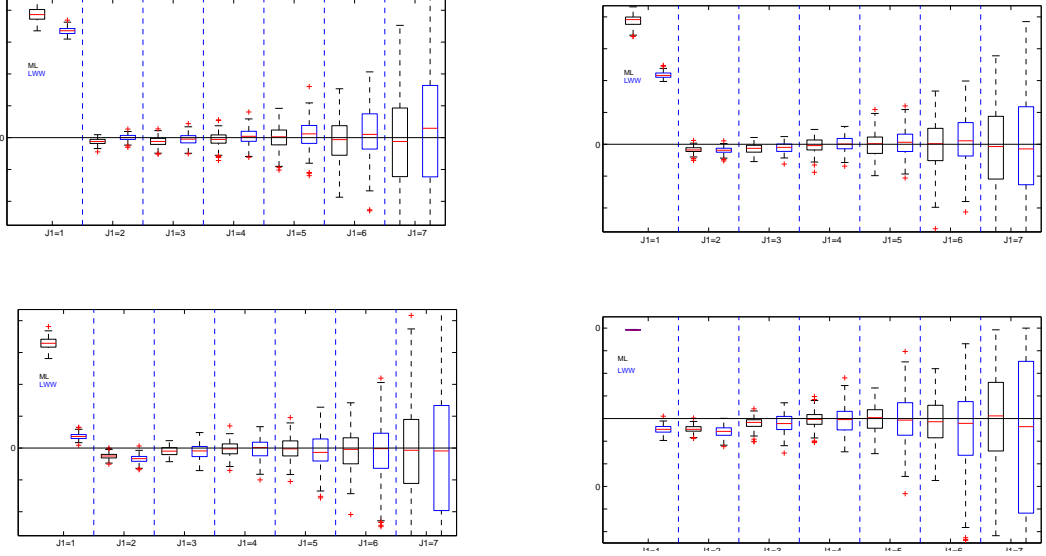


Figure 4.2: Monte Carlo simulation to compare $\hat{d}^{\text{LW}}(\ell)$ and $\hat{d}^{\text{LWW}}(\ell)$ in a semi parametric setting. (Top left) plot is a DARFIMA(1,-0.4,0), (top right) is a DARFIMA(1,0.2,0), (bottom left) is a DARFIMA(1,2,0) and the (bottom right) is a DARFIMA(1,2.6,0).

4.4.3 Change point test statistic.

Let X_1, \dots, X_n be n observations of a time series, and denote by $W_{j,k}$ for $(j,k) \in \mathcal{I}_n$ with \mathcal{I}_n defined in (2.53) the associated wavelet coefficients. If X is an $M(d)$ process and the number of vanishing moments M of the wavelet satisfies $M > d - 1/2$, then $\Delta^M X$ is weakly stationary and it follows that the wavelet variance at each given scale j should be constant. If this hypothesis is not fulfilled, then it can be expected that the wavelet variance will change either gradually or abruptly. This motivates a test for the constancy of the variance of the wavelet coefficients. If $\Delta^M X$ is weakly stationary but not centered, it becomes centered by replacing M by $M + 1$; hence, without loss of generality, the hypothesis that $\Delta^M X$ is weakly stationary can be replaced by the hypothesis that $\Delta^M X$ is weakly stationary and centered, which is considered as our *null hypothesis* in the following. Under the null hypothesis, the wavelet coefficients $\{W_{j,k}, k \in \mathbb{Z}\}$ is a centered and covariance stationary sequence. The cumulative sum (CUSUM) of squares used in Inclin and Tiao [1994] to detect change points in the variance can thus be adapted to obtain a test of the null hypothesis based on CUSUM of squared wavelet coefficients. This idea have been developed in Kouamo et al. [2010], where, moreover, changes in wavelet variances occurring simultaneously at multiple time-scales are considered, resulting in a *multiple scale* procedure. Let us briefly describe the procedure and the main theoretical result which it relies on (see Kouamo et al. [2010] for details). Given $J_1 \leq J_2$, define the

multi-scale scalogram by

$$Y_{J_1, J_2}[i] = \left[W_{J_2, i}^2, \sum_{u=1}^2 W_{J_2-1, 2(i-1)+u}^2, \dots, \sum_{u=1}^{2^{(J_2-J_1)}} W_{J_1, 2^{(J_2-J_1)}(i-1)+u}^2 \right]^T. \quad (4.29)$$

and denote the corresponding partial sum process by

$$S_{J_1, J_2}(t) = \frac{1}{\sqrt{n_{J_2}}} \sum_{i=0}^{\lfloor n_{J_2} t \rfloor} Y_{J_1, J_2}[i]. \quad (4.30)$$

Then for a wide class of Gaussian processes, under the null hypothesis and provided that the analyzing wavelet has M vanishing moments, it is shown that

$$\Gamma_{J_1, J_2}^{-1/2} (S_{J_1, J_2}(t) - \mathbb{E}[S_{J_1, J_2}(t)]) \xrightarrow{d} B(t) = (B_{J_1}(t), \dots, B_{J_2}(t)), \quad (4.31)$$

where the weak convergence holds in the space $D^{J_2-J_1+1}[0, 1]$ of càdlàg functions defined on $[0, 1]$ and valued in $\mathbb{R}^{J_2-J_1+1}$, $\{B_j(t), t \geq 0\}_{j=J_1, \dots, J_2}$ are independent *Brownian motions* and Γ_{J_1, J_2} denotes the asymptotic covariance matrix of $S_{J_1, J_2}(1)$. As classical in convergence results for càdlàg functions, $D^{J_2-J_1+1}[0, 1]$ is equipped with the Skorokhod metric.

Since the matrix Γ_{J_1, J_2} is unknown, it should be estimated. Denote $\widehat{\Gamma}_{J_1, J_2}$ a consistent estimator of Γ_{J_1, J_2} . The test is based on the statistics $T_{J_1, J_2} : [0, 1] \rightarrow \mathbb{R}_+$ defined by

$$T_{J_1, J_2}(t) \stackrel{\text{def}}{=} (S_{J_1, J_2}(t) - tS_{J_1, J_2}(1))^T \widehat{\Gamma}_{J_1, J_2}^{-1} (S_{J_1, J_2}(t) - tS_{J_1, J_2}(1)), \quad t \in [0, 1]. \quad (4.32)$$

As a consequence of (4.31), $T_{J_1, J_2}(t)$ also converges weakly in the Skorokhod space $D([0, 1])$,

$$T_{J_1, J_2}(t) \xrightarrow{d} \sum_{\ell=J_1}^{J_2} [B_\ell^0(t)]^2 \quad (4.33)$$

where $\{B_j^0(t), t \geq 0\}_{j=J_1, \dots, J_2}$ are independent Brownian bridges. For any continuous function $F : D[0, 1] \rightarrow \mathbb{R}$, the continuous mapping Theorem implies that

$$F[T_{J_1, J_2}(\cdot)] \xrightarrow{d} F \left[\sum_{\ell=1}^{J_2-J_1-1} [B_\ell^0(\cdot)]^2 \right].$$

We may for example apply either integral or sup functionals, resulting in Cramér-Von Mises or Kolmogorov-Smirnov statistics, respectively defined by

$$\text{CVM}(J_1, J_2) \stackrel{\text{def}}{=} \int_0^1 T_{J_1, J_2}(t) dt, \quad (4.34)$$

$$\text{KSM}(J_1, J_2) \stackrel{\text{def}}{=} \sup_{0 \leq t \leq 1} T_{J_1, J_2}(t). \quad (4.35)$$

We refer to Kouamo et al. [2010] for details on the computation of asymptotic quantiles for these statistics.

Our goal here is to provide new estimators of Γ_{J_1, J_2} . Because wavelet coefficients are correlated along and across scales, it has been suggested in Kouamo et al. [2010] to use the *Bartlett estimator* $\widehat{\Gamma}_{J_1, J_2}^B$ of the covariance matrix of the square wavelet's coefficients for scales $\{J_1, \dots, J_2\}$ which is given by (2.122). The Bartlett estimator corresponds to a nonparametric approach which avoids assuming a precise model under the null hypothesis. On the other hand, the choice of the window size q in the Bartlett estimator is difficult in practice, and may lead to very different estimates. Here we consider alternative estimators of Γ_{J_1, J_2} . The first is based on an estimator of the asymptotic covariance matrix of the GFBM square wavelet coefficients defined in (2.130) and the second is based on a pre-estimation of a semi-parametric model. In this second case, once the model is estimated, the estimator of Γ_{J_1, J_2} can be obtained by two different ways, based on 1) the auto-covariance and cross-covariance of wavelet coefficients among scales and 2) the with-in and between scale spectral density. Indexing $\Gamma_{J_1, J_2}[i, j]$ by scale indices i and j , we have, for $J_1 \leq i \leq j \leq J_2$,

$$\begin{aligned} \Gamma_{J_1, J_2}[i, j] &= \sum_{\tau \in \mathbb{Z}} \text{Cov} \left(\sum_{l=0}^{2^{J_2-j}-1} W_{j,l}^2, \sum_{u=0}^{2^{J_2-i}-1} W_{i, 2^{J_2-i}\tau+u}^2 \right) \\ &= 2 \sum_{\tau \in \mathbb{Z}} \sum_{l=0}^{2^{J_2-j}-1} \sum_{u=0}^{2^{J_2-i}-1} \text{Cov}^2(W_{j,l}, W_{i, 2^{J_2-i}\tau+u}) \\ &= 2^{J_2-j+1} \sum_{s \in \mathbb{Z}} \delta_{i,j}^2[s]. \end{aligned}$$

This expression is useful when the covariance of the estimated model is easy to compute, yielding a fast computation of $\delta_{i,j}$ by Proposition 4.3.1. However only an approximation of the infinite sum above can be computing by truncation. The computation can also be done in the frequency domain, using the Parseval inequality,

$$\begin{aligned} 2^{J_2-j+1} \sum_{s \in \mathbb{Z}} \delta_{i,j}^2[s] &= 2^{J_2-j+1} \sum_{t \in \mathbb{Z}} \sum_{u=0}^{2^{j-i}-1} \delta_{i,j}^2[2^{j-i}t + u] \\ &= \frac{2^{J_2-j}}{\pi} \int_{-\pi}^{\pi} \|\mathbf{D}_{j,u}(\lambda)\|^2 d\lambda. \end{aligned}$$

Again the obtained formula cannot be computed exactly but can be approximated to an arbitrary level of accuracy approximating the integral $\int_{-\pi}^{\pi} \|\mathbf{D}_{j,j-i}(\lambda; f)\|^2 d\lambda$ with a Riemann sum.

Alternatively, once a specific model has been estimated, it is legitimate to compute

the *non-asymptotic* covariance matrix of $S_{J_1, J_2}(1)$ in (4.30), given by

$$\begin{aligned}\tilde{\Gamma}_{J_1, J_2}[i, j] &= \frac{1}{n_{J_2}} \sum_{k, k'=0}^{n_{J_2}-1} \text{Cov} \left(\sum_{l=0}^{2^{J_2-j}-1} W_{j, 2^{J_2-j}k+l}^2, \sum_{u=0}^{2^{J_2-i}-1} W_{i, 2^{J_2-i}k'+u}^2 \right) \\ &= \frac{2}{n_{J_2}} \sum_{s, s'=0}^{2^{J_2-j}n_{J_2}-1} \sum_{v=0}^{2^{j-i}-1} \text{Cov}^2(W_{j, s}, W_{i, s'2^{j-i}+v}),\end{aligned}$$

where we set $s = 2^{J_2-j}k + l$ and $2^{j-i}s' + v = 2^{J_2-i}k' + u$, and relied on a Gaussian assumption on the wavelet coefficients. The latter expression can be expressed using the covariance function $\delta_{i, j}$, namely,

$$\tilde{\Gamma}_{J_1, J_2}[i, j] = 2 \sum_{\tau=-2^{J_2-j}n_{J_2}+1}^{2^{J_2-j}n_{J_2}-1} (2^{J_2-j} - |\tau|/n_{J_2}) \sum_{v=0}^{2^{j-i}-1} \delta_{i, j}[2^{j-i}\tau + v]; \quad (4.36)$$

Observe that this covariance matrix can be computed exactly since it involves a finite number of covariances $\delta_{i, j}[t]$.

Having this expression at hand, suppose now that we are interested in changes of the spectral density at *large scales*, or, equivalently, at *small frequencies*, that is, take $J_2 \geq J_1$ large enough. Note however that we do not take $J_2 = J_n$ here since the Gaussian approximation (4.31) is valid as $n_{J_2} \sim n2^{-J_2} \rightarrow \infty$ and $n_{J_n} = O(1)$ by definition. Hence J_1, J_2 and n_{J_2} have to be large simultaneously. Not surprisingly, detecting changes in low frequencies, that is, when 2^{J_2} large, obviously requires long times series, that is, n has to be even larger. We propose to rely on the approximation of the second order properties of the wavelet coefficients of $M(d)$ processes at large scales that has been already used for semiparametric wavelet estimation of d in Section 4.4.1. More precisely an alternative to the Bartlett estimator is obtained through the following steps.

Step 1. Estimate the memory parameter d using $\hat{d} = \hat{d}^{\text{LLW}}(J_1, J_2)$, that is the same estimator as $\hat{d}^{\text{LLW}}(\ell)$ but using the scale indices $J_1 \leq j \leq J_2$ instead of $\ell \leq j \leq J_n$. We also estimate the corresponding scaling parameter σ^2 of the ARFIMA(0, d , 0) model as in (4.24).

Step 2. Compute the auto-covariance $\hat{\nu}_0[k]$, of the ARFIMA(0, $\hat{d} - M, 0$) model, with the scaling parameter given by the estimated one. From this, compute the auto-covariance $\hat{\delta}_j[k]$ for a scale j and the cross-covariance $\hat{\delta}_{i, j}[k]$ of the wavelet coefficients for two given scales i and j included in $\{J_1, \dots, J_2\}$ as in Proposition 4.3.2.

Step 3. Finally estimate $\tilde{\Gamma}_{J_1, J_2}$ by $\hat{\Gamma}_{J_1, J_2}^{\text{LLW}}$ by replacing $\delta_{i, j}[k]$ in (4.36) by their estimates $\hat{\delta}_{i, j}[k]$, that is, for two scales i, j with $J_1 \leq i \leq j \leq J_2$, we set

$$\hat{\Gamma}_{J_1, J_2}^{\text{LLW}}[i, j] = 2 \sum_{\tau=-2^{J_2-j}n_{J_2}+1}^{2^{J_2-j}n_{J_2}-1} (2^{J_2-j} - |\tau|/n_{J_2}) \sum_{v=0}^{2^{j-i}-1} \hat{\delta}_{i, j}[2^{j-i}\tau + v].$$

We denote by $\text{KSM}^{\text{B}}(J_1, J_2)$, $\text{KSM}^{(d)}(J_1, J_2)$ and $\text{KSM}^{\text{LLW}}(J_1, J_2)$ the test statistics $\text{KSM}(J_1, J_2)$ normalized by $\widehat{\Gamma}_{J_1, J_2}^{\text{B}}$, $\widehat{\Gamma}_{J_1, J_2}^{(d)}$ and $\widehat{\Gamma}_{J_1, J_2}^{\text{LLW}}$ respectively. The test statistics $\text{CVM}^{\text{B}}(J_1, J_2)$, $\text{CVM}^{(d)}(J_1, J_2)$ and $\text{CVM}^{\text{LLW}}(J_1, J_2)$ are defined accordingly for the CVM statistic. Each corresponding test rejects the null hypothesis if the statistic exceeds the $(1 - \alpha)$ -th quantile of the corresponding asymptotic distributions. We now investigate the finite sample false alarm probabilities of the SIX resulting KSM's and CVM's tests when the level is set to $\alpha = 0.05$, when the observed signal is the same ARFIMA(1, d , 0) as in Section 4.4.2, analyzed with the same Daubechies wavelets with 4 vanishing moments. Here, we consider series with sample length $n = 15000$, which, with the chosen wavelet, gives $J_n = 10$. We set the coarsest scale to $J_2 = 8$ and vary the finest scale J_1 from 4 to 6.

The results, obtained through Monte Carlo simulations with 1000 independent replications, are displayed in Table 4.3. It appears that the normalization of the test by $\widehat{\Gamma}_{J_1, J_2}^{\text{LLW}}$ yields more accurate levels compared to the normalization by $\widehat{\Gamma}_{J_1, J_2}^{\text{B}}$ and $\widehat{\Gamma}_{J_1, J_2}^{(d)}$. In particular the accuracy deteriorates more significantly when the process is not stationary ($d > 1/2$) for CVM^{B} , KSM^{B} , $\text{CVM}^{(d)}$ and $\text{KSM}^{(d)}$ see Table 4.3. However, when $d = 2, 2.6$, the empirical levels are much higher to the asymptotic one in all cases. We also notice that in general the empirical levels of CVM are more accurate than the ones for KSM test.

		$J_2 = 8$	$n = 15000$ ARFIMA(1, d , 0)							
		d	-0.8	0	0.2	0.6	1	1.6	2	2.6
$J_1 = 4$	KSM^{LLW}		0.057	0.072	0.046	0.09	0.087	0.12	0.26	0.334
	$\text{KSM}^{(d)}$		0.07	0.086	0.063	0.11	0.24	0.26	0.33	0.54
	KSM^{B}		0.083	0.10	0.22	0.42	0.385	0.52	0.58	0.69
	CVM^{LLW}		0.03	0.045	0.061	0.053	0.043	0.064	0.091	0.14
	$\text{CVM}^{(d)}$		0.053	0.056	0.076	0.098	0.109	0.196	0.23	0.42
	CVM^{B}		0.031	0.068	0.091	0.15	0.29	0.25	0.31	0.49
$J_1 = 5$	KSM^{LLW}		0.079	0.08	0.061	0.13	0.112	0.17	0.253	0.613
	$\text{KSM}^{(d)}$		0.09	0.078	0.089	0.21	0.27	0.25	0.45	0.71
	KSM^{B}		0.13	0.18	0.23	0.26	0.31	0.39	0.43	0.75
	CVM^{LLW}		0.056	0.049	0.039	0.06	0.087	0.085	0.12	0.26
	$\text{CVM}^{(d)}$		0.051	0.067	0.064	0.106	0.16	0.197	0.23	0.61
	CVM^{B}		0.091	0.097	0.094	0.15	0.21	0.26	0.31	0.63
$J_1 = 6$	KSM^{LLW}		0.08	0.073	0.091	0.16	0.157	0.23	0.218	0.584
	$\text{KSM}^{(d)}$		0.067	0.089	0.18	0.27	0.31	0.31	0.381	0.786
	KSM^{B}		0.108	0.123	0.328	0.321	0.519	0.61	0.716	0.857
	CVM^{LLW}		0.06	0.071	0.062	0.078	0.067	0.096	0.108	0.247
	$\text{CVM}^{(d)}$		0.078	0.068	0.091	0.11	0.097	0.18	0.23	0.56
	CVM^{B}		0.101	0.11	0.172	0.328	0.303	0.411	0.515	0.702

Table 4.3: Empirical level of KSM – CVM on 15000 observations of different ARFIMA(1, d , 0) processes using two estimators of the scalogram covariance matrix.

4.4.4 Large scale multiple change detection for stock market data

Let us apply the above change point detection method to the S&P daily 500 stock market absolute log returns from 03-01-1950 to 27-10-2010, namely 15303 observations. The dataset is available from website <http://finance.yahoo.com/>. Such analysis is motivated

by previous works where long memory of such data were exhibited, see Ding et al. [1993] and Granger and Ding [1996]. Moreover, in [Granger and Ding, 1996, Section 5], it is claimed that the long memory parameter may change along time, and that such changes may correspond to economic structural changes. Their claims are illustrated in their Figure 6, which indicates that the long memory parameter started increasing around 1960, stabilized around 1979 and then decreased again until the late 1980's. Their study relies on estimations of this parameter over successive 10 periods from 1928 to 1991, each of them being 6 years long. The daily returns that we analyzed are displayed in Figure 4.3 along with the autocorrelation of their absolute values. The largest absolute return appears on the Black Monday stock market crash in 1987. It is well known since Ding et al. [1993], that although the returns themselves contain little serial correlation, their absolute values exhibit significantly positive serial correlation, which is confirmed on the right plot of Figure 4.3, where the blue dotted lines represent the 95% confidence interval for the autocorrelations of an i.i.d. process.

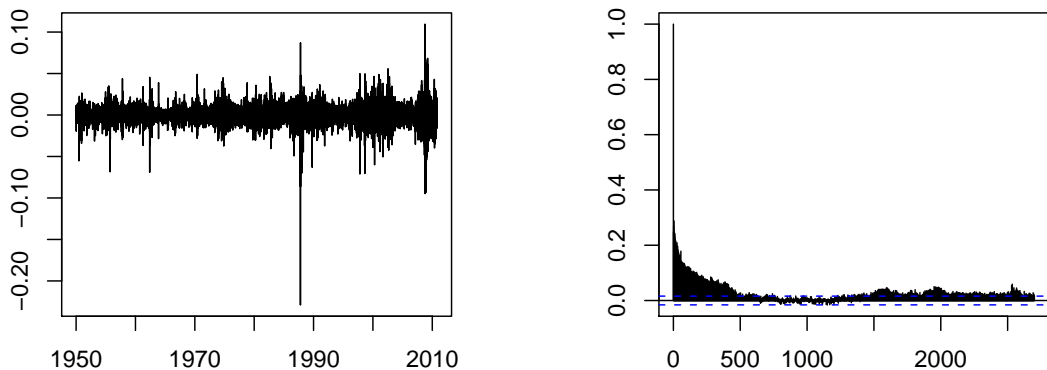


Figure 4.3: Standard and Poor daily log returns r_t 03/01/1950-27/10/2010 (left) and Sample autocorrelation of $|r_t|$ (right).

The test procedures CVM^{LLW} and CVM^{B} are now applied to detect changes in low frequencies of the S&P 500 absolute returns between 1950 and 2010. We set the coarsest scale $J_2 = 7$, hence we are able to detect changes every $2^7 = 128$ working days, that is, around every 6 months. We vary the finest scale from $J_1 = 3$ to $J_1 = 6$.

The results are displayed in Table 4.4. Note that for $J_1 = 3, 4, 5$, the statistics CVM^{B} and CVM^{LLW} exceed each their corresponding asymptotic 95% quantile which are respectively 1.0014, 1.297, and 1.4491 see Kouamo et al. [2010]. This confirms the presence of a change point. The corresponding change point, that maximizes $T_{J_1, J_2}(k/n_{J_2})$ is also displayed in the data. In view of Table 4.4, we would choose the change point around

J_1	3	4	5	6
CVM ^{LLW}	1.89	1.54	1.27	0.74
Change ^{LLW}	9527	9463	9479	9771
CVM ^B	2.01	1.62	1.39	0.79
Change ^B	9527	9402	9503	10270

Table 4.4: Change point in the S&P 500 absolute returns .

9500 since for $J_1 = 3 : 5$ the changes are around these value for the two statistics CVM^B and CVM^{LLW}. This ensures us that the change occurs in year 1987, which correspond to the market crash.

The preceding approach assumes at most one possible change point in the time series. If one is interested in *multiple change points detection*, the Iterated Cumulative Sums of Squares (ICSS) Algorithm proposed by Inclan and Tiao [1994] is easily adapted to our context. Let us denote by $\tilde{T}_{J_1, J_2}^{k_1: k_2}(k)$ the statistic $T_{J_1, J_2}(t)$ based on the observations $Y_{J_1, J_2}[k_1 : k_2] = \{Y_{J_1, J_2}(k), k_1 \leq k < k_2\}$ corresponding to rescaled time $t = 0, \dots, 1$. In other words, we set, for $k_1 \leq k \leq k_2$,

$$\tilde{T}_{J_1, J_2}^{k_1: k_2}(k) = \left(\tilde{S}_{J_1, J_2}(k) - \frac{k - k_1}{k_2 - k_1} \tilde{S}_{J_1, J_2}(k_2) \right)^T \left(\hat{\Gamma}_{J_1, J_2}^{k_1: k_2} \right)^{-1} \left(\tilde{S}_{J_1, J_2}(k) - \frac{k - k_1}{k_2 - k_1} \tilde{S}_{J_1, J_2}(k_2) \right),$$

where $\hat{\Gamma}_{J_1, J_2}^{k_1: k_2}$ is the estimator $\hat{\Gamma}_{J_1, J_2}^{\text{LLW}}$ based on $Y_{J_1, J_2}[k_1 : k_2]$ and

$$\tilde{S}_{J_1, J_2}(k) = \frac{1}{\sqrt{k_2 - k_1}} \sum_{i=k_1}^k Y_{J_1, J_2}[i].$$

Finally we denote by CVM $_{J_1, J_2}^{k_1: k_2}$ the corresponding test statistic,

$$\text{CVM}_{J_1, J_2}^{k_1: k_2} = \frac{1}{k_2 - k_1} \sum_{k=k_1}^{k_2-1} \tilde{T}_{J_1, J_2}^{k_1: k_2}(k).$$

The *multiple change points detections algorithm* of [Inclan and Tiao, 1994, Section 3] can be readily applied to $\tilde{T}_{J_1, J_2}^{k_1: k_2}(k)$ with the appropriate quantiles. To this end we denote by \mathbf{c} the asymptotic quantile associated to probability 0.95. We quote this algorithm hereafter for convenience.

Step 0 Let $k_1 = 1$

Step 1 Calculate $\tilde{T}_{J_1, J_2}^{k_1: n_{J_2}}(k)$ for $k_1 \leq k < n_{J_2}$. Let $k_{k_1: n_{J_2}}^*$ be the maximizing point of this sequence. If $\text{CVM}_{J_1, J_2}^{k_1: n_{J_2}} > \mathbf{c}$, consider that there is change point at $k_{k_1: n_{J_2}}^*$ and proceed to Step 2a. Otherwise, there is no evidence of variance changes in the series. The algorithm stops.

- Step 2a** Let $k_2 = k_{k_1:n_{J_2}}^*$. Evaluate $\tilde{T}_{J_1, J_2}^{k_1:k_2}(k)$ for $k_1 \leq k < k_2$. If $\text{CVM}_{J_1, J_2}^{k_1:k_2} > \mathbf{c}$, then we have a new point of change and should repeat Step 2a until $\text{CVM}_{J_1, J_2}^{k_1:k_2} \leq \mathbf{c}$. When this occurs, we can say that there is no evidence change in $k = k_1, \dots, k_2$ and, therefore, the first change point is then $k_{\text{first}} = k_2$.
- Step 2b** Now do a similar search starting from the first change point found in step 1 toward the end of the series. Define a new value for k_1 : let $k_1 = k_{k_1:n_{J_2}}^*$. Evaluate $\tilde{T}_{J_1, J_2}^{k_1:k_2}(k)$ for $k_1 \leq k < k_2$ and repeat step 2b until $\text{CVM}_{J_1, J_2}^{k_1:k_2} \leq \mathbf{c}$. Then let $k_{\text{last}} = k_1 - 1$
- Step 2c** If $k_{\text{first}} = k_{\text{last}}$, there is just one change point. The algorithm stops there. If $k_{\text{first}} < k_{\text{last}}$, keep both values as possible change points and repeat Step 1 and Step 2 on the middle part of the series; that is, $k_1 = k_{\text{first}}$ and $k_2 = k_{\text{last}}$. Each time that Steps 2a and 2b are repeated, the result can be one or two more points. Call N the number of change points found so far.
- Step 3** If there are two or more possible change points, make sure they are in increasing order. Let cp be the vector of all the possible changepoints found so far. Define the two extreme values $cp_0 = 0$ and $cp_{N+1} = n_{J_2}$. Check each possible change point by calculating $\tilde{T}_{J_1, J_2}^{cp_{j-1}:cp_{j+1}}$ $j = 1, \dots, N$. If $\text{CVM}_{J_1, J_2}^{cp_{j-1}:cp_{j+1}} > \mathbf{c}$, then cp_j will be replaced by the corresponding change point; otherwise eliminate it. Repeat Step 3 until the number of change points does not change and the points found in each new pass are "close" to those on the previous pass. We consider that if each cp_j has not increase nor decrease by more than one from the previous iteration, then the algorithm has converged.

Applying this algorithm, we obtain a (possibly empty) set of changepoints cp_1, \dots, cp_N corresponding to indices at scale J_2 ; hence the corresponding time indices are $cp_1 2^{J_2}, \dots, cp_N 2^{J_2}$. At the end of Step 2 the obtained change points correspond to the following dates: Dec-1973, Dec-1987, Nov-1990, Aug-1997, Mar-2001, Dec-2009. Up to Step 2, the procedure performs a systematic search for possible points of change within previously found change points. But newly discovered change points may make past ones less relevant. Step 3 provides a global check of all change points. After applying this step, the following dates are finally obtained: Mar-1973, Sept-1987, Sept-1996, Nov-2001, May-2009. It is interesting to note that all these dates are close to major events that implies economic structural changes, namely the oil crisis (October 1973), "black Monday" (October 1987), Asian crisis (July 1997), the collapse of the twin towers of the World Trade Center in New-york (September 2001), bankrupt of Lehman brothers (September 2008). In Figure 4.4(b), the change points are displayed using vertical red dashed lines on the absolute returns time series. We observe that over the whole period (1950–2010), the change dates are getting closer, which can be interpreted as an increasing instability of the market system.

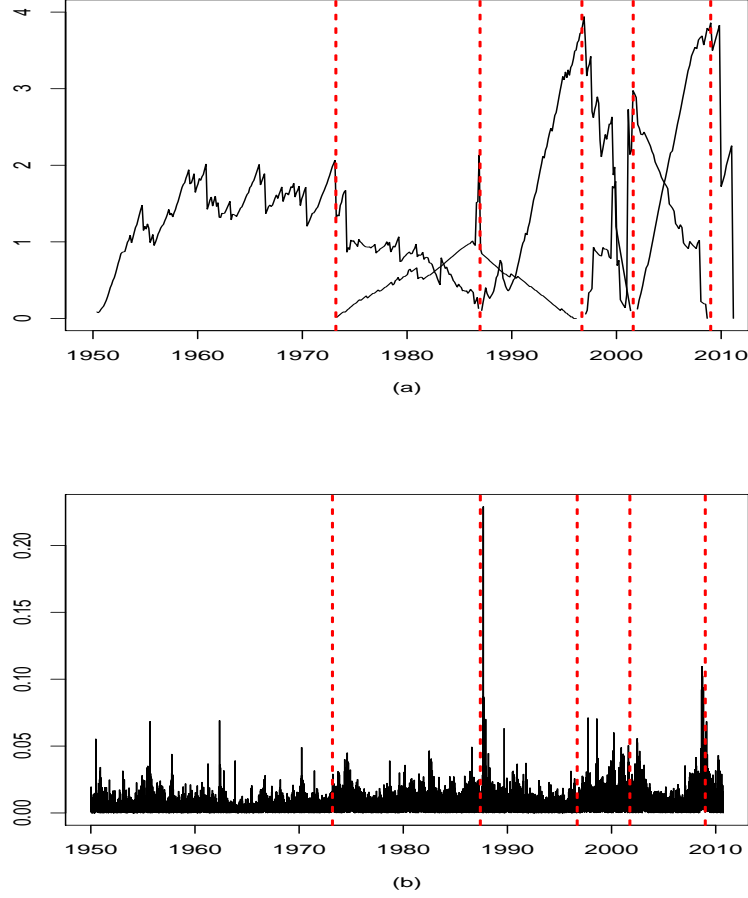


Figure 4.4: Figure (a) indicates the statistic $\tilde{T}_{J_1, J_2}^{k_1: k_2}(k)$ based on time index. The red dotted lines where $\tilde{T}_{J_1, J_2}^{k_1: k_2}(k)$ is maximal indicate years of changes. Figure (b) corresponds to the change points in the absolute log returns S&P 500 $|r_t|$.

4.5 Postponed proofs

Recall that B denotes the backward shift operator defined for any sequence $x \stackrel{\text{def}}{=} (x_k)_{k \in \mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}}$ by $[Bx]_k = x_{k-1}$ for any $k \in \mathbb{Z}$ and \downarrow^m denotes the down-sampling operator by a factor of m that is $\downarrow^m x = (x_{km})_{k \in \mathbb{Z}}$.

Proof of Lemma (2.2.1). We recall the first Noble Identity, see [Strang and Nguyen, 1997, Chapter 3],

$$\left(\sum_{k \in \mathbb{Z}} \alpha[k] B^k \right) \circ \downarrow^m = \downarrow^m \circ \left(\sum_{k \in \mathbb{Z}} \alpha[k] B^{km} \right), \quad (4.37)$$

where $\{\alpha[k], k \in \mathbb{Z}\}$ is arbitrary and \circ denotes the composition of operators, $A \circ B(x) = A(B(x))$. Applying (4.37) M times successively, we get

$$(I - B)^M \circ \downarrow^2 = \downarrow^2 \circ (I - B^2)^M = \downarrow^2 \circ (I + B)^M \circ (I - B)^M.$$

Applying $\Delta^M = (I - B)^M$ to both sides of (2.50) and using the definition of $A_{j,k}^{(M)}$ thus yields

$$A_{j,k}^{(M)} = \left[\downarrow^2 \circ (I + B)^M (h \star A_{j-1, \cdot}^{(M)}) \right]_k = \left[\downarrow^2 (\tilde{h} \star A_{j-1, \cdot}^{(M)}) \right]_k ,$$

by definition of \tilde{h} in (2.56). That is, we obtain (2.59). As for (2.60), it is a direct consequence of (2.51), (2.55) and the definition of $A_{j,k}^{(M)}$. \square

We preface the proofs of Propositions 4.3.1 and 4.3.2 by two lemmas on the effect of down-sampling on second order properties.

Lemma 4.5.1. *Let $m \geq 1$, $r \in \mathbb{Z}$ and $\{[Y_k, X_k]^T\}_{k \in \mathbb{Z}}$ be a weakly stationary $\mathbb{R} \times \mathbb{R}$ valued process with cross-spectral density f , with the convention for all $k, \tau \in \mathbb{Z}$,*

$$\text{Cov}(Y_{k+\tau}, X_k) = \int_{-\pi}^{\pi} f(\lambda) e^{i\tau\lambda} d\lambda .$$

Then for all $r = 0, \dots, m-1$, the process $[\downarrow^m \circ B^{-r} \circ Y, \downarrow^m \circ X]^T$ is weakly stationary with cross-spectral density given by

$$f_{m,r}(\lambda) = \frac{1}{m} \sum_{j=0}^{m-1} e^{i\left(\frac{\lambda+2\pi j}{m}\right)r} f\left(\frac{\lambda+2\pi j}{m}\right) , \quad (4.38)$$

that is, for all $k, \tau \in \mathbb{Z}$,

$$\text{Cov}(Y_{m(k+\tau)+r}, X_{mk}) = \int_{-\pi}^{\pi} f_{m,r}(\lambda) e^{i\lambda\tau} d\lambda . \quad (4.39)$$

Proof of Lemma 4.5.1. By definition of f we have, for all $k, k' \in \mathbb{Z}$,

$$\text{Cov}(Y_{mk'+r}, X_{mk}) = \int_0^{2\pi} e^{i\lambda m(k'-k)} e^{i\lambda r} f(\lambda) d\lambda .$$

By the change of variable $\xi = m\lambda$, we get

$$\begin{aligned} \text{Cov}(Y_{mk'+r}, X_{mk}) &= \frac{1}{m} \int_0^{2\pi m} e^{i\xi(k'-k)} e^{i\frac{\xi}{m}r} f\left(\frac{\xi}{m}\right) d\xi \\ &= \frac{1}{m} \sum_{l=0}^{m-1} \int_{2\pi l}^{2\pi(l+1)} e^{i\xi(k'-k)} e^{i\frac{\xi}{m}r} f\left(\frac{\xi}{m}\right) d\xi \\ &= \frac{1}{m} \int_0^{2\pi} e^{i\xi(k'-k)} \left(\sum_{l=0}^{m-1} e^{i\left(\frac{\xi+2\pi l}{m}\right)r} f\left(\frac{\xi+2\pi l}{m}\right) \right) d\xi . \end{aligned}$$

The function between parentheses is (2π) -periodic because f is. Note indeed that

$$\begin{aligned} \sum_{l=0}^{m-1} e^{i\left(\frac{\xi+2\pi(l+1)}{m}\right)r} f\left(\frac{\xi+2\pi(l+1)}{m}\right) &= \sum_{l=1}^{m-1} e^{i\left(\frac{\xi+2\pi l}{m}\right)r} f\left(\frac{\xi+2\pi l}{m}\right) + e^{i\frac{\xi}{m}r} f\left(\frac{\xi}{m}\right) \\ &= \sum_{l=0}^{m-1} e^{i\left(\frac{\xi+2\pi l}{m}\right)r} f\left(\frac{\xi+2\pi l}{m}\right) . \end{aligned}$$

Hence the last integral $\int_0^{2\pi}$ can be replaced by $\int_{-\pi}^{\pi}$ which gives (4.38). \square

Lemma 4.5.2. *Let p be a positive integer and a be a filter with FIR $\{a[k]\}_{k \in \mathbb{Z}}$. Consider a weakly stationary $\mathbb{R}^p \times \mathbb{R}$ valued process $\left\{ \left[\underline{\mathbf{Y}}_k^T, X_k \right]^T \right\}_{k \in \mathbb{Z}}$ with p -dimensional cross-spectral density given by $f(\lambda) = [f_0(\lambda), \dots, f_{p-1}(\lambda)]^T$, that is, for all $k, \tau \in \mathbb{Z}$,*

$$\text{Cov}(\underline{\mathbf{Y}}_{k+\tau}, X_k) = \int_{-\pi}^{\pi} f(\lambda) e^{i\lambda\tau} d\lambda.$$

Consider the decimated filtered process $Z_k = \sum_{n \in \mathbb{Z}} a[2k-n] X_n$. Then, the $\mathbb{R}^{2p} \times \mathbb{R}$ process $\left\{ \left[\left[\underline{\mathbf{Y}}_{2k}^T, \underline{\mathbf{Y}}_{2k+1}^T \right]^T, Z_k \right]^T \right\}_{k \in \mathbb{Z}}$ is covariance stationary with $(2p)$ -dimensional cross-spectral density $\tilde{f}(\lambda) = \left[\tilde{f}_0(\lambda), \dots, \tilde{f}_{2p-1}(\lambda) \right]^T$, given by

$$\tilde{f}_v(\lambda) = \frac{1}{2} \left[e^{i\lambda \langle \frac{v}{2} \rangle} A^* \left(\frac{\lambda}{2} \right) f_{\lfloor \frac{v}{2} \rfloor} \left(\frac{\lambda}{2} \right) + e^{i(\lambda+2\pi) \langle \frac{v}{2} \rangle} A^* \left(\frac{\lambda}{2} + \pi \right) f_{\lfloor \frac{v}{2} \rfloor} \left(\frac{\lambda}{2} + \pi \right) \right], \quad (4.40)$$

for all $v = 0, \dots, 2p-1$, where $A(\lambda) = \sum_{k \in \mathbb{Z}} a[k] e^{-i\lambda k}$ is the transfer function of a , $\lfloor x \rfloor$ denotes the integer part of x and $\langle x \rangle = x - \lfloor x \rfloor$ its fractional part.

Proof of Lemma (4.5.2). We denote by $\{\bar{Z}_k\}$ the undecimated filtered process $\bar{Z}_k = \sum_{l \in \mathbb{Z}} a[l] X_{k-l}$ so that $Z = \downarrow^2 \bar{Z}$. Then the cross-spectral density of the process $\left[\underline{\mathbf{Y}}_k^T, \bar{Z}_k \right]^T$ is given by $A^*(\lambda) f(\lambda)$. That is, for all $k, \tau \in \mathbb{Z}$,

$$\text{Cov}(\underline{\mathbf{Y}}_{k+\tau}, Z_k) = \int_{-\pi}^{\pi} e^{i\lambda\tau} A^*(\lambda) f(\lambda) d\lambda.$$

Then by Lemma 4.5.1, the two spectral densities $\tilde{f}^{(\epsilon)}$ of the two $\mathbb{R}^p \times \mathbb{R}$ processes $\left[\underline{\mathbf{Y}}_{2k+\epsilon}^T, Z_k \right]^T$, $\epsilon = 0, 1$ are given by

$$\tilde{f}^{(\epsilon)}(\lambda) = \frac{1}{2} \left\{ e^{i\frac{\lambda}{2}\epsilon} A^* \left(e^{i\frac{\lambda}{2}} \right) f \left(\frac{\lambda}{2} \right) + e^{i(\frac{\lambda}{2}+\pi)\epsilon} A^* \left(e^{i(\frac{\lambda}{2}+\pi)} \right) f \left(\frac{\lambda}{2} + \pi \right) \right\}.$$

We get (4.40) by noting that $\tilde{f}_v(\lambda) = \left[\tilde{f}_{\lfloor \frac{v}{2} \rfloor}^{(2\langle \frac{v}{2} \rangle)}(\lambda) \right]^T$ for all $v = 0, \dots, 2p-1$. \square

Proof of Proposition 4.3.1. Define $\check{A}_{j,k}^{(M)} = \sum_{l \in \mathbb{Z}} \tilde{h}[k-l] A_{j-1,l}^M$ and $\check{W}_{j,k} = \sum_{l \in \mathbb{Z}} \tilde{g}[k-l] A_{j-1,l}^M$. They respectively are the differenced undecimated approximation coefficient and the undecimated wavelet coefficient at scale j , since, by Lemma 2.2.1), $A_{j,\cdot}^{(M)} = \downarrow^2 \check{A}_{j,k}^{(M)}$ and $W_{j,\cdot} = \downarrow^2 \check{W}_{j,k}$.

The spectral densities of the processes $\{\check{A}_{j,k}^{(M)}\}_{k \in \mathbb{Z}}$ and $\{\check{W}_{j,k}\}_{k \in \mathbb{Z}}$ are given respectively by

$$\left| \tilde{H}(\lambda) \right|^2 \mathbf{A}_{j-1}(\lambda) \text{ and } \left| \tilde{G}(\lambda) \right|^2 \mathbf{A}_{j-1}(\lambda),$$

where \tilde{H} and \tilde{G} are given by in(2.57) and (2.54).

Then by applying Lemma 4.5.1 with $m = 2$, $r = 0$ and $f(\lambda) = \left| \tilde{H}(\lambda) \right|^2 \mathbf{A}_{j-1}(\lambda)$, we get (4.12). Similarly, this time with $f(\lambda) = \left| \tilde{G}(e^{i\lambda}) \right|^2 \mathbf{A}_{j-1}(\lambda)$, we get (4.13).

Now, the cross spectral density of the process $\{[\check{W}_{j,k}, \check{A}_{j,k}^{(M)}]^T\}_{k \in \mathbb{Z}}$ is given by $\mathbf{A}_{j-1}(\lambda) \tilde{H}^*(\lambda) \tilde{G}(\lambda)$, that is, for all $k, \tau \in \mathbb{Z}$,

$$\text{Cov} \left(\check{W}_{j,k+\tau}, \check{A}_{j,k}^{(M)} \right) = \int_{-\pi}^{\pi} \mathbf{A}_{j-1}(\lambda) \tilde{H}^*(\lambda) \tilde{G}(\lambda) e^{i\lambda\tau} d\lambda.$$

Hence applying Lemma 4.5.1 with $m = 2$, $r = 0$, and $f(\lambda) = \mathbf{A}_{j-1}(\lambda) \tilde{H}^*(\lambda) \tilde{G}(\lambda)$, we get (4.14).

To conclude the proof, we apply Lemma 4.5.2 with $\underline{\mathbf{Y}}_k = \mathbf{W}_{j+u,k}(u)$, $X_k = A_{j+u,k}^{(M)}$, $f(\lambda) = \mathbf{B}_{j+u,u}(\lambda)$ and a respectively equal to \tilde{h} and \tilde{g} . We respectively get (4.15) and (4.16). \square

Proof of Proposition 4.3.2. Using the definition of ν_j and (2.50), we have

$$\begin{aligned} \nu_j[k] &= \text{Cov} \left(A_{j,k}^{(M)} ; A_{j,0}^{(M)} \right) \\ &= \sum_{n,n' \in \mathbb{Z}} \tilde{h}[2k-n] \tilde{h}[-n'] \nu_{j-1}[n-n'] \\ &= \sum_{l,l' \in \mathbb{Z}} \tilde{h}[l] \tilde{h}[l'] \nu_{j-1}[2k-l+l']; \end{aligned}$$

which prove (4.18). Relations (4.19) and (4.20) are proved similarly using (2.60).

Using again (2.50), we have

$$\gamma_{i,j+1}[k] = \sum_{n \in \mathbb{Z}} \tilde{h}[-n] \text{Cov} \left(W_{i,k}, A_{j,n}^{(M)} \right).$$

On the other hand, since $\{[\mathbf{W}_{j,k}^T(j-i), A_{j,k}^{(M)}]^T\}_{k \in \mathbb{Z}}$ is weakly stationary, and $W_{i,k}$ is the ℓ -th entry of $\mathbf{W}_{j,k'}$ where $k = 2^{j-i}k' + \ell$, we have

$$\text{Cov} \left(W_{i,k}, A_{j,n}^{(M)} \right) = \text{Cov} \left(W_{i,k-2^{j-i}n}, A_{j,0}^{(M)} \right) = \gamma_{i,j}[k-2^{j-i}n].$$

Inserting this in the previous display and setting $l = -n$, we get (4.21). Relation (4.22) follows similarly by replacing \tilde{h} by \tilde{g} . \square

Chapter 5

Central Limit Theorem for the robust wavelet regression estimator

The material of this chapter is the article entitled "Central limit theorem for the robust log-regression wavelet estimation of the memory parameter in the Gaussian semi-parametric context " which is submitted to *Bernoulli*.

5.1 Introduction

Long-range dependent processes are characterized by hyperbolically slowly decaying correlations or by a spectral density exhibiting a fractional pole at zero frequency. During the last decades, long-range dependence (and the closely related self-similarity phenomena) has been observed in many different fields, including financial econometrics, hydrology or analysis of Internet traffic. In most of these applications, however, the presence of atypical observations is quite common. These outliers might be due to gross errors in the observations but also to unmodeled disturbances; see for example Stoev et al. [2006] and Stoev and Taqqu [2005] for possible explanations of the presence of outliers in Internet traffic analysis. It is well-known that even a few atypical observations can severely affect estimators, leading to incorrect conclusions. Hence, defining robust estimators of the memory parameter which are less sensitive to the presence of additive outliers is a challenging practical problem.

In this paper, we consider the class of fractional processes, denoted $M(d)$ defined in subsection 2.1.3. In the sequel, we assume that f^* in (2.11) belongs to $\mathcal{H}(\beta, L)$ with $0 < \beta \leq 2$ and $L > 0$ where $\mathcal{H}(\beta, L)$ denotes the set of non-negative and symmetric

functions g satisfying, for all $\lambda \in (-\pi, \pi)$,

$$|g(\lambda) - g(0)| \leq L g(0) |\lambda|^\beta . \quad (5.1)$$

Our setting is semi-parametric in that both d and f^* in (2.11) are unknown. Here, f^* can be seen as a nuisance parameter whereas d is the parameter of interest. This assumption on f^* is typical in the semi-parametric estimation setting; see for instance Robinson [1995b] and Moulines and Soulier [2003] and the references therein.

Different approaches have been proposed for building robust estimators of the memory parameter for $M(d)$ processes in the semi-parametric setting outlined above. Stoev et al. [2006] have proposed a robustified wavelet based-regression estimator developed by Abry and Veitch [1998]; the robustification is achieved by replacing the estimation of the wavelet coefficients variance at different scales by the median of the square of the wavelet coefficients. Another technique to robustify the wavelet regression technique has been outlined in Park and Park [2009] which consists in regressing the logarithm of the square of the wavelet coefficients at different scales. Molinares et al. [2009] proposed a robustified version of the log-periodogram regression estimator introduced in Geweke and Porter-Hudak [1983]. The method replaces the log-periodogram of the observation by a robust estimator of the spectral density in the neighborhood of the zero frequency, obtained as the discrete Fourier transform of a robust autocovariance estimator defined in Ma and Genton [2000]; the procedure is appealing and has been found to work well but also lacks theoretical support in the semi-parametric context (note however that the consistency and the asymptotic normality of the robust estimator of the covariance have been discussed in Lévy-Leduc et al. [2009]).

In the related context of the estimation of the fractal dimension of locally self-similar Gaussian processes Coeurjolly [2008b] has proposed a robust estimator of the Hurst coefficient; instead of using the variance of the generalized discrete variations of the process (which are closely related to the wavelet coefficients, despite the facts that the motivations are quite different), this author proposes to use the empirical quantiles and the trimmed-means. The consistency and asymptotic normality of this estimator is established for a class of locally self-similar processes, using a Bahadur-type representation of the sample quantile; see also Coeurjolly [2008a]. Shen et al. [2007] proposes to replace the classical regression of the wavelet coefficients by a robust regression approach, based on Huberized M-estimators.

The two robust estimators of d that we propose consist in regressing the logarithm of robust variance estimators of the wavelet coefficients of the process X on a range of scales. We use as robust variance estimators the square of the scale estimator proposed by Rousseeuw and Croux [1993] and the square of the *mean absolute deviation* (MAD). These estimators are a robust alternative to the estimator of d proposed by Moulines et al.

[2007a] which uses the same method but with the classical variance estimator. Here, we derive a Central Limit Theorem (CLT) for the two robust estimators of d and, by the way, we give another methodology for obtaining a Central Limit Theorem for the estimator of d proposed by Moulines et al. [2007a]. In this paper, we have also established new results on the empirical process of array of stationary Gaussian processes by extending [Arcones, 1994, Theorem 4] and the Theorem of Csörgő and Mielniczuk [1996] to arrays of stationary Gaussian processes. These new results were very helpful in establishing the CLT for the three estimators of d that we propose.

The chapter is organized as follows. In Section 5.2, we introduce the wavelet setting and define the wavelet based regression estimators of d . Section 5.3 is dedicated to the asymptotic properties of the robust estimators of d . In this section, we derive asymptotic expansions of the wavelet spectrum estimators and provide a CLT for the estimators of d . In Section 5.4, some Monte-Carlo experiments are presented in order to support our theoretical claims. The Nile River data and two Internet traffic packet counts datasets collected from the University of North Carolina, Chapel are studied as an application in Section 5.5. Sections 5.6 and 5.7 detail the proofs of the theoretical results stated in Section 5.3.

5.2 Definition of the wavelet-based regression estimators of the memory parameter d .

5.2.1 The wavelet setting

The wavelet setting involves two functions ϕ and ψ in $L^2(\mathbb{R})$ and their Fourier transforms

$$\widehat{\phi}(\xi) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \phi(t)e^{-i\xi t} dt \quad \text{and} \quad \widehat{\psi}(\xi) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \psi(t)e^{-i\xi t} dt. \quad (5.2)$$

Let us assume that conditions (W-1)-(W-4) hold. Condition (W-2) ensures that the Fourier transform $\widehat{\psi}$ decreases quickly to zero. Condition (W-3) ensures that ψ oscillates and that its scalar product with continuous-time polynomials up to degree $M - 1$ vanishes. It is equivalent to asserting that the first $M - 1$ derivatives of $\widehat{\psi}$ vanish at the origin and hence

$$|\widehat{\psi}(\lambda)| = O(|\lambda|^M), \quad \text{as } \lambda \rightarrow 0. \quad (5.3)$$

Daubechies wavelets (with $M \geq 2$) and the Coiflets satisfy these conditions, see Moulines et al. [2007a]. Viewing the wavelet $\psi(t)$ as a basic template, define the family $\{\psi_{j,k}, j \in \mathbb{Z}, k \in \mathbb{Z}\}$ of translated and dilated functions

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k), \quad j \in \mathbb{Z}, k \in \mathbb{Z}. \quad (5.4)$$

Positive values of k translate ψ to the right, negative values to the left. The *scale index* j dilates ψ so that large values of j correspond to coarse scales and hence to low frequencies. We suppose throughout the paper that

$$(1 + \beta)/2 - \alpha < d \leq M. \quad (5.5)$$

We now describe how the wavelet coefficients are defined in discrete time, that is for a real-valued sequence $\{x_k, k \in \mathbb{Z}\}$ and for a finite sample $\{x_k, k = 1, \dots, n\}$. Using the scaling function ϕ , we first interpolate these discrete values to construct the following continuous-time functions

$$\mathbf{x}_n(t) \stackrel{\text{def}}{=} \sum_{k=1}^n x_k \phi(t - k) \quad \text{and} \quad \mathbf{x}(t) \stackrel{\text{def}}{=} \sum_{k \in \mathbb{Z}} x_k \phi(t - k), \quad t \in \mathbb{R}. \quad (5.6)$$

Without loss of generality we may suppose that the support of the scaling function ϕ is included in $[-T, 0]$ for some integer $T \geq 1$. Then

$$\mathbf{x}_n(t) = \mathbf{x}(t) \quad \text{for all} \quad t \in [0, n - T + 1].$$

We may also suppose that the support of the wavelet function ψ is included in $[0, T]$. With these conventions, the support of $\psi_{j,k}$ is included in the interval $[2^j k, 2^j(k + T)]$. The wavelet coefficient $W_{j,k}$ at scale $j \geq 0$ and location $k \in \mathbb{Z}$ is formally defined as the scalar product in $L^2(\mathbb{R})$ of the function $t \mapsto \mathbf{x}(t)$ and the wavelet $t \mapsto \psi_{j,k}(t)$:

$$W_{j,k} \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \mathbf{x}(t) \psi_{j,k}(t) dt = \int_{-\infty}^{\infty} \mathbf{x}_n(t) \psi_{j,k}(t) dt, \quad j \geq 0, k \in \mathbb{Z}, \quad (5.7)$$

when $[2^j k, 2^j k + T] \subseteq [0, n - T + 1]$, that is, for all $(j, k) \in \mathcal{I}_n$, where where \mathcal{I}_n is defined in (2.32)

If $\Delta^M X$ is stationary, then from (2.43) the process $\{W_{j,k}\}_{k \in \mathbb{Z}}$ of wavelet coefficients at scale $j \geq 0$ is stationary but the two-dimensional process $\{[W_{j,k}, W_{j',k}]^T\}_{k \in \mathbb{Z}}$ of wavelet coefficients at scales j and j' , with $j \geq j'$, is not stationary. Here T denotes the transposition. This is why we consider instead the stationary *between-scale* process

$$\{[W_{j,k}, \mathbf{W}_{j,k}(j - j')]^T\}_{k \in \mathbb{Z}},$$

where $\mathbf{W}_{j,k}(j - j')$ is defined as follows:

$$\mathbf{W}_{j,k}(j - j') \stackrel{\text{def}}{=} \left[W_{j', 2^{j-j'} k}, W_{j', 2^{j-j'} k + 1}, \dots, W_{j', 2^{j-j'} k + 2^{j-j'} - 1} \right]^T.$$

For all $j, j' \geq 1$, the covariance function of the between scale process is given by

$$\text{Cov}(\mathbf{W}_{j,k'}(j - j'), W_{j,k}) = \int_{-\pi}^{\pi} e^{i\lambda(k-k')} \mathbf{D}_{j,j-j'}(\lambda; f) d\lambda, \quad (5.8)$$

where $\mathbf{D}_{j,j-j'}(\lambda; f)$ stands for the cross-spectral density function of this process. The case $j = j'$ corresponds to the spectral density function of the *within-scale* process $\{W_{j,k}\}_{k \in \mathbb{Z}}$.

In the sequel, we shall use that the within- and between-scale spectral densities $\mathbf{D}_{j,j-j'}(\lambda; d)$ of the process X with memory parameter $d \in \mathbb{R}$ can be approximated by the corresponding spectral density of the generalized fractional Brownian motion $B_{(d)}$ defined in (2.84).

5.2.2 Definition of the robust estimators of d

Let us now define robust estimators of the memory parameter d of the $M(d)$ process X from the observations X_1, \dots, X_n . These estimators are derived from the Abry and Veitch [1998] construction, and consists in regressing estimators of the scale spectrum

$$\sigma_j^2 \stackrel{\text{def}}{=} \text{Var}(W_{j,0}) \quad (5.9)$$

with respect to the scale index j . More precisely, if $\widehat{\sigma}_j^2$ is an estimator of σ_j^2 based on $W_{j,0:n_j-1} = (W_{j,0}, \dots, W_{j,n_j-1})$ then an estimator of the memory parameter d is obtained by regressing $\log(\widehat{\sigma}_j^2)$ for a finite number of scale indices $j \in \{J_0, \dots, J_0 + \ell\}$ where $J_0 = J_0(n) \geq 0$ is the lower scale and $1 + \ell \geq 2$ is the number of scales in the regression. The regression estimator can be expressed formally as

$$\widehat{d}_n(J_0, \mathbf{w}) \stackrel{\text{def}}{=} \sum_{j=J_0}^{J_0+\ell} w_{j-J_0} \log(\widehat{\sigma}_j^2) , \quad (5.10)$$

where the vector $\mathbf{w} \stackrel{\text{def}}{=} [w_0, \dots, w_\ell]^T$ of weights satisfies $\sum_{i=0}^{\ell} w_i = 0$ and $2 \log(2) \sum_{i=0}^{\ell} i w_i = 1$, see Abry and Veitch [1998] and Moulines et al. [2007b]. For $J_0 \geq 1$ and $\ell > 1$, one may choose for example \mathbf{w} corresponding to the least squares regression matrix, defined by $\mathbf{w} = DB(B^T DB)^{-1} \mathbf{b}$ where

$$\mathbf{b} \stackrel{\text{def}}{=} \begin{bmatrix} 0 & (2 \log(2))^{-1} \end{bmatrix}, \quad B \stackrel{\text{def}}{=} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 0 & 1 & \dots & \ell \end{bmatrix}^T$$

is the design matrix and D is an arbitrary positive definite matrix. The best choice of D depends on the memory parameter d . However a good approximation of this optimal matrix D is the diagonal matrix with diagonal entries $D_{i,i} = 2^{-i}$, $i = 0 \dots, \ell$; see Faÿ et al. [2009] and the references therein. We will use this choice of the design matrix in the numerical experiments. A heuristic justification for this choice is that by [Moulines et al., 2007a, Eq. (28)],

$$\sigma_j^2 \sim C 2^{2jd} , \text{ as } j \rightarrow \infty , \quad (5.11)$$

where C is a positive constant.

In the sequel, we shall consider three different estimators of d based on three different estimators of the scale spectrum σ_j^2 with respect to the scale index j which are defined below.

Classical scale estimator

This estimator has been considered in the original contribution of Abry and Veitch [1998] and consists in estimating the scale spectrum σ_j^2 with respect to the scale index j by the empirical variance

$$\hat{\sigma}_{\text{CL},j}^2 = \frac{1}{n_j} \sum_{i=1}^{n_j} W_{j,i}^2, \quad (5.12)$$

where for any j , n_j denotes the number of available wavelet coefficients at scale index j defined in (2.32).

Median absolute deviation

This estimator is well-known to be a robust estimator of the scale and as mentioned by Rousseeuw and Croux [1993] it has several appealing properties: it is easy to compute and has the best possible breakdown point (50%). Since the wavelet coefficients $W_{j,i}$ are centered Gaussian observations, the square of the median absolute deviation of $W_{j,0:n_j-1}$ is defined by

$$\hat{\sigma}_{\text{MAD},j}^2 = \left(m(\Phi) \operatorname{med}_{0 \leq i \leq n_j-1} W_{j,i} \right)^2, \quad (5.13)$$

where Φ denotes the c.d.f of a standard Gaussian random variable and

$$m(\Phi) = 1/\Phi^{-1}(3/4) = 1.4826. \quad (5.14)$$

The use of the median estimator to estimate the scalogram has been suggested to estimate the memory parameter in Stoev et al. [2005] (see also [Percival and Walden, 2006, p. 420]). A closely related technique is considered in Coeurjolly [2008a] and Coeurjolly [2008b] to estimate the Hurst coefficient of locally self-similar Gaussian processes. Note that the use of the median of the squared wavelet coefficients has been advocated to estimate the variance at a given scale in wavelet denoising applications; this technique is mentioned in Donoho and Johnstone [1994] to estimate the scalogram of the noise in the i.i.d. context; Johnstone and Silverman [1997] proposed to use this method in the long-range dependent context; the use of these estimators has not been however rigorously justified.

The Croux and Rousseeuw estimator

This estimator is another robust scale estimator introduced in Rousseeuw and Croux [1993]. Its asymptotic properties in several dependence contexts have been further studied in Lévy-Leduc et al. [2009] and the square of this estimator is defined by

$$\hat{\sigma}_{\text{CR},j}^2 = \left(c(\Phi) \{ |W_{j,i} - W_{j,k}|; 0 \leq i, k \leq n_j - 1 \}_{(k_{n_j})} \right)^2, \quad (5.15)$$

where $c(\Phi) = 2.21914$ and $k_{n_j} = \lfloor n_j^2/4 \rfloor$. That is, up to the multiplicative constant $c(\Phi)$, $\hat{\sigma}_{\text{CR},j}$ is the k_{n_j} th order statistics of the n_j^2 distances $|W_{j,i} - W_{j,k}|$ between all the pairs of observations.

5.3 Asymptotic properties of the robust estimators of d

5.3.1 Properties of the scale spectrum estimators

The following proposition gives an asymptotic expansion for $\hat{\sigma}_{\text{CL},j}^2$, $\hat{\sigma}_{\text{MAD},j}^2$ and $\hat{\sigma}_{\text{CR},j}^2$ defined in (5.12), (5.13) and (5.15), respectively. These asymptotic expansions are used for deriving Central Limit Theorems for the different estimators of d .

Proposition 5.3.1. *Assume that X is a Gaussian $M(d)$ process with generalized spectral density function defined in (2.11) such that $f^* \in \mathcal{H}(\beta, L)$ for some $L > 0$ and $0 < \beta \leq 2$. Assume that (W-1)-(W-4) hold with d , α and M satisfying (5.5). Let $W_{j,k}$ be the wavelet coefficients associated to X defined by (2.31). If $n \mapsto J_0(n)$ is an integer valued sequence satisfying $J_0(n) \rightarrow \infty$ and $n2^{-J_0(n)} \rightarrow \infty$, as $n \rightarrow \infty$, then $\hat{\sigma}_{*,j}^2$ defined in (5.12), (5.13) and (5.15), satisfies the following asymptotic expansion, as $n \rightarrow \infty$, for any given $\ell \geq 1$*

$$\max_{J_0(n) \leq j \leq J_0(n) + \ell} \left| \sqrt{n_j} (\hat{\sigma}_{*,j}^2 - \sigma_j^2) - \frac{2\sigma_j^2}{\sqrt{n_j}} \sum_{i=0}^{n_j-1} \text{IF} \left(\frac{W_{j,i}}{\sigma_j}, *, \Phi \right) \right| = o_P(1), \quad (5.16)$$

where $*$ denotes CL, CR and MAD, σ_j^2 is defined in (5.9) and IF is given by

$$\text{IF}(x, \text{CL}, \Phi) = \frac{1}{2} H_2(x), \quad (5.17)$$

$$\text{IF}(x, \text{CR}, \Phi) = c(\Phi) \left(\frac{1/4 - \Phi(x + 1/c(\Phi)) + \Phi(x - 1/c(\Phi))}{\int_{\mathbb{R}} \varphi(y) \varphi(y + 1/c(\Phi)) dy} \right), \quad (5.18)$$

$$\text{IF}(x, \text{MAD}, \Phi) = -m(\Phi) \left(\frac{(\mathbb{1}_{\{x \leq 1/m(\Phi)\}} - 3/4) - (\mathbb{1}_{\{x \leq -1/m(\Phi)\}} - 1/4)}{2\varphi(1/m(\Phi))} \right), \quad (5.19)$$

where φ denotes the p.d.f of the standard Gaussian random variable, $m(\Phi)$ and $c(\Phi)$ being defined in (5.14) and (5.15), respectively and $H_2(x) = x^2 - 1$ is the second Hermite polynomial.

The proof is postponed to Section 5.6.

We deduce from Proposition 5.3.1 and Theorem 5.6.3 given and proved in Section 5.6 the following multivariate Central Limit Theorem for the wavelet coefficient scales.

Theorem 5.3.1. *Under the assumptions of Proposition 5.3.1, $(\hat{\sigma}_{*,J_0}^2, \dots, \hat{\sigma}_{*,J_0+\ell}^2)^T$, where $\hat{\sigma}_{*,j}^2$ is defined in (5.12), (5.13) and (5.15), satisfies the following multivariate Central*

Limit Theorem

$$\sqrt{n2^{-J_0}2^{-2J_0d}} \left(\begin{bmatrix} \hat{\sigma}_{*,J_0}^2 \\ \hat{\sigma}_{*,J_0+1}^2 \\ \vdots \\ \hat{\sigma}_{*,J_0+\ell}^2 \end{bmatrix} - \begin{bmatrix} \sigma_{*,J_0}^2 \\ \sigma_{*,J_0+1}^2 \\ \vdots \\ \sigma_{*,J_0+\ell}^2 \end{bmatrix} \right) \xrightarrow{d} \mathcal{N}(0, \mathbf{U}_*(d)) , \quad (5.20)$$

where

$$\begin{aligned} \mathbf{U}_{*,i,j}(d) &= 4(f^*(0))^2 \sum_{p \geq 2} \frac{c_p^2(\text{IF}_*)}{p! K(d)^{p-2}} 2^{d(2+p)i \vee j} 2^{d(2-p)i \wedge j + i \wedge j} \\ &\quad \times \sum_{\tau \in \mathbb{Z}} \sum_{r=0}^{2^{i-j} - 1} \left(\int_{-\pi}^{\pi} \mathbf{D}_{\infty,|i-j|}^{(r)}(\lambda; d) e^{i\lambda\tau} d\lambda \right)^p, \quad 0 \leq i, j \leq \ell. \end{aligned} \quad (5.21)$$

In (5.21), $K(d)$ is defined in (2.90), $\mathbf{D}_{\infty,|i-j|}(\cdot; d)$ is the cross-spectral density defined in (2.84), $c_p(\text{IF}_*) = \mathbb{E}[\text{IF}(X, *, \Phi) H_p(X)]$, where H_p is the p th Hermite polynomial and $\text{IF}(\cdot, *, \Phi)$ is defined in (5.17), (5.18) and (5.19).

The proof of Theorem 5.3.1 is postponed to Section 5.6.

Remark 5.3.1. Since for $* = \text{CL}$, $\text{IF}(\cdot) = H_2(\cdot)/2$, Theorem 5.3.1 gives an alternative proof to [Moulines et al., 2007a, Theorem 2] of the limiting covariance matrix of $(\hat{\sigma}_{\text{CL},J_0}^2, \dots, \hat{\sigma}_{\text{CL},J_0+\ell}^2)^T$ which is given, for $0 \leq i, j \leq \ell$, by

$$\mathbf{U}_{\text{CL},i,j}(d) = 4\pi (f^*(0))^2 2^{4d(i \vee j) + i \wedge j} \int_{-\pi}^{\pi} |\mathbf{D}_{\infty,|i-j|}(\lambda; d)|^2 d\lambda.$$

Thus, for $* = \text{CR}$ and $* = \text{MAD}$, we deduce the following

$$\frac{\mathbf{U}_{\text{CL},i,i}(d)}{\mathbf{U}_{*,i,i}(d)} \geq \frac{1/2}{\mathbb{E}[\text{IF}_*^2(Z)]}, \quad (5.22)$$

where Z is a standard Gaussian random variable. With Lemma 5.7.2, we deduce from the inequality (5.22) that the asymptotic relative efficiency of $\hat{\sigma}_{*,j}^2$ is larger than 36.76% when $* = \text{MAD}$ and larger than 82.27% when $* = \text{CR}$.

5.3.2 CLT for the robust wavelet-based regression estimator

Based on the results obtained in the previous section, we derive a Central Limit Theorem for the robust wavelet-based regression estimators of d defined by

$$\hat{d}_{*,n}(J_0, \mathbf{w}) \stackrel{\text{def}}{=} \sum_{j=J_0}^{J_0+\ell} w_{j-J_0} \log(\hat{\sigma}_{*,j}^2), \quad (5.23)$$

where $\hat{\sigma}_{*,j}^2$ are given for $* = \text{CL}$, MAD and CR by (5.12), (5.13) and (5.15), respectively.

Theorem 5.3.2. *Under the same assumptions as in Proposition 5.3.1 and if*

$$n2^{-(1+2\beta)J_0(n)} \rightarrow 0, \text{ as } n \rightarrow \infty, \quad (5.24)$$

then, $\widehat{d}_{*,n}(J_0, \mathbf{w})$ satisfies the following Central Limit Theorem:

$$\sqrt{n2^{-J_0(n)}} \left(\widehat{d}_{*,n}(J_0, \mathbf{w}) - d \right) \xrightarrow{d} \mathcal{N} \left(0, \mathbf{w}^T \mathbf{V}_*(d) \mathbf{w} \right), \quad (5.25)$$

where $\mathbf{V}_*(d)$ is the $(1 + \ell) \times (1 + \ell)$ matrix defined by

$$\mathbf{V}_{*,i,j}(d) = \sum_{p \geq 2} \frac{4c_p^2(\text{IF}_*)}{p! \mathbf{K}(d)^p} 2^{pd|i-j|+i \wedge j} \sum_{\tau \in \mathbb{Z}} \sum_{r=0}^{2^{|i-j|}-1} \left(\int_{-\pi}^{\pi} \mathbf{D}_{\infty,|i-j|}^{(r)}(\lambda; d) e^{i\lambda\tau} d\lambda \right)^p, \quad 0 \leq i, j \leq \ell. \quad (5.26)$$

In (5.26), $\mathbf{K}(d)$ is defined in (2.90), $\mathbf{D}_{\infty,|i-j|}(\cdot; d)$ is the cross-spectral density defined in (2.84), $c_p(\text{IF}_*) = \mathbb{E}[\text{IF}(X, *, \Phi) H_p(X)]$, where H_p is the p th Hermite polynomial and $\text{IF}(\cdot, *, \Phi)$ is defined in (5.17), (5.18) and (5.19).

The proof of Theorem 5.3.2 is a straightforward consequence of [Moulines et al., 2007a, Proposition 3] and Theorem 5.3.1 and is thus not detailed here.

Remark 5.3.2. *Since it is difficult to provide a theoretical lower bound for the asymptotic relative efficiency (ARE) of $\widehat{d}_{*,n}(J_0, \mathbf{w})$ defined by*

$$\text{ARE}_*(d) = \mathbf{w}^T \mathbf{V}_{\text{CL}}(d) \mathbf{w} / \mathbf{w}^T \mathbf{V}_*(d) \mathbf{w}, \quad (5.27)$$

where $*$ = CR or MAD, we propose to compute this quantity empirically. We know from Theorem 5.3.2 that the expression of the limiting covariance matrix $\mathbf{V}_{*,i,j}(d)$ is valid for all Gaussian $M(d)$ processes satisfying the assumptions given in Proposition 5.3.1, thus it is enough to compute $\text{ARE}_*(d)$ in the particular case of a Gaussian ARFIMA(0,d,0) process (X_t) . We propose to evaluate $\text{ARE}_*(d)$ when d belongs to $[-0.8; 3]$. With such a choice of d , both stationary and non-stationary processes are considered. The empirical values of $\text{ARE}_*(d)$ are given in Table 5.1. The results were obtained from the observations X_1, \dots, X_n where $n = 2^{12}$ and 1000 independent replications. We used Daubechies wavelets with $M = 2$ vanishing moments when $d \leq 2$ and $M = 4$ when $d > 2$ which ensures that condition (5.5) is satisfied. The smallest scale is chosen to be $J_0 = 3$ and $J_0 + \ell = 8$.

From Table 5.1, we can see that $\widehat{d}_{n,\text{CR}}$ is more efficient than $\widehat{d}_{n,\text{MAD}}$ and that its asymptotic relative efficiency ARE_{CR} ranges from 0.63 to 0.79. These results indicate empirically that the the loss of efficiency of the robust estimator $\widehat{d}_{n,\text{CR}}$ is moderate and makes it an attractive robust procedure to the non-robust estimator $\widehat{d}_{n,\text{CL}}$.

d	-0.8	-0.4	-0.2	0	0.2	0.6	1	1.6	2	2.2	2.6	3
$\text{ARE}_{\text{CR}}(d)$	0.72	0.67	0.63	0.65	0.70	0.63	0.75	0.75	0.79	0.74	0.77	0.74
$\text{ARE}_{\text{MAD}}(d)$	0.48	0.39	0.38	0.36	0.43	0.39	0.47	0.50	0.48	0.5	0.49	0.49

Table 5.1: Asymptotic relative efficiency of $\hat{d}_{n,\text{CR}}$ and $\hat{d}_{n,\text{MAD}}$ with respect to $\hat{d}_{n,\text{CL}}$ for $n = 2^{12}$ observations of an ARFIMA(0, d , 0) process.

5.4 Numerical experiments

In this section the robustness properties of the different estimators of d , namely $\hat{d}_{\text{CL},n}(J_0, \mathbf{w})$, $\hat{d}_{\text{CR},n}(J_0, \mathbf{w})$ and $\hat{d}_{\text{MAD},n}(J_0, \mathbf{w})$, that are defined in Section 5.2.2 are investigated using Monte Carlo experiments. In the sequel, the memory parameter d is estimated from $n = 2^{12}$ observations of a Gaussian ARFIMA(0, d , 0) process defined in (2.17) when $d=0.2$ and 1.2 eventually corrupted by additive outliers. We use the Daubechies wavelets with $M = 2$ vanishing moments which ensures that condition (5.5) is satisfied.

Let us first explain how to choose the parameters J_0 and $J_0 + \ell$. With $n = 2^{12}$, the maximal available scale is equal to 10. Choosing J_0 too small may introduce a bias in the estimation of d by Theorem 5.3.2. However, at coarse scales (large values of J_0), the number of observations may be too small and thus choosing J_0 too large may yield a large variance. Since at scales $j = 9$ and $j = 10$, we have respectively 5 and 1 observations, we chose $J_0 + \ell = 8$. For the choice of J_0 , we proposed to use the empirical rule explained illustrated in Figure 5.1. In this figure, we display the estimates $\hat{d}_{n,\text{CL}}$, $\hat{d}_{n,\text{CR}}$ and $\hat{d}_{n,\text{MAD}}$ of the memory parameter d as well as their respective 95% confidence intervals from $J_0 = 1$ to $J_0 = 7$ with $J_0 + \ell = 8$. We propose to choose $J_0 = 3$ in both cases ($d = 0.2$ and $d = 1.2$) since the successive confidence intervals starting from $J_0 = 3$ to $J_0 = 7$ are such that the smallest one is included in the largest one. We shall take $J_0 = 3$ in the sequel.

In the left panels of Figures 5.2 and 5.3 the empirical distribution of $\sqrt{n2^{-J_0}}(\hat{d}_{*,n} - d)$ are displayed when $*$ = CL, MAD and CR for the ARFIMA(0, d , 0) model with $d = 0.2$ (Figure 5.2) and $d = 1.2$ (Figure 5.3), respectively. They were computed using 5000 replications; their shapes are close to the Gaussian density (the standard deviations are of course different). In the right panels of Figures 5.2 and 5.3, the empirical distribution of $\sqrt{n2^{-J_0}}(\hat{d}_{*,n} - d)$ are displayed when outliers are present. We introduce 1% of additive outliers in the observations; these outliers are obtained by choosing uniformly at random a time index and to add to the selected observation 5 times the standard error of the raw observations. The empirical distribution of $\sqrt{n2^{-J_0}}(\hat{d}_{\text{CL},n} - d)$ is clearly located far away from zero especially in the non stationary ARFIMA(0, 1.2, 0) model. One can also observe the considerable increase in the variance of the classical estimator. In sharp contrast, the distribution of the robust estimators $\sqrt{n2^{-J_0}}(\hat{d}_{\text{MAD},n} - d)$ and $\sqrt{n2^{-J_0}}(\hat{d}_{\text{CR},n} - d)$ stay symmetric and the variance stays constant.

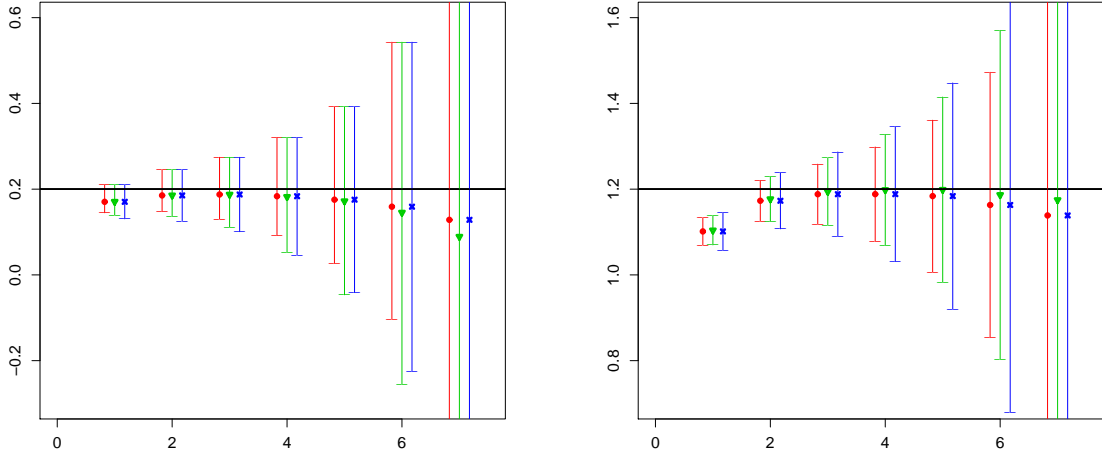


Figure 5.1: Confidence intervals of the estimates $\hat{d}_{n,CL}$, $\hat{d}_{n,CR}$ and $\hat{d}_{n,MAD}$ of an ARFIMA(0, d , 0) process with $d = 0.2$ (left) and $d = 1.2$ (right) for $J_0 = 1, \dots, 7$ and $J_0 + \ell = 8$. For each J_0 , are displayed confidence interval associated to $\hat{d}_{n,CL}$ (red), interval $\hat{d}_{n,CR}$ (green) and $\hat{d}_{n,MAD}$ (blue), respectively.

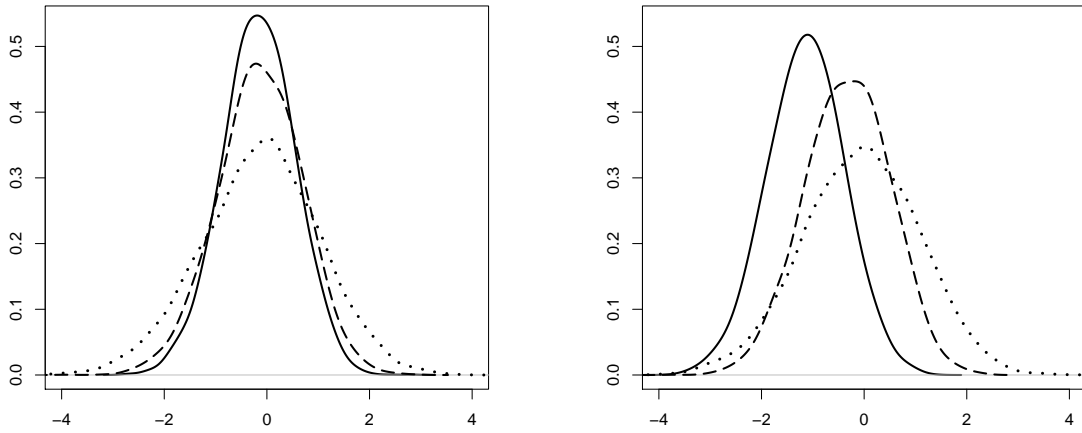


Figure 5.2: Empirical densities of the quantities $\sqrt{n2^{-J_0}}(\hat{d}_{*,n} - d)$, with $*$ = CL (solid line), $*$ = CR (dashed line) and $*$ = MAD (dotted line) of the ARFIMA(0,0.2,0) model without outliers (left) and with 1% of outliers (right).

5.5 Application to real Data

In this section, we compare the performance of the different estimators of the long memory parameter d introduced in Section 5.2.2 on two different real datasets.

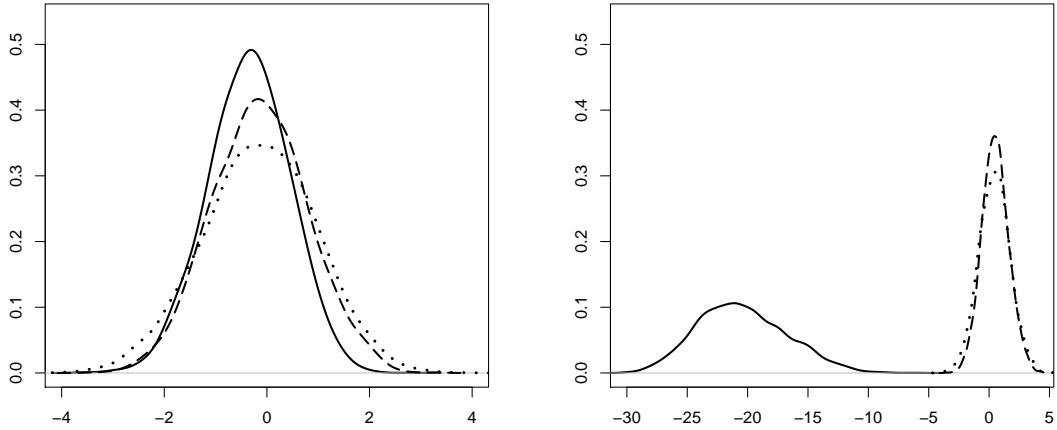


Figure 5.3: Empirical densities of the quantities $\sqrt{n2^{-J_0}}(\hat{d}_{*,n} - d)$, with $*$ = CL (solid line), $*$ = CR (dashed line) and $*$ = MAD (dotted line) of the ARFIMA(0,1,2,0) model without outliers (left) and with 1% of outliers (right).

5.5.1 Nile River data

The Nile River dataset is a well-known time series, which has been extensively analyzed; see [Beran, 1994, Section 1.4, p. 20]. The data consists of yearly minimal water levels of the Nile river measured at the Roda gauge, near Cairo, for the years 622–1284 AD and contains 663 observations; The units for the data as presented by Beran [1994] are centimeters. The empirical mean and the standard deviation of the data are equal to 1148 and 89.05, respectively. The question has been raised as to whether the Nile time series contains outliers; see for example Beran [1992], Robinson [1995b], Chareka et al. [2006] and Molinares et al. [2009]. The test procedure developed by Chareka et al. [2006] suggests the presence of outliers at 646 AD (p -value 0.0308) and at 809 (p -value 0.0007). Another possible outliers is at 878 AD. Since the number of observations is small, in the estimation of d , we took $J_0 = 1$ and $J_0 + \ell = 6$. With this choice, we observe a significant difference between the classical estimators $\hat{d}_{n,CL} = 0.28$ (with 95% confidence interval [0.23, 0.32]) and the robust estimators $\hat{d}_{n,CR} = 0.408$ (with 95% confidence interval [0.34, 0.46]) and $\hat{d}_{n,MAD} = 0.414$ (with 95% confidence interval [0.34, 0.49]). Thus, to better understand the influence of outliers on the estimated memory parameter in practical situations, a new dataset with artificial outliers was generated. Here, we replaced the presumed outliers of Chareka et al. [2006] by the value of the observation plus 10 times the standard deviation. The new memory parameter estimators are $\hat{d}_{n,CL} = 0.12$, $\hat{d}_{n,CR} = 0.4$ and $\hat{d}_{n,MAD} = 0.392$. As was expected, the values of the robust estimators remained stable. However, the classical estimator of d was significantly affected. A robust estimate of d for the Nile data is also given in Agostinelli [2003] and in Molinares et al. [2009]. The authors found 0.412 and 0.416, respectively. These values are very close to $\hat{d}_{n,CR} = 0.408$ and $\hat{d}_{n,MAD} = 0.414$.

5.5.2 Internet traffic packet counts data

In this section, two Internet traffic packet counts datasets collected at the University of North Carolina, Chapel (UNC) are analyzed. These datasets are available from the website http://netlab.cs.unc.edu/public/old_research/net_lrd/. These datasets have been studied by Park and Park [2009].

Figure 5.4 (left) displays a packet count time series measured at the link of UNC on April 13, Saturday, from 7:30 p.m. to 9:30 p.m., 2002 (Sat1930). Figure 5.4 (right) displays the same type of time series but on April 11, a Thursday, from 1 p.m. to 3 p.m., 2002 (Thu1300). These packet counts were measured every 1 millisecond but, for a better display, we aggregated them at 1 second.

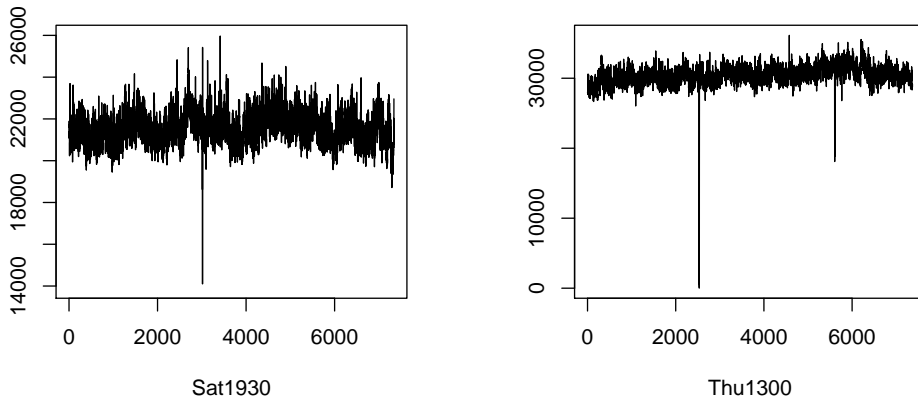


Figure 5.4: Packet counts of aggregated traffic every 1 second.

The maximal available scale for the two datasets is 20. Since we have less than 4 observations at this scale, we set the coarse scale $J_0 + \ell = 19$ and vary the finest scale J_0 from 1 to 17. The values of the three estimators of d are stored in Table 5.2.

Following Park and Park [2009], the finest scale J_0 is set to 9 for (Thu1300) and J_0 is to 10 for (Sat1930). Furthermore, in view of Figure 5.5, we would choose for the two datasets the same value of J_0 as Park and Park [2009] since for $J_0 = 9$ for (Thu) and $J_0 = 10$ for (Sat) the successive intervals starting from $J_0 = 9$ to $J_0 = 14$ for (Thu) and $J_0 = 10$ to $J_0 = 14$ are such that the smaller one is included in the larger one. For these values of the finest scale J_0 , we obtain $\hat{d}_{n,CL} = 0.43$ (with 95% confidence interval [0.412, 0.443]), $\hat{d}_{n,CR} = 0.37$ (with 95% confidence interval [0.358, 0.385]) and $\hat{d}_{n,MAD} = 0.38$ with (95% confidence interval [0.362, 0.397]) for Thu1300 and $\hat{d}_{n,CL} = 0.36$ (with 95% confidence interval [0.345, 0.374]), $\hat{d}_{n,CR} = \hat{d}_{n,MAD} = 0.38$ (with 95% confidence interval [0.361, 0.398] for CR and [0.357, 0.402] for MAD) for Sat1930. These values are similar to the one found by Park and Park [2009].

With this choice of J_0 in Thu1300, we observe a significant difference between the

Thu1300							
J_0	1	2	3	4	5	6	7
$\widehat{d}_{n,CL}$	0.08	0.09	0.11	0.15	0.19	0.23	0.29
S.E _{CL}	(0.52)	(0.56)	(0.51)	(0.52)	(0.57)	(0.52)	(0.56)
$\widehat{d}_{n,CR}$	0.08	0.07	0.07	0.09	0.13	0.19	0.28
S.E _{CR}	(0.55)	(0.58)	(0.61)	(0.63)	(0.59)	(0.6)	(0.67)
$\widehat{d}_{n,MAD}$	0.08	0.08	0.07	0.09	0.13	0.19	0.27
S.E _{MAD}	(0.74)	(0.87)	(0.78)	(0.83)	(0.86)	(0.84)	(0.91)
J_0	8	9	10	11	12	13	14
$\widehat{d}_{n,CL}$	0.36	0.43	0.49	0.55	0.52	0.41	0.21
S.E _{CL}	(1.45)	(0.64)	(0.61)	(0.71)	(0.81)	(1.08)	(1.09)
$\widehat{d}_{n,CR}$	0.34	0.37	0.40	0.42	0.43	0.48	0.45
S.E _{CR}	(0.90)	(0.81)	(0.70)	(0.88)	(0.96)	(1.21)	(1.26)
$\widehat{d}_{n,MAD}$	0.33	0.38	0.40	0.43	0.43	0.5	0.48
S.E _{MAD}	(1.49)	(0.98)	(1.04)	(1.07)	(1.15)	(1.18)	(1.2)
Sat1930							
J_0	1	2	3	4	5	6	7
$\widehat{d}_{n,CL}$	0.05	0.06	0.08	0.11	0.14	0.17	0.23
S.E _{CL}	(0.41)	(0.47)	(0.43)	(0.48)	(0.47)	(0.48)	(0.46)
$\widehat{d}_{n,CR}$	0.06	0.06	0.06	0.09	0.12	0.16	0.23
S.E _{CR}	(0.51)	(0.47)	(0.54)	(0.48)	(0.48)	(0.53)	(0.56)
$\widehat{d}_{n,MAD}$	0.06	0.06	0.07	0.09	0.11	0.16	0.23
S.E _{MAD}	(0.59)	(0.77)	(0.72)	(0.81)	(0.70)	(0.89)	(0.82)
J_0	8	9	10	11	12	13	14
$\widehat{d}_{n,CL}$	0.28	0.33	0.36	0.37	0.39	0.42	0.42
S.E _{CL}	(0.89)	(0.54)	(0.61)	(0.70)	(0.80)	(1.11)	(1.24)
$\widehat{d}_{n,CR}$	0.3	0.34	0.38	0.4	0.42	0.44	0.42
S.E _{CR}	(0.90)	(0.81)	(0.70)	(0.88)	(0.96)	(1.21)	(1.26)
$\widehat{d}_{n,MAD}$	0.29	0.33	0.38	0.4	0.43	0.45	0.4
S.E _{MAD}	(0.64)	(1.13)	(0.99)	(1.10)	(1.34)	(1.49)	(1.38)

Table 5.2: Estimators of d from $J_0 = 1$ to $J_0 = 17$ and $J_0 + \ell = 19$ obtained from Thu1300 and Sat1930. Here S.E denotes the standard error of $(\sqrt{n2^{-J_0}}(\widehat{d} - d))$

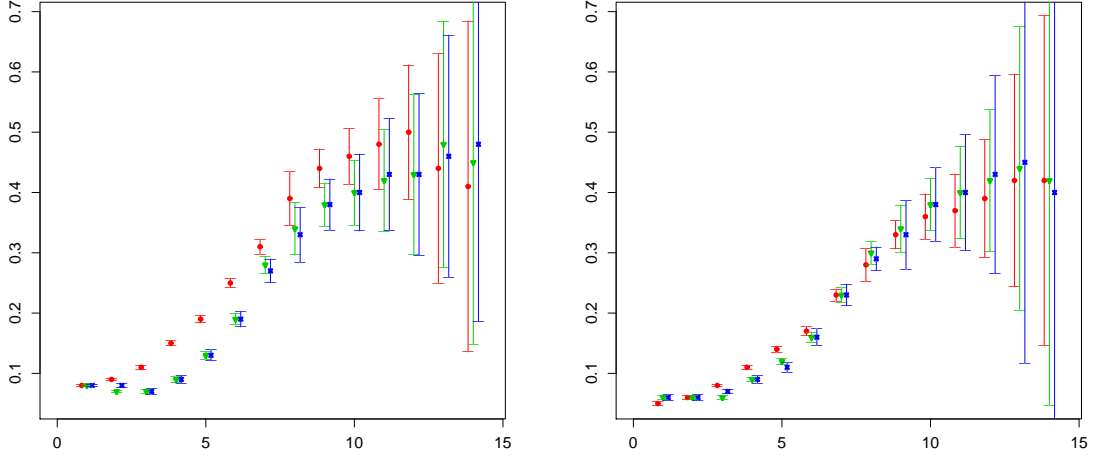


Figure 5.5: Confidence intervals of the estimates $\hat{d}_{n,CL}$, $\hat{d}_{n,CR}$ and $\hat{d}_{n,MAD}$ on the data Thu (left) and Sat (right) for $J_0 = 1, \dots, 14$ and $J_0 + \ell = 19$. For each J_0 , are displayed confidence interval associated to $\hat{d}_{n,CL}$ (red), interval $\hat{d}_{n,CR}$ (green) and $\hat{d}_{n,MAD}$ (blue), respectively.

classical estimator and the robust estimators. Thus to better understand the influence of outliers on the estimated memory parameter a new dataset with artificial outliers was generated. The Thu1300 time series shows two spikes shooting down. Especially, the first downward spike hits zero. Park et al. [2007] have shown that this dropout lasted 8 seconds. Outliers are introduced by dividing by 6 the 8000 observations in this period. The new memory parameter estimators are $\hat{d}_{n,CL} = 0.445$, $\hat{d}_{n,CR} = 0.375$ and $\hat{d}_{n,MAD} = 0.377$. As for the Nile River data, the classical estimator was affected while the robust estimators remain stable.

5.6 Proofs

Theorem 5.6.1 is an extension of [Arcones, 1994, Theorem 4] to arrays of stationary Gaussian processes in the unidimensional case and Theorem 5.6.2 extends the result of Csörgő and Mielniczuk [1996] to arrays of stationary Gaussian processes. These two theorems are useful for the proof of Proposition 5.3.1.

Theorem 5.6.1. *Let $\{X_{j,i}, j \geq 1, i \geq 0\}$ be an array of standard stationary Gaussian processes such that for a fixed $j \geq 1$, $(X_{j,i})_{i \geq 0}$ has a spectral density f_j and an autocorrelation function ρ_j defined by $\rho_j(k) = \mathbb{E}(X_{j,0}X_{j,k})$, for all $k \geq 0$. Assume also that there exists a non increasing sequence $\{u_j\}_{j \geq 1}$ such that for all $j \geq 1$*

$$\sup_{\lambda \in (-\pi, \pi)} |f_j(\lambda) - g_\infty(\lambda)| \leq u_j, \quad (5.28)$$

where g_∞ is a 2π -periodic function which is bounded on $(-\pi, \pi)$ and continuous at the origin. Let h be a function on \mathbb{R} with Hermite rank $\tau \geq 1$. We assume that h is either bounded or is a finite linear combination of Hermite polynomials. Let $\{n_j\}_{j \geq 1}$, be a sequence of integers such that n_j tends to infinity as j tends to infinity. Then,

$$\frac{1}{\sqrt{n_j}} \sum_{i=1}^{n_j} h(X_{j,i}) \xrightarrow{d} \mathcal{N}(0, \tilde{\sigma}^2), \text{ as } j \rightarrow \infty, \quad (5.29)$$

where

$$\tilde{\sigma}^2 = \lim_{n \rightarrow \infty} \text{Var} \left(\frac{1}{\sqrt{n_j}} \sum_{i=1}^{n_j} h(X_{j,i}) \right) = \sum_{\ell \geq \tau} \frac{c_\ell^2}{\ell!} g_\infty^{\star \ell}(0).$$

In the previous equality, $c_\ell = \mathbb{E}[h(X)H_\ell(X)]$, where H_ℓ is the ℓ -th Hermite polynomial and X is a standard Gaussian random variable.

Proof of Theorem 5.6.1. Let us first prove that

$$\frac{\sum_{i=1}^{n_j} \sum_{l \geq \tau} \frac{c_l}{l!} H_l(X_{j,i})}{\sqrt{\text{Var} \left(\sum_{i=1}^{n_j} \sum_{l \geq \tau} \frac{c_l}{l!} H_l(X_{j,i}) \right)}} \xrightarrow{d} \mathcal{N}(0, 1), \text{ as } n \rightarrow \infty. \quad (5.30)$$

Using Mehler's formula, see Eq. (2.1) of Breuer and Major [1983], we have

$$\begin{aligned} \text{Var} \left(\sum_{i=1}^{n_j} \sum_{l \geq \tau} \frac{c_l}{l!} H_l(X_{j,i}) \right) &= \sum_{i_1, i_2=1}^{n_j} \sum_{l_1, l_2 \geq \tau} \frac{c_{l_1} c_{l_2}}{l_1! l_2!} \mathbb{E} [H_{l_1}(X_{j,i_1}) H_{l_2}(X_{j,i_2})] \\ &= \sum_{l \geq \tau} \frac{c_l^2}{l!} \left[\sum_{i_1, i_2=1}^{n_j} \rho_j^l(i_2 - i_1) \right]. \end{aligned}$$

In order to prove (5.30), it is enough to prove that for $p \geq 1$,

$$\frac{\mathbb{E} \left[\left(\sum_{i=1}^{n_j} \sum_{l \geq \tau} \frac{c_l}{l!} H_l(X_{j,i}) \right)^{2p+1} \right]}{\left(\sum_{l \geq \tau} \frac{c_l^2}{l!} \left[\sum_{i_1, i_2=1}^{n_j} \rho_j^l(i_2 - i_1) \right] \right)^{\frac{2p+1}{2}}} \rightarrow 0, \text{ as } n \rightarrow \infty \text{ and} \quad (5.31)$$

$$\frac{\mathbb{E} \left[\left(\sum_{i=1}^{n_j} \sum_{l \geq \tau} \frac{c_l}{l!} H_l(X_{j,i}) \right)^{2p} \right]}{\left(\sum_{l \geq \tau} \frac{c_l^2}{l!} \left[\sum_{i_1, i_2=1}^{n_j} \rho_j^l(i_2 - i_1) \right] \right)^p} \rightarrow \frac{(2p)!}{p! 2^p}, \text{ as } n \rightarrow \infty. \quad (5.32)$$

For all $m \in \mathbb{N}^*$,

$$\mathbb{E} \left[\left(\sum_{i=1}^{n_j} \sum_{l \geq \tau} \frac{c_l}{l!} H_l(X_{j,i}) \right)^m \right] = \sum_{1 \leq i_1, \dots, i_m \leq n_j} \sum_{l_1, \dots, l_m \geq \tau} \frac{c_{l_1} \dots c_{l_m}}{l_1! \dots l_m!} \mathbb{E} [H_{l_1}(X_{j,i_1}), \dots, H_{l_m}(X_{j,i_m})].$$

1) We start with the case where $m = 2p + 1$.

a) Let us first assume that $|\{i_1, \dots, i_{2p+1}\}| = 2p + 1$ and that

$$\forall i, \rho_j(i) \leq \rho^* < 1/(2p). \quad (5.33)$$

By [Taqqu, 1977, Lemma 3.2 P. 210], $\mathbb{E}[H_{l_1}(X_{j,i_1}), \dots, H_{l_m}(X_{j,i_m})]$ is zero if $l_1 + \dots + l_m$ is odd. Otherwise it is bounded by a constant times a sum of products of $(l_1 + \dots + l_m)/2$ correlations. Bounding, in each product, all of them but $p+1$, by $\rho^* < 1/(2p)$, we get that $\mathbb{E}[H_{l_1}(X_{j,i_1}), \dots, H_{l_{2p+1}}(X_{j,i_{2p+1}})]$ is bounded by a finite number of terms of the following form

$$(\rho^*)^{\frac{l_1 + \dots + l_{2p+1}}{2} - (p+1)} \rho_j(i_2 - i_1) \rho_j(i_4 - i_3) \dots \rho_j(i_{2p} - i_{2p-1}) \rho_j(i_{2p+1} - i_{2p}) |\mathbb{E}(H_{l_1}(X) \dots H_{l_{2p+1}}(X))|,$$

where X is a standard Gaussian random variable. Note also that the hypercontractivity [Taqqu, 1977, Lemma 3.1 P.210] yields

$$|\mathbb{E}[H_{l_1}(X) \dots H_{l_{2p+1}}(X)]| \leq (2p)^{\frac{l_1 + \dots + l_{2p+1}}{2}} \sqrt{l_1! \dots l_{2p+1}!}.$$

Thus, using the Cauchy-Schwarz inequality and that $\rho^* < \frac{1}{2p}$, there exists a positive constant C such that

$$\begin{aligned} & \sum_{l_1, \dots, l_{2p+1} \geq \tau} \frac{|c_{l_1} \dots c_{l_{2p+1}}|}{l_1! \dots l_{2p+1}!} (\rho^*)^{\frac{l_1 + \dots + l_{2p+1}}{2} - (p+1)} |\mathbb{E}(H_{l_1}(X) \dots H_{l_{2p+1}}(X))| \\ & \leq \sum_{l_1, \dots, l_{2p+1} \geq \tau} \frac{|c_{l_1}| \dots |c_{l_{2p+1}}|}{\sqrt{l_1! \dots l_{2p+1}!}} (2p\rho^*)^{\frac{l_1 + \dots + l_{2p+1}}{2} - (p+1)} \leq (2p\rho^*)^{-1} \left(\sum_{l \geq \tau} \frac{|c_l|}{\sqrt{l!}} [(2p\rho^*)]^{\frac{l}{2} - \frac{p}{2p+1}} \right)^{2p+1} \\ & \leq C \left(\sum_{l \geq \tau} \frac{c_l^2}{l!} \right)^{\frac{2p+1}{2}} \left(\sum_{l \geq \tau} (2p\rho^*)^{l - \frac{2p}{2p+1}} \right)^{\frac{2p+1}{2}} < \infty. \end{aligned}$$

To conclude the proof of (5.31), it remains to prove that

$$\frac{\sum_{\substack{1 \leq i_1, \dots, i_{2p+1} \leq n_j \\ |\{i_1, \dots, i_{2p+1}\}| = 2p+1}} \rho_j(i_2 - i_1) \rho_j(i_4 - i_3) \dots \rho_j(i_{2p} - i_{2p-1}) \rho_j(i_{2p+1} - i_{2p})}{\left(\sum_{l \geq \tau} \frac{c_l^2}{l!} \left[\sum_{i_1, i_2=1}^{n_j} \rho_j^l(i_2 - i_1) \right] \right)^{p + \frac{1}{2}}} \rightarrow 0, \text{ as } n_j \rightarrow \infty. \quad (5.34)$$

Let us first study the numerator in the l.h.s of (5.34).

$$\begin{aligned} & \sum_{\substack{1 \leq i_1, \dots, i_{2p+1} \leq n_j \\ |\{i_1, \dots, i_{2p+1}\}| = 2p+1}} \rho_j(i_2 - i_1) \rho_j(i_4 - i_3) \dots \rho_j(i_{2p} - i_{2p-1}) \rho_j(i_{2p+1} - i_{2p}) \\ & = \left(\sum_{1 \leq i_1 \neq i_2 \leq n_j} \rho_j(i_2 - i_1) \right)^{p-1} \sum_{\substack{1 \leq i_{2p-1}, i_{2p}, i_{2p+1} \leq n_j \\ |\{i_{2p-1}, i_{2p}, i_{2p+1}\}| = 3}} \rho_j(i_{2p} - i_{2p-1}) \rho_j(i_{2p+1} - i_{2p}) \\ & = \left(\sum_{1 \leq i_1 \neq i_2 \leq n_j} \rho_j(i_2 - i_1) \right)^{p-1} \sum_{i_{2p}=1}^{n_j} \left(\sum_{1 \leq i_{2p} \neq i_{2p+1} \leq n_j} \rho_j(i_{2p+1} - i_{2p}) \right)^2. \end{aligned}$$

To prove (5.34), we start by proving that

$$\sum_{r=1}^{n_j} \left(\sum_{1 \leq s \leq n_j} \rho_j(r - s) \right)^2 = O(n_j). \quad (5.35)$$

Using the notation $D_{n_j}(\lambda) = \sum_{r=1}^{n_j} e^{i\lambda r}$, we get

$$\begin{aligned} \sum_{r=1}^{n_j} \left(\sum_{1 \leq s \leq n_j} \rho_j(r-s) \right)^2 &= \sum_{r=1}^{n_j} \left(\int_{-\pi}^{\pi} e^{i\lambda r} \sum_{1 \leq s \leq n_j} e^{-i\lambda s} f_j(\lambda) d\lambda \right)^2 \\ &= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} D_{n_j}(\lambda - \lambda') D_{n_j}(\lambda) \overline{D_{n_j}(\lambda')} f_j(\lambda) f_j(\lambda') d\lambda d\lambda'. \end{aligned}$$

Using (5.28), the boundedness of g_∞ and that u_j is bounded, there exists a positive constant C such that

$$\begin{aligned} |f_j(\lambda) f_j(\lambda')| &\leq |f_j(\lambda) - g_\infty(\lambda)| |f_j(\lambda') - g_\infty(\lambda')| + |g_\infty(\lambda')| |f_j(\lambda) - g_\infty(\lambda)| \\ &\quad + |g_\infty(\lambda)| |f_j(\lambda') - g_\infty(\lambda')| + |g_\infty(\lambda)| |g_\infty(\lambda')| \leq C. \end{aligned}$$

Then, using that there exists a positive constant c such that $|D_{n_j}(\lambda)| \leq cn_j/(1 + n_j|\lambda|)$, for all λ in $[-\pi, \pi]$,

$$\sum_{r=1}^{n_j} \left(\sum_{1 \leq s \leq n_j} \rho_j(r-s) \right)^2 \leq c^3 n_j \int_{\mathbb{R}^2} \frac{1}{1 + |\mu - \mu'|} \frac{1}{1 + |\mu|} \frac{1}{1 + |\mu'|} d\mu d\mu'. \quad (5.36)$$

The result (5.35) thus follows from the convergence of the integral in (5.36) which is proved in Lemma 5.7.3. Let us now prove that

$$\frac{1}{n_j} \sum_{1 \leq r, s \leq n_j} \rho_j(r-s) \rightarrow g_\infty(0), \text{ as } n \rightarrow \infty. \quad (5.37)$$

Using that F_j defined by $F_j(\lambda) = (2\pi n_j)^{-1} \left| \sum_{r=1}^{n_j} e^{i\lambda r} \right|^2$, for all λ in $[-\pi, \pi]$ satisfies $\int_{-\pi}^{\pi} F_j(\lambda) d\lambda = 1$, we obtain

$$\frac{1}{n_j} \left(\sum_{1 \leq r, s \leq n_j} \rho_j(r-s) \right) - g_\infty(0) = \int_{-\pi}^{\pi} (f_j(\lambda) - g_\infty(\lambda)) F_j(\lambda) d\lambda + \int_{-\pi}^{\pi} (g_\infty(\lambda) - g_\infty(0)) F_j(\lambda) d\lambda. \quad (5.38)$$

Using that $\int_{-\pi}^{\pi} F_j(\lambda) d\lambda = 1$ and (5.28), the first term in the r.h.s of (5.38) tends to zero as n tends to infinity. The second term in the r.h.s of (5.38) can be upper bounded as follows. For $0 < \eta \leq \pi$,

$$\begin{aligned} \left| \int_{-\pi}^{\pi} (g_\infty(\lambda) - g_\infty(0)) F_j(\lambda) d\lambda \right| &\leq \int_{-\pi}^{-\eta} |g_\infty(\lambda) - g_\infty(0)| F_j(\lambda) d\lambda \\ &\quad + \int_{-\eta}^{\eta} |g_\infty(\lambda) - g_\infty(0)| F_j(\lambda) d\lambda + \int_{\eta}^{\pi} |g_\infty(\lambda) - g_\infty(0)| F_j(\lambda) d\lambda. \end{aligned} \quad (5.39)$$

Since there exists a positive constant C such that $F_j(\lambda) \leq C/(n_j|\lambda|^2)$, for all λ in $[-\pi, \pi]$, the first and last terms in the r.h.s of (5.39) are bounded by $C\pi/(n_j\eta^2)$. The continuity of g_∞ at 0 and the fact that $\int_{-\eta}^{\eta} F_j(\lambda) d\lambda \leq \int_{-\pi}^{\pi} F_j(\lambda) d\lambda = 1$ ensure that the second term in the r.h.s of (5.39) tends to zero as n tends to infinity. This concludes the proof of (5.37).

Using the same arguments as those used to prove (5.37) and the fact that ρ_j^l is the autocorrelation associated to f_j^{*l} which is the l -th self-convolution of f_j , we get that

$$\frac{1}{n_j} \sum_{r,s=1}^{n_j} \rho_j^l(r-s) \rightarrow g_\infty^{*l}(0), \text{ as } n \rightarrow \infty. \quad (5.40)$$

Let us now prove that the denominator in (5.34) is $O(n_j^{p+\frac{1}{2}})$ as $n \rightarrow \infty$. We aim at applying Lemma 5.7.6 with f_n, g_n, f and g defined hereafter.

$$f_{n_j}(s, l) = \frac{c_l^2}{l!} \mathbb{1}_{\{|s| < n_j\}} \left(1 - \frac{|s|}{n_j}\right) \rho_j^l(s).$$

Observe that $|f_{n_j}(s, l)| \leq g_{n_j}(s, l)$ where

$$g_{n_j}(s, l) = \frac{c_l^2}{l!} \mathbb{1}_{\{|s| < n_j\}} \left(1 - \frac{|s|}{n_j}\right) \rho_j^2(s).$$

Using (5.28) and the fact that the spectral density associated to ρ_j^l is f_j^{*l} , we get, as $n \rightarrow \infty$,

$$f_{n_j}(s, l) \rightarrow f(s, l) = \frac{c_l^2}{l!} \int_{-\pi}^{\pi} g_\infty^{*l}(\lambda) e^{i\lambda s} d\lambda \text{ and } g_{n_j}(s, l) \rightarrow g(s, l) = \frac{c_l^2}{l!} \int_{-\pi}^{\pi} g_\infty^{*2}(\lambda) e^{i\lambda s} d\lambda.$$

Using [Moulines et al., 2007b, Lemma 1], we get

$$\sum_{l \geq \tau} \sum_{s \in \mathbb{Z}} g_{n_j}(s, l) \rightarrow \sum_{l \geq \tau} \frac{c_l^2}{l!} g_\infty^{*2}(0).$$

Then, Lemma 5.7.6 yields

$$\lim_{n \rightarrow \infty} \frac{1}{n_j} \text{Var} \left(\sum_{i=1}^{n_j} \sum_{l \geq \tau} \frac{c_l}{l!} H_l(X_{j,i}) \right) = \lim_{n \rightarrow \infty} \frac{1}{n_j} \sum_{l \geq \tau} \frac{c_l^2}{l!} \left[\sum_{i_1, i_2=1}^{n_j} \rho_j^l(i_2 - i_1) \right] = \sum_{l \geq \tau} \frac{c_l^2}{l!} g_\infty^{*l}(0).$$

Hence we get (5.34) by noticing that the numerator in (5.34) is $O(n_j^p)$.

If Condition (5.33) is not satisfied then let k_0 be such that $\rho_j(k) \leq \rho^* < 1/(2p)$, for all $k > k_0$. In the case where h is a linear combination of L Hermite polynomials, the same arguments as those used previously are valid with $\rho^* = 1$. In the case where h is bounded, there exists a positive constant C such that

$$\mathbb{E} \left[\left(\sum_{i=1}^{n_j} h(X_{j,i}) \right)^{2p+1} \right] \leq C \sum_{1 \leq i_1, \dots, i_q \leq n_j} \mathbb{E} [|h|(X_{j,i_1}) \dots |h|(X_{j,i_q})], \quad (5.41)$$

where i_1, \dots, i_q are such that $|i_k - i_l| > k_0$, for all k, l in $\{1, \dots, q\}$ with $q \leq 2p + 1$. By expanding $|h|$ onto the basis of Hermite polynomials, we can conclude with the same arguments as those used when Condition (5.33) is valid.

b) Let us now assume that $|\{i_1, \dots, i_{2p+1}\}| = r \leq 2p$. In the case where h is bounded, the inequality (5.41) is valid with $q \leq r$ which gives that the numerator of (5.31) is

$O(n_j^{\lfloor r/2 \rfloor})$. In the case where h is a linear combination of L Hermite polynomials, we use the same arguments as those used in a) with $\rho^* = 1$ which implies that the numerator of (5.31) is $O(n_j^{\lfloor r/2 \rfloor})$.

2) Let us now study the case where m is even that is $m = 2p$ with $p \geq 1$.

$$\mathbb{E} \left[\left(\sum_{i=1}^{n_j} \sum_{l \geq \tau} \frac{c_l}{l!} H_l(X_{j,i}) \right)^{2p} \right] = \sum_{1 \leq i_1, \dots, i_{2p} \leq n_j} \sum_{l_1, \dots, l_{2p} \geq \tau} \frac{c_{l_1} \dots c_{l_{2p}}}{l_1! \dots l_{2p}!} \mathbb{E} [H_{l_1}(X_{j,i_1}) \dots H_{l_{2p}}(X_{j,i_{2p}})] . \quad (5.42)$$

By [Rosenblatt, 1985, Formula (33), P.69], we have

$$\mathbb{E} [H_{l_1}(X_{j,i_1}) \dots H_{l_{2p}}(X_{j,i_{2p}})] = l_1! \dots l_{2p}! \sum_{\{l_1, \dots, l_{2p}\}} \frac{\rho_j^{\nu}}{\nu!}, \quad (5.43)$$

where it is understood that $\rho_j^{\nu} = \prod_{1 \leq q < k \leq 2p} \rho_j^{\nu_{q,k}} (q - k)$, $\nu! = \prod_{1 \leq q < k \leq 2p} \nu_{q,k}!$, and $\sum_{\{l_1, \dots, l_{2p}\}}$ indicates that we are to sum over all symmetric matrices ν with nonnegative integer entries, $\nu_{ii} = 0$ and the row sums equal to l_1, \dots, l_{2p} .

We shall prove that among all the terms in the r.h.s of (5.43), the leading ones correspond to the case where we have p pairs of equal indices in the set $\{l_1, \dots, l_{2p}\}$, that is, for instance, $l_1 = l_2, l_3 = l_4, \dots, l_{2p-1} = l_{2p}$ and $\nu_{1,2} = l_1, \nu_{3,4} = l_3, \dots, \nu_{2p-1,2p} = l_{2p-1}$ the others $\nu_{i,j}$ being equal to zero. This gives

$$(l_2!)^2 \dots (l_{2p}!)^2 \frac{\rho_j^{(i_2 - i_1)^{l_2}} \rho_j^{(i_4 - i_3)^{l_4}} \dots \rho_j^{(i_{2p} - i_{2p-1})^{l_{2p}}}}{l_2! \dots l_{2p}!} .$$

The corresponding term in (5.42) is given by

$$\begin{aligned} \sum_{1 \leq i_1, \dots, i_{2p} \leq n_j} \sum_{l_2, l_4, \dots, l_{2p} \geq \tau} \frac{c_{l_2}^2 c_{l_4}^2 \dots c_{l_{2p}}^2}{l_2! l_4! \dots l_{2p}!} \rho_j^{(i_2 - i_1)^{l_2}} \rho_j^{(i_4 - i_3)^{l_4}} \dots \rho_j^{(i_{2p} - i_{2p-1})^{l_{2p}}} \\ = \left[\sum_{l \geq \tau} \frac{c_l^2}{l!} \left(\sum_{i_1, i_2=1}^{n_j} \rho_k^l (i_2 - i_1) \right) \right]^p , \end{aligned}$$

which corresponds to the denominator in the l.h.s of (5.32). Since there exists exactly $(2p)!/(2^p p!)$ possibilities to have pairs of equal indices among $2p$ indices we obtain (5.32) if we prove that the other terms can be neglected.

Let us first consider the case where

$$\forall i, \rho_j(i) \leq \rho^* < \frac{1}{2p-1} \quad (5.44)$$

and $|\{i_1, \dots, i_{2p}\}| = 2p$. By [Taqqu, 1977, Lemma 3.2 P. 210], $\mathbb{E} [H_{l_1}(X_{j,i_1}), \dots, H_{l_m}(X_{j,i_m})]$ is zero if $l_1 + \dots + l_m$ is odd. Otherwise it is bounded by a constant times a sum of products of $(l_1 + \dots + l_m)/2$ correlations. Bounding, in each product, all of them but $p + 1$, by

$\rho^* < 1/(2p-1)$, we get that $\mathbb{E} [H_{l_1}(X_{j,i_1}), \dots, H_{l_{2p}}(X_{j,i_{2p}})]$ is bounded by a finite number of terms of the following form

$$(\rho^*)^{\frac{l_1+\dots+l_{2p}}{2}-(p+1)} \rho_j(i_2-i_1) \rho_j(i_4-i_3) \dots \rho_j(i_{2p}-i_{2p-1}) \rho_j(i_{2p}-i_1) |\mathbb{E} (H_{l_1}(X) \dots H_{l_{2p}}(X))| .$$

where X is a standard Gaussian random variable. Using the same arguments as in the case where m was odd, we have

$$\sum_{l_1, \dots, l_{2p} \geq \tau} \frac{|c_{l_1} \dots c_{l_{2p}}|}{l_1! \dots l_{2p}!} (\rho^*)^{\frac{l_1+\dots+l_{2p}}{2}-(p+1)} |\mathbb{E} (H_{l_1}(X) \dots H_{l_{2p}}(X))| < \infty .$$

To have the result (5.32), it remains to show that

$$\frac{\sum_{\substack{1 \leq i_1, \dots, i_{2p} \leq n_j \\ |\{i_1, \dots, i_{2p}\}| = 2p}} \rho_j(i_2-i_1) \rho_j(i_4-i_3) \dots \rho_j(i_{2p}-i_{2p-1}) \rho_j(i_{2p}-i_1)}{\left[\sum_{l \geq \tau} \frac{c_l^2}{l!} \left(\sum_{i_1, i_2=1}^{n_j} \rho_j^l(i_2-i_1) \right) \right]^p} \rightarrow 0, \text{ as } n \rightarrow \infty . \quad (5.45)$$

The numerator of (5.45) can be rewritten as

$$\begin{aligned} & \sum_{\substack{1 \leq i_1, \dots, i_{2p} \leq n_j \\ |\{i_1, \dots, i_{2p}\}| = 2p}} \rho_j(i_2-i_1) \rho_j(i_4-i_3) \dots \rho_j(i_{2p}-i_{2p-1}) \rho_j(i_{2p}-i_1) \\ &= \left(\sum_{1 \leq i_3 \neq i_4 \leq n_j} \rho_j(i_4-i_3) \right)^{p-2} \left[\sum_{\substack{1 \leq i_1, i_2, i_{2p-1}, i_{2p} \leq n_j \\ |\{i_1, i_2, i_{2p-1}, i_{2p}\}| = 4}} \rho_j(i_2-i_1) \rho_j(i_{2p}-i_{2p-1}) \rho_j(i_{2p}-i_1) \right]. \end{aligned}$$

Using (5.37), we have $\left(\sum_{1 \leq i_3 \neq i_4 \leq n_j} \rho_j(i_4-i_3) \right)^{p-2} = O(n_j^{p-2})$. Let us now prove that

$$\sum_{1 \leq i_1, i_2, i_3, i_4 \leq n_j} \rho_j(i_2-i_1) \rho_j(i_3-i_4) \rho_j(i_3-i_1) = O(n_j) . \quad (5.46)$$

Using the notation $D_{n_j}(\lambda) = \sum_{r=1}^{n_j} e^{i\lambda r}$,

$$\begin{aligned} & \sum_{1 \leq i_1, i_2, i_3, i_4 \leq n_j} \rho_j(i_2-i_1) \rho_j(i_3-i_4) \rho_j(i_3-i_1) \\ &= \sum_{1 \leq i_1, i_2, i_3, i_4 \leq n_j} \left(\int_{-\pi}^{\pi} e^{i\lambda(i_2-i_1)} f_j(\lambda) d\lambda \right) \left(\int_{-\pi}^{\pi} e^{i\mu(i_3-i_4)} f_j(\mu) d\mu \right) \left(\int_{-\pi}^{\pi} e^{i\xi(i_3-i_1)} f_j(\xi) d\xi \right) \\ &= \int_{-\pi}^{\pi} f_j(\xi) \left(\int_{-\pi}^{\pi} \overline{D_{n_j}(\mu)} D_{n_j}(\mu + \xi) f_j(\mu) d\mu \int_{-\pi}^{\pi} D_{n_j}(\lambda) \overline{D_{n_j}(\lambda + \xi)} f_j(\lambda) d\lambda \right) d\xi \\ &\leq \int_{-\pi}^{\pi} \left(\int_{-\pi}^{\pi} |D_{n_j}(\lambda)| |D_{n_j}(\lambda + \xi)| f_j(\lambda) d\lambda \right)^2 f_j(\xi) d\xi . \end{aligned}$$

Using (5.28) and that g_∞ is bounded, (5.46) will follow if we prove that $\int_{-\pi}^{\pi} \left(\int_{-\pi}^{\pi} |D_{n_j}(\lambda)| |D_{n_j}(\lambda + \xi)| d\lambda \right)^2 d\xi = O(n_j)$. Since there exists a positive constant c such that $|D_{n_j}(\lambda)| \leq cn_j/(1 +$

$n_j|\lambda|)$, for all λ in $[-\pi, \pi]$,

$$\begin{aligned} & \int_{-\pi}^{\pi} \left(\int_{-\pi}^{\pi} |D_{n_j}(\lambda)| |D_{n_j}(\lambda + \xi)| f_j(\lambda) d\lambda \right)^2 f_j(\xi) d\xi \\ & \leq c^4 n_j \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \frac{1}{1 + |\mu|} \frac{1}{|1 + \mu + \mu'|} d\mu \right)^2 d\mu' \end{aligned} \quad (5.47)$$

The result (5.46) thus follows from the convergence of the last integral in (5.47) which is proved in Lemma 5.7.4. Hence we get (5.45) since the numerator of the l.h.s of (5.45) is $O(n_j^{p-1})$ and the denominator is $O(n_j^p)$ by the same arguments as those used to find the order of the denominator of (5.34). If Condition (5.44) is not satisfied or if $|\{i_1, \dots, i_{2p}\}| < 2p$, we can use similar arguments as those used in 1)a) and 1)b) to conclude the proof. \square

Theorem 5.6.2. *Let $\{X_{j,i}, j \geq 1, i \geq 0\}$ be an array of standard stationary Gaussian processes such that for a fixed $j \geq 1$, $(X_{j,i})_{i \geq 0}$ has a spectral density f_j and an autocorrelation function ρ_j defined by $\rho_j(k) = \mathbb{E}(X_{j,0}X_{j,k})$, for all $k \geq 0$. Let F_j be the c.d.f of $X_{j,1}$ and F_{n_j} the empirical c.d.f computed from $X_{j,1}, \dots, X_{j,n_j}$. If Condition (5.28) holds,*

$$\sqrt{n_j}(F_{n_j} - F_j) \xrightarrow{d} W \quad \text{in } D([-\infty, \infty]), \quad (5.48)$$

where W is a Gaussian process and $D([-\infty, \infty])$ denotes the Skorokhod space on $[-\infty, \infty]$.

Proof of Theorem 5.6.2. Let $S_j(x) = n_j^{-1/2} \sum_{i=1}^{n_j} (\mathbb{1}_{\{X_{j,i} \leq x\}} - F_j(x))$, for all x in \mathbb{R} . We shall first prove that for x_1, \dots, x_Q and a_1, \dots, a_Q in \mathbb{R}

$$\sum_{q=1}^Q a_q S_j(x_q) \xrightarrow{d} \mathcal{N} \left(0, \sum_{l \geq 1} \frac{c_l^2}{l!} g_{\infty}^{*l}(0) \right), \quad \text{as } n \rightarrow \infty, \quad (5.49)$$

where c_l is the l -th Hermite coefficient of the function h defined by

$$h(\cdot) = \sum_{q=1}^Q a_q (\mathbb{1}_{\{\cdot \leq x_q\}} - \mathbb{E}(\mathbb{1}_{\{\cdot \leq x_q\}})) .$$

Thus, $\sum_{q=1}^Q a_q S_j(x_q) = n_j^{-1/2} \sum_{i=1}^{n_j} h(X_{j,i})$, where h is bounded and of Hermite rank $\tau \geq 1$ since for all t in \mathbb{R} , $\mathbb{E}(X \mathbb{1}_{X \leq t}) = \int_{\mathbb{R}} x \mathbb{1}_{x \leq t} \varphi(x) dx = \int_{-\infty}^t (-\varphi(x))' dx = -\varphi(t) \neq 0$, and the CLT (5.49) follows from Theorem 5.6.1.

Let us now prove that there exists a positive constant C and $\beta > 1$ such that for all $r \leq s \leq t$,

$$\mathbb{E}(|S_j(s) - S_j(r)|^2 |S_j(t) - S_j(s)|^2) \leq C|t - r|^\beta . \quad (5.50)$$

The convergence (5.48) then follows from (5.49), (5.50) and [Billingsley, 1999, Theorem 13.5]. Note that

$$\begin{aligned} & \mathbb{E}(|S_j(s) - S_j(r)|^2 |S_j(t) - S_j(s)|^2) \\ & = \frac{1}{n_j^2} \sum_{i,i'=1}^{n_j} \sum_{l,l'=1}^{n_j} \mathbb{E}((f_s - f_r)(X_{j,i})(f_s - f_r)(X_{j,i'})(f_t - f_s)(X_{j,l})(f_t - f_s)(X_{j,l'})) , \end{aligned}$$

where $f_t(X) = \mathbb{1}_{\{X \leq t\}} - \mathbb{E}(\mathbb{1}_{\{X \leq t\}})$. By developing each difference of functions in Hermite polynomials, we get

$$\mathbb{E}(|S_j(s) - S_j(r)|^2 |S_j(t) - S_j(s)|^2) = \frac{1}{n_j^2} \sum_{i,i'=1}^{n_j} \sum_{l,l'=1}^{n_j} \sum_{p_1, \dots, p_4 \geq 1} \frac{c_{p_1}(f_s - f_r) c_{p_2}(f_s - f_r) c_{p_3}(f_t - f_s) c_{p_4}(f_t - f_s)}{p_1! \dots p_4!} \mathbb{E}(H_{p_1}(X_{j,i}) H_{p_2}(X_{j,i'}) H_{p_3}(X_{j,l}) H_{p_4}(X_{j,l'})) .$$

Using the same arguments as in the case where m is even in the proof of Theorem 5.6.1, we obtain

$$\begin{aligned} \mathbb{E}(|S_j(s) - S_j(r)|^2 |S_j(t) - S_j(s)|^2) &= \frac{1}{n_j^2} \sum_{p_1, p_2 \geq 1} \sum_{i, i', l, l'=1}^{n_j} \left[\frac{c_{p_1}^2(f_t - f_s) c_{p_2}^2(f_s - f_r)}{p_1! p_2!} \right. \\ &\rho_j^{p_1}(i' - i) \rho_j^{p_2}(l' - l) + \frac{c_{p_1}(f_t - f_s) c_{p_1}(f_s - f_r) c_{p_2}(f_t - f_s) c_{p_2}(f_s - f_r)}{p_1! p_2!} \rho_j^{p_1}(l - i) \rho_j^{p_2}(l' - i') \\ &\left. + \frac{c_{p_1}(f_t - f_s) c_{p_1}(f_s - f_r) c_{p_2}(f_t - f_s) c_{p_2}(f_s - f_r)}{p_1! p_2!} \rho_j^{p_1}(l' - i) \rho_j^{p_2}(l - i') \right] + O(n_j^{-1}) . \end{aligned}$$

Let $\|\cdot\|_2 = (\mathbb{E}(\cdot)^2)^{1/2}$ and $\langle f, g \rangle = \mathbb{E}[f(X)g(X)]$, where X is a standard Gaussian random variable. Since, by (5.40), $\sum_{i, i', l, l'=1}^{n_j} \rho_j^{p_1}(l - i) \rho_j^{p_2}(l' - i') = O(n_j^2)$, we get with the Cauchy-Schwarz inequality that there exists a positive constant C such that

$$\begin{aligned} &\mathbb{E}(|S_j(s) - S_j(r)|^2 |S_j(t) - S_j(s)|^2) \\ &\leq C \sum_{p_1, p_2 \geq 1} \left[\frac{c_{p_1}^2(f_t - f_s) c_{p_2}^2(f_s - f_r)}{p_1! p_2!} + \frac{c_{p_1}(f_t - f_s) c_{p_1}(f_s - f_r) c_{p_2}(f_t - f_s) c_{p_2}(f_s - f_r)}{p_1! p_2!} \right] \\ &\leq C \left(\|f_t - f_s\|_2^2 \|f_s - f_r\|_2^2 + |\langle f_t - f_s, f_s - f_r \rangle|^2 \right) \leq C \|f_t - f_s\|_2^2 \|f_s - f_r\|_2^2 . \end{aligned}$$

Note that $\|f_t - f_s\|_2^2 \leq 2(\|\mathbb{1}_{\{X \leq t\}} - \mathbb{1}_{\{X \leq s\}}\|_2^2 + \|\mathbb{E}(\mathbb{1}_{\{X \leq s\}}) - \mathbb{E}(\mathbb{1}_{\{X \leq t\}})\|_2^2)$. Since $s \leq t$, $\|\mathbb{1}_{\{X \leq t\}} - \mathbb{1}_{\{X \leq s\}}\|_2^2 = \Phi(t) - \Phi(s) \leq C|t - s|$, where Φ denotes the c.d.f of a standard Gaussian random variable. Moreover, $\|\mathbb{E}(\mathbb{1}_{\{X \leq s\}}) - \mathbb{E}(\mathbb{1}_{\{X \leq t\}})\|_2^2 \leq C|t - s|^2$, which concludes the proof of (5.50). \square

Proof of Proposition 5.3.1. We first prove (5.16) for $*$ = CL.

$$\sqrt{n_j}(\hat{\sigma}_{\text{CL},j}^2 - \sigma_j^2) = \frac{1}{\sqrt{n_j}} \sum_{i=0}^{n_j-1} (W_{j,i}^2 - \sigma_j^2) = \frac{2\sigma_j^2}{\sqrt{n_j}} \sum_{i=0}^{n_j-1} \frac{1}{2} \left(\frac{W_{j,i}^2}{\sigma_j^2} - 1 \right) .$$

Let us now prove (5.16) for $*$ = MAD. Let us denote by F_{n_j} the empirical c.d.f of $W_{j,0:n_j-1}$ and by F_j the c.d.f of $W_{j,0}$. Note that

$$\hat{\sigma}_{\text{MAD},j} = m(\Phi) T_0(F_{n_j}) ,$$

where $T_0 = T_2 \circ T_1$ with $T_1 : F \mapsto \{r \mapsto \int_{\mathbb{R}} \mathbb{1}_{\{|x| \leq r\}} dF(x)\}$ and $T_2 : U \mapsto U^{-1}(1/2)$. To prove (5.16), we start by proving that $\sqrt{n_j}(F_{n_j} - F_j)$ converges in distribution in

the space of cadlag functions equipped with the topology of uniform convergence. This convergence follows by applying Theorem 5.6.2 to $X_{j,i} = W_{j,i}/\sigma_j$ which is an array of zero mean stationary Gaussian processes by [Moulines et al., 2007b, Corollary 1]. The spectral density f_j of $(X_{j,i})_{i \geq 0}$ is given by $f_j(\lambda) = \mathbf{D}_{j,0}(\lambda; f)/\sigma_j^2$ where $\mathbf{D}_{j,0}(\cdot; f)$ is the within scale spectral density of the process $\{W_{j,k}\}_{k \geq 0}$ defined in (5.8) and σ_j^2 is the wavelet spectrum defined in (5.9). Here, $g_\infty(\lambda) = \mathbf{D}_{\infty,0}(\lambda; d)/\mathbf{K}(d)$, with $\mathbf{D}_{\infty,0}(\cdot; d)$ defined in (2.84) and $\mathbf{K}(d)$ defined in (2.90) since by [Moulines et al., 2007b, (26) and (29) in Theorem 1]

$$\left| \frac{\mathbf{D}_{j,0}(\lambda; f)}{f^*(0)\mathbf{K}(d)2^{2dj}} - \frac{\mathbf{D}_{\infty,0}(\lambda; d)}{\mathbf{K}(d)} \right| \leq C L \mathbf{K}(d)^{-1} 2^{-\beta j} \rightarrow 0, \text{ as } n \rightarrow \infty,$$

$$\left| \frac{\sigma_j^2}{f^*(0)\mathbf{K}(d)2^{2dj}} - 1 \right| \leq C L 2^{-\beta j} \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Note also that, by [Moulines et al., 2007b, Theorem 1], $g_\infty(\lambda)$ is a continuous and 2π -periodic function on $(-\pi, \pi)$. Moreover, $g_\infty(\lambda)$ is bounded on $(-\pi, \pi)$ by Lemma 5.7.5 and

$$u_j = C_1 \frac{2^{-\beta j}}{\sigma_j^2/2^{2dj}} \left(2^{-\beta j} + C_2 \frac{\sigma_j^2}{2^{2dj}} \right) \rightarrow 0, \text{ as } n \rightarrow \infty,$$

where C_1 and C_2 are positive constants. The asymptotic expansion (5.16) for $\widehat{\sigma}_{\text{MAD},j}$ can be deduced from the functional Delta method stated *e.g* in [van der Vaart, 1998, Theorem 20.8] and the classical Delta Method stated *e.g* in [van der Vaart, 1998, Theorem 3.1]. To show this, we have to prove that $T_0 = T_1 \circ T_2$ is Hadamard differentiable and that the corresponding Hadamard differential is defined and continuous on the whole space of cadlag functions. We prove first the Hadamard differentiability of the functional T_1 . Let (g_t) be a sequence of cadlag functions with bounded variations such that $\|g_t - g\|_\infty \rightarrow 0$, as $t \rightarrow 0$, where g is a cadlag function. For any non negative r , we consider

$$\begin{aligned} \frac{T_1(F_j + tg_t)[r] - T_1(F_j)[r]}{t} &= \frac{(F_j + tg_t)(r) - (F_j + tg_t)(-r) - F_j(r) + F_j(-r)}{t} \\ &= \frac{tg_t(r) - tg_t(-r)}{t} = g_t(r) - g_t(-r) \rightarrow g(r) - g(-r), \end{aligned}$$

since $\|g_t - g\|_\infty \rightarrow 0$, as $t \rightarrow 0$. The Hadamard differential of T_1 at g is given by :

$$(DT_1(F_j).g)(r) = g(r) - g(-r).$$

By [van der Vaart, 1998, Lemma 21.3], T_2 is Hadamard differentiable. Finally, using the Chain rule [van der Vaart, 1998, Theorem 20.9], we obtain the Hadamard differentiability of T_0 with the following Hadamard differential :

$$DT_0(F_j).g = -\frac{(DT_1(F_j).g)(T_0(F_j))}{(T_1(F_j))'[T_0(F_j)]} = -\frac{g(T_0(F_j)) - g(-T_0(F_j))}{(T_1(F_j))'[T_0(F_j)]}.$$

In view of the last expression, $DT_0(F_j)$ is a continuous function of g and is defined on the whole space of cadlag functions. Thus by [van der Vaart, 1998, Theorem 20.8], we obtain :

$$m(\Phi)\sqrt{n_j} (T_0(F_{n_j}) - T_0(F_j)) = m(\Phi)DT_0(F_j) \left\{ \sqrt{n_j}(F_{n_j} - F_j) \right\} + o_P(1),$$

where $m(\Phi)$ is the constant defined in (5.14). Since $T_0(F_j) = \sigma_j/m(\Phi)$ and $(T_1(F_j))'(r) = 2\sigma_j^{-1}\varphi(r/\sigma_j)$, where φ is the p.d.f of a standard Gaussian random variable, we get

$$\sqrt{n_j} (\widehat{\sigma}_{\text{MAD},j} - \sigma_j) = \frac{\sigma_j}{\sqrt{n_j}} \sum_{i=0}^{n_j-1} \text{IF} \left(\frac{W_{j,i}}{\sigma_j}, \text{MAD}, \Phi \right) + o_P(1)$$

and the expansion (5.16) for $*$ = MAD follows from the classical Delta method applied with $f(x) = x^2$. We end the proof of Proposition 5.3.1 by proving the asymptotic expansion (5.16) for $*$ = CR. We use the same arguments as those used previously. In this case the Hadamard differentiability comes from [Lévy-Leduc et al., 2009, Lemma 1]. \square

The following theorem is an extension of [Arcones, 1994, Theorem 4] to arrays of stationary Gaussian processes in the multidimensional case.

Theorem 5.6.3. *Let $\underline{X}_{J,i} = \{X_{J,i}^{(0)}, \dots, X_{J,i}^{(d)}\}$ be an array of standard stationary Gaussian processes such that for j, j' in $\{0, \dots, d\}$, the vector $\{X_{J,i}^{(j)}, X_{J,i}^{(j')}\}$ has a cross-spectral density $f_j^{(j,j')}$ and a cross-correlation function $\rho_j^{(j,j')}$ defined by $\rho_j^{(j,j')}(k) = \mathbb{E}(X_{J,i}^{(j)} X_{J,i+k}^{(j')})$, for all $k \geq 0$. Assume also that there exists a non increasing sequence $\{u_J\}_{J \geq 1}$ such that u_J tends to zero as J tends to infinity and for all $J \geq 1$,*

$$\sup_{\lambda \in (-\pi, \pi)} |f_j^{(j,j')}(\lambda) - g_\infty^{(j,j')}(\lambda)| \leq u_J, \quad (5.51)$$

where $g_\infty^{(j,j')}$ is a 2π -periodic function which is bounded on $(-\pi, \pi)$ and continuous at the origin. Let h be a function on \mathbb{R} with Hermite rank $\tau \geq 1$ which is either bounded or is a finite linear combination of Hermite polynomials. Let $\beta = \{\beta_0, \dots, \beta_d\}$ in \mathbb{R}^{d+1} and $\mathcal{H} : \mathbb{R}^{d+1} \rightarrow \mathbb{R}$ the real valued function defined by $\mathcal{H}(\mathbf{x}) = \sum_{j=0}^d \beta_j h(x_j)$. Let $\{n_J\}_{J \geq 1}$ be a sequence of integers such that n_J tends to infinity as J tends to infinity. Then

$$\frac{1}{\sqrt{n_J}} \sum_{i=1}^{n_J} \mathcal{H}(\underline{X}_{J,i}) \xrightarrow{d} \mathcal{N}(0, \tilde{\sigma}^2), \quad \text{as } J \rightarrow \infty, \quad (5.52)$$

where

$$\tilde{\sigma}^2 = \lim_{n \rightarrow \infty} \text{Var} \left(\frac{1}{\sqrt{n_J}} \sum_{i=1}^{n_J} \mathcal{H}(\underline{X}_{J,i}) \right) = \sum_{\ell \geq \tau} \frac{c_\ell^2}{\ell!} \sum_{0 \leq j, j' \leq d} \beta_j \beta_{j'} (g_\infty^{(j,j')})^{*\ell}(0).$$

In the previous equality, $c_\ell = \mathbb{E}[h(X)H_\ell(X)]$, where H_ℓ is the ℓ -th Hermite polynomial and X is a standard Gaussian random variable.

The proof of Theorem 5.6.3 follows the same lines as the one of Theorem 5.6.1 and is thus omitted.

Proof of Theorem 5.3.1. Without loss of generality, we set $f^*(0) = 1$. In order to prove (5.20), let us first prove that for $\boldsymbol{\alpha} = (\alpha_0, \dots, \alpha_\ell)$ where the α_i 's are in \mathbb{R} ,

$$\sqrt{n2^{-J_0}2^{-2J_0d}} \sum_{j=0}^{\ell} \alpha_j \left(\widehat{\sigma}_{*,J_0+j}^2(W_{J_0+j,0:n_{J_0+j-1}}) - \sigma_{*,J_0+j}^2 \right) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\alpha}^T \mathbf{U}_*(d) \boldsymbol{\alpha}). \quad (5.53)$$

By Proposition 5.3.1,

$$\begin{aligned} & \sqrt{n2^{-J_0}2^{-2J_0d}} \sum_{j=0}^{\ell} \alpha_j \left(\widehat{\sigma}_{*,J_0+j}^2(W_{J_0+j,0:n_{J_0+j-1}}) - \sigma_{*,J_0+j}^2 \right) \\ &= \sum_{j=0}^{\ell} \frac{\sqrt{n2^{-J_0}2^{-2J_0d}}}{n_{J_0+j}} 2\alpha_j \sigma_{J_0+j}^2 \sum_{i=0}^{n_{J_0+j}-1} \text{IF} \left(\frac{W_{J_0+j,i}}{\sigma_{J_0+j}}, *, \Phi \right) + o_P(1). \end{aligned} \quad (5.54)$$

Thus, proving (5.53) amounts to proving that

$$\frac{2^{-\ell/2} f^*(0) \mathbf{K}(d)}{\sqrt{n_{J_0+\ell}}} \sum_{j=0}^{\ell} 2\alpha_j 2^{2dj+j} \sum_{i=0}^{n_{J_0+j}-1} \text{IF} \left(\frac{W_{J_0+j,i}}{\sigma_{J_0+j}}, *, \Phi \right) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\alpha}^T \mathbf{U}_*(d) \boldsymbol{\alpha}), \quad (5.55)$$

since $\sigma_{J_0+j}^2 \sqrt{n2^{-J_0}2^{-2J_0d}}/n_{J_0+j} \sim 2^{2dj-\ell/2+j} \mathbf{K}(d) f^*(0) / \sqrt{n_{J_0+\ell}}$, as n tends to infinity, by [Moulines et al., 2007b, (29) in Theorem 1]. Note that

$$\begin{aligned} \sum_{i=0}^{n_{J_0+j}-1} \text{IF} \left(\frac{W_{J_0+j,i}}{\sigma_{J_0+j}}, *, \Phi \right) &= \sum_{i=0}^{n_{J_0+\ell}-1} \sum_{v=0}^{2^{\ell-j}-1} \text{IF} \left(\frac{W_{j+J_0, 2^{\ell-j}i+v}}{\sigma_{J_0+j}}, *, \Phi \right) \\ &\quad + \sum_{q=n_{J_0+j}-(T-1)(2^{\ell-j}-1)}^{n_{J_0+j}-1} \text{IF} \left(\frac{W_{j+J_0,q}}{\sigma_{J_0+j}}, *, \Phi \right) \end{aligned}$$

Using the notation: $\beta_j = 2\alpha_j 2^{2dj-\ell/2+j} \mathbf{K}(d) f^*(0)$ and that IF is either bounded or equal to $H_2/2$,

$$\begin{aligned} & \frac{1}{\sqrt{n_{J_0+\ell}}} \sum_{j=0}^{\ell} \beta_j \sum_{i=0}^{n_{J_0+j}-1} \text{IF} \left(\frac{W_{J_0+j,i}}{\sigma_{J_0+j}}, *, \Phi \right) \\ &= \frac{1}{\sqrt{n_{J_0+\ell}}} \sum_{j=0}^{\ell} \beta_j \sum_{i=0}^{n_{J_0+\ell}-1} \sum_{v=0}^{2^{\ell-j}-1} \text{IF} \left(\frac{W_{j+J_0, 2^{\ell-j}i+v}}{\sigma_{J_0+j}}, *, \Phi \right) + o_P(1) \\ &= \frac{1}{\sqrt{n_{J_0+\ell}}} \sum_{i=0}^{n_{J_0+\ell}-1} \mathbf{F}(Y_{J_0,\ell,i}, *) + o_P(1), \end{aligned}$$

where

$$\mathbf{F}(Y_{J_0,\ell,i}, *) = \sum_{j=0}^{\ell} \beta_j \sum_{v=0}^{2^{\ell-j}-1} \text{IF} \left(\frac{W_{j+J_0, 2^{\ell-j}i+v}}{\sigma_{J_0+j}}, *, \Phi \right)$$

and

$$Y_{J_0, \ell, i} = \left(\frac{W_{J_0+\ell, i}}{\sigma_{J_0+\ell}}, \frac{W_{J_0+\ell-1, 2i}}{\sigma_{J_0+\ell-1}}, \frac{W_{J_0+\ell-1, 2i+1}}{\sigma_{J_0+\ell-1}}, \dots, \frac{W_{J_0+j, 2^{\ell-j}i}}{\sigma_{J_0+j}}, \dots, \frac{W_{J_0+j, 2^{\ell-j}i+2^{\ell-j}-1}}{\sigma_{J_0+j}}, \dots, \frac{W_{J_0, 2^\ell i}}{\sigma_{J_0}}, \dots, \frac{W_{J_0, 2^\ell i+2^\ell-1}}{\sigma_{J_0}} \right)^T$$

is a $2^{\ell+1} - 1$ stationary Gaussian vector. By Lemma 5.7.1, \mathbf{F} is of Hermite rank larger than 2. Hence, from Theorem 5.6.3 applied to $\mathcal{H}(\cdot) = \mathbf{F}(\cdot)$, $\underline{\mathbf{X}}_{J, i} = Y_{J_0, \ell, i}$ and $h(\cdot) = \text{IF}(\cdot)$, we get

$$\frac{1}{\sqrt{n_{J_0+\ell}}} \sum_{i=0}^{n_{J_0+\ell}-1} \mathbf{F}(Y_{J_0, \ell, i}, *) \xrightarrow{d} \mathcal{N}(0, \tilde{\sigma}_*^2), \quad (5.56)$$

where $\tilde{\sigma}_*^2 = \lim_{n \rightarrow \infty} n_{J_0+\ell}^{-1} \text{Var} \left(\sum_{i=0}^{n_{J_0+\ell}-1} \mathbf{F}(Y_{J_0, \ell, i}, *) \right)$. By [Moulines et al., 2007b, (26) and (29)] and by using the same arguments as those used in the proof of Proposition 5.3.1, Condition (5.51) of Theorem 5.6.3 holds with $f_j^{(j, j')}(\lambda) = \mathbf{D}_{J_0+j, j-j'}^{(r)}(\lambda; f) / \sigma_{J_0+j} \sigma_{J_0+j'}$ and $g_\infty^{(j, j')} = \mathbf{D}_{\infty, j-j'}^{(r)}(\lambda; d) / K(d)$, where $0 \leq r \leq 2^{j-j'} - 1$ and $\mathbf{D}_{J_0+j, j-j'}(\cdot; f)$ is the cross-spectral density of the stationary between scale process defined in (5.8). Lemma 5.7.5 and [Moulines et al., 2007b, Theorem 1] ensure that $\mathbf{D}_{\infty, j-j'}^{(r)}(\cdot; d)$ is a bounded, continuous and 2π -periodic function.

By using Mehler's formula [Breuer and Major, 1983, Eq. (2.1)] and the expansion of IF onto the Hermite polynomials basis given by: $\text{IF}(x, *, \Phi) = \sum_{p \geq 2} c_p(\text{IF}_*) H_p(x) / p!$, where $c_p(\text{IF}_*) = \mathbb{E}[\text{IF}(X, *, \Phi) H_p(X)]$, H_p being the p th Hermite polynomial, we get

$$\begin{aligned} & \frac{1}{n_{J_0+\ell}} \text{Var} \left(\sum_{i=0}^{n_{J_0+\ell}-1} \mathbf{F}(Y_{J_0, \ell, i}, *) \right) \\ &= \frac{1}{n_{J_0+\ell}} \sum_{j, j'=1}^{\ell} \beta_j \beta_{j'} \sum_{i, i'=0}^{n_{J_0+\ell}-1} \sum_{v=0}^{2^{\ell-j}-1} \sum_{v'=0}^{2^{\ell-j'}-1} \mathbb{E} \left[\text{IF} \left(\frac{W_{J_0+j, 2^{\ell-j}i+v}}{\sigma_{J_0+j}}, *, \Phi \right) \text{IF} \left(\frac{W_{J_0+j', 2^{\ell-j'}i'+v'}}{\sigma_{J_0+j'}}, *, \Phi \right) \right] \\ &= \frac{1}{n_{J_0+\ell}} \sum_{j, j'=1}^{\ell} \beta_j \beta_{j'} \sum_{i=0}^{n_{J_0+j}-1} \sum_{i'=0}^{n_{J_0+j'}-1} \mathbb{E} \left[\text{IF} \left(\frac{W_{J_0+j, i}}{\sigma_{J_0+j}}, *, \Phi \right) \text{IF} \left(\frac{W_{J_0+j', i'}}{\sigma_{J_0+j'}}, *, \Phi \right) \right] + o(1) \\ &= \frac{1}{n_{J_0+\ell}} \sum_{j, j'=1}^{\ell} \beta_j \beta_{j'} \sum_{i=0}^{n_{J_0+j}-1} \sum_{i'=0}^{n_{J_0+j'}-1} \sum_{p \geq 2} \frac{c_p^2(\text{IF}_*)}{p!} \mathbb{E} \left[\frac{W_{J_0+j, i}}{\sigma_{J_0+j}} \frac{W_{J_0+j', i'}}{\sigma_{J_0+j'}} \right]^p + o(1). \quad (5.57) \end{aligned}$$

Without loss of generality, we shall assume in the sequel that $j \geq j'$. (5.57) can be rewritten as follows by using that $i' = 2^{j-j'}q + r$, where $q \in \mathbb{N}$ and $r \in \{0, 1, \dots, 2^{j-j'} - 1\}$

and Eq. (18) in Moulines et al. [2007b]

$$\begin{aligned}
& \frac{1}{n_{J_0+\ell}} \sum_{j,j'=1}^{\ell} \beta_j \beta_{j'} \sum_{i=0}^{n_{J_0+j}-1} \sum_{q=0}^{n_{J_0+j'}-1} \sum_{r=0}^{2^{j-j'}-1} \sum_{p \geq 2} \frac{c_p^2(\text{IF}_*)}{p!} \mathbb{E} \left[\frac{W_{J_0+j,0}}{\sigma_{J_0+j}} \frac{W_{J_0+j',2^{j-j'}(q-i)+r}}{\sigma_{J_0+j'}} \right]^p + o(1) \\
&= \frac{n_{J_0+j}}{n_{J_0+\ell}} \sum_{j,j'=1}^{\ell} \beta_j \beta_{j'} \sum_{|\tau| < n_{J_0+j}} \sum_{r=0}^{2^{j-j'}-1} \sum_{p \geq 2} \frac{c_p^2(\text{IF}_*)}{p!} \left(1 - \frac{|\tau|}{n_{J_0+j}}\right) \mathbb{E} \left[\frac{W_{J_0+j,0}}{\sigma_{J_0+j}} \frac{W_{J_0+j',2^{j-j'}\tau+r}}{\sigma_{J_0+j'}} \right]^p + o(1) \\
&= \frac{n_{J_0+j}}{n_{J_0+\ell}} \sum_{j,j'=1}^{\ell} \beta_j \beta_{j'} \sum_{|\tau| < n_{J_0+j}} \sum_{r=0}^{2^{j-j'}-1} \sum_{p \geq 2} \frac{c_p^2(\text{IF}_*)}{p!} \left(1 - \frac{|\tau|}{n_{J_0+j}}\right) \left(\int_{-\pi}^{\pi} \frac{\mathbf{D}_{J_0+j,j-j'}^{(r)}(\lambda; f) e^{i\lambda\tau}}{\sigma_{J_0+j} \sigma_{J_0+j'}} d\lambda \right)^p + o(1),
\end{aligned}$$

where $\mathbf{D}_{J_0+j,j-j'}(\cdot; f)$ is the cross-spectral density of the stationary between scale process defined in (5.8). We aim at applying Lemma 5.7.6 with f_n, g_n, f and g defined hereafter.

$$f_{n_{J_0+j}}(\tau, p) = \frac{c_p^2(\text{IF}_*)}{p!} \sum_{r=0}^{2^{j-j'}-1} \mathbb{1}\{|\tau| < n_{J_0+j}\} \left(1 - \frac{|\tau|}{n_{J_0+j}}\right) \mathbb{E} \left[\frac{W_{J_0+j,0}}{\sigma_{J_0+j}} \frac{W_{J_0+j',2^{j-j'}\tau+r}}{\sigma_{J_0+j'}} \right]^p.$$

Observe that $|f_{n_{J_0+j}}| \leq g_{n_{J_0+j}}$, where

$$g_{n_{J_0+j}}(\tau, p) = \frac{c_p^2(\text{IF}_*)}{p!} \sum_{r=0}^{2^{j-j'}-1} \mathbb{1}\{|\tau| < n_{J_0+j}\} \left(1 - \frac{|\tau|}{n_{J_0+j}}\right) \mathbb{E} \left[\frac{W_{J_0+j,0}}{\sigma_{J_0+j}} \frac{W_{J_0+j',2^{j-j'}\tau+r}}{\sigma_{J_0+j'}} \right]^2.$$

Using [Moulines et al., 2007b, (26) and (29) in Theorem 1] we get that

$$\lim_{n \rightarrow \infty} \frac{\mathbf{D}_{J_0+j,j-j'}(\lambda; f)}{\sigma_{J_0+j} \sigma_{J_0+j'}} = \frac{2^{d(j-j')}}{\mathbf{K}(d)} \mathbf{D}_{\infty,j-j'}(\lambda; d).$$

This implies that $\lim_{n \rightarrow \infty} f_{n_{J_0+j}}(\tau, p) = f(\tau, p)$ where

$$f(\tau, p) = \frac{c_p^2(\text{IF}_*)}{p!} \sum_{r=0}^{2^{j-j'}-1} \left(\frac{2^{d(j-j')}}{\mathbf{K}(d)} \int_{-\pi}^{\pi} \mathbf{D}_{\infty,j-j'}^{(r)}(\lambda; d) e^{i\lambda\tau} d\lambda \right)^p.$$

Futhermore, $\lim_{n \rightarrow \infty} g_{n_{J_0+j}}(\tau, p) = g(\tau, p)$ where

$$g(\tau, p) = \frac{c_p^2(\text{IF}_*)}{p!} \frac{2^{2d(j-j')}}{\mathbf{K}(d)^2} \left| \int_{-\pi}^{\pi} \mathbf{D}_{\infty,j-j'}(\lambda; d) e^{i\lambda\tau} d\lambda \right|_2^2,$$

and $\|\mathbf{x}\|_2^2 = \sum_{k=1}^r x_k^2$ for $\mathbf{x} = (x_1, \dots, x_r) \in \mathbb{R}^r$. Using (63)-(65) in Moulines et al. [2007b] we get

$$\sum_{p \geq 2} \sum_{\tau \in \mathbb{Z}} g_{n_{J_0+j}}(\tau, p) \longrightarrow \left(\sum_{p \geq 2} \frac{c_p^2(\text{IF}_*)}{p!} \right) \frac{2^{2d(j-j')}}{\mathbf{K}(d)^2} 2\pi \int_{-\pi}^{\pi} |\mathbf{D}_{\infty,j-j'}(\lambda; d)|_2^2 d\lambda, \text{ as } n \rightarrow \infty,$$

Then, with Lemma 5.7.6, we obtain

$$\tilde{\sigma}_*^2 = \sum_{p \geq 2} \frac{c_p^2(\text{IF}_*) (f^*(0))^2}{p! \mathbf{K}(d)^{p-2}} \sum_{j,j'=0}^{\ell} 4\alpha_j \alpha_{j'} 2^{dj(2+p)} 2^{dj'(2-p)+j'} \sum_{\tau \in \mathbb{Z}} \sum_{r=0}^{2^{j-j'}-1} \left(\int_{-\pi}^{\pi} \mathbf{D}_{\infty,j-j'}^{(r)}(\lambda; d) e^{i\lambda\tau} d\lambda \right)^p.$$

□

5.7 Technical Lemmas

Lemma 5.7.1. *Let X be a standard Gaussian random variable. The influence functions IF defined in Proposition 5.3.1 have the following properties*

$$\mathbb{E}[\text{IF}(X, *, \Phi)] = 0, \quad (5.58)$$

$$\mathbb{E}[X\text{IF}(X, *, \Phi)] = 0, \quad (5.59)$$

$$\mathbb{E}[X^2\text{IF}(X, *, \Phi)] \neq 0. \quad (5.60)$$

Proof of Lemma 5.7.1. We only have to prove the result for $*$ = MAD since the result for $*$ = CR follows from [Lévy-Leduc et al., 2009, Lemma 12]. (5.58) comes from $\mathbb{E}(\mathbb{1}_{\{X \leq 1/m(\Phi)\}}) = \mathbb{E}(\mathbb{1}_{\{X \leq \Phi^{-1}(3/4)\}}) = 3/4$ and $\mathbb{E}(\mathbb{1}_{\{X \leq -1/m(\Phi)\}}) = 1/4$, where X is a standard Gaussian random variable. (5.59) follows from $\int_{\mathbb{R}} x \mathbb{1}_{\{x \leq \Phi^{-1}(3/4)\}} \varphi(x) dx - \int_{\mathbb{R}} x \mathbb{1}_{\{x \leq -\Phi^{-1}(3/4)\}} \varphi(x) dx = -\varphi(\Phi^{-1}(3/4)) + \varphi(-\Phi^{-1}(3/4)) = 0$, where φ is the p.d.f. of a standard Gaussian random variable and the fact that $\mathbb{E}(X) = 0$. Let us now compute $\mathbb{E}[X^2\text{IF}(X, \text{MAD}, \Phi)]$. Integrating by parts, we get $\int_{\mathbb{R}} x^2 \mathbb{1}_{\{x \leq \Phi^{-1}(3/4)\}} \varphi(x) dx - 3/4 - \int_{\mathbb{R}} x^2 \mathbb{1}_{\{x \leq -\Phi^{-1}(3/4)\}} \varphi(x) dx + 1/4 = -2\varphi(\Phi^{-1}(3/4))$. Thus, $\mathbb{E}[X^2\text{IF}(X, \text{MAD}, \Phi)] = 2 \neq 0$, which concludes the proof. \square

Lemma 5.7.2. *Let X be a standard Gaussian random variable. The influence functions IF defined in Lemma 5.3.1 have the following properties*

$$\mathbb{E}[\text{IF}^2(X, \text{MAD}, \Phi)] = \frac{m^2(\Phi)}{16\varphi(\Phi^{-1}(3/4)^2)} = 1.3601, \quad (5.61)$$

$$\mathbb{E}[\text{IF}^2(X, \text{CR}, \Phi)] \approx 0.6077. \quad (5.62)$$

Proof of Lemma 5.7.2. Eq (5.62) comes from Rousseeuw and Croux [1993]. Since ,

$$\mathbb{E}[\text{IF}^2(X, \text{MAD}, \Phi)] = \frac{m^2(\Phi)}{4\varphi(\Phi^{-1}(3/4)^2)} \text{Var}(\mathbb{1}_{\{|X| \leq \Phi^{-1}(3/4)\}}),$$

where $\mathbb{1}_{\{|X| \leq \Phi^{-1}(3/4)\}}$ is a Bernoulli random variable with parameter 1/2, (5.61) follows. \square

Lemma 5.7.3.

$$\int_{\mathbb{R}^2} \frac{1}{1 + |\mu - \mu'|} \frac{1}{1 + |\mu|} \frac{1}{1 + |\mu'|} d\mu d\mu' < \infty.$$

Proof of Lemma 5.7.3. Let us set $I = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\mu, \mu') d\mu d\mu'$, with

$$p(\mu, \mu') = \frac{1}{1 + |\mu - \mu'|} \frac{1}{1 + |\mu|} \frac{1}{1 + |\mu'|}.$$

Note that $I = I_1 + I_2 + I_3 + I_4$, where $I_1 = \int_0^{\infty} \int_0^{\infty} p(\mu, \mu') d\mu d\mu'$, $I_2 = \int_0^{\infty} \int_{-\infty}^0 p(\mu, \mu') d\mu d\mu'$, $I_3 = \int_{-\infty}^0 \int_0^{\infty} p(\mu, \mu') d\mu d\mu'$ and $I_4 = \int_{-\infty}^0 \int_{-\infty}^0 p(\mu, \mu') d\mu d\mu'$. It is easy to see that $I_1 = I_4$

and $I_2 = I_3$. Let us now compute I_1 . Using partial fraction decomposition,

$$\begin{aligned} I_1 &= \int_0^\infty \frac{1}{1+\mu'} \left(\int_{\mu'}^\infty \frac{1}{1+\mu-\mu'} \frac{1}{1+\mu} d\mu \right) d\mu' + \int_0^\infty \frac{1}{1+\mu'} \left(\int_0^{\mu'} \frac{1}{1-\mu+\mu'} \frac{1}{1+\mu} d\mu \right) d\mu' \\ &= \int_0^\infty \frac{\log(1+\mu')}{\mu'(1+\mu')} d\mu' + 2 \int_0^\infty \frac{\log(1+\mu')}{(2+\mu')(1+\mu')} d\mu' < \infty, \end{aligned}$$

since in the neighborhood of 0, $\log(1+\mu')/\{\mu'(1+\mu')\} \sim 1/(1+\mu')$, $\log(1+\mu')/\{(2+\mu')(1+\mu')\} \sim \{-1/(1+\mu') + 2/(2+\mu')\}$ and in the neighborhood of ∞ , $\log(1+\mu')/\{\mu'(1+\mu')\} \sim \log(\mu')/\mu'^2$ and $\log(1+\mu')/\{(2+\mu')(1+\mu')\} \sim \log(\mu')/\mu'^2$. Let us now compute I_2 . Using the same arguments as previously, we get

$$I_2 = \int_0^\infty \frac{1}{1+\mu'} \left(\int_0^\infty \frac{1}{1+\mu+\mu'} \frac{1}{1+\mu} d\mu \right) d\mu' = \int_0^\infty \frac{\log(1+\mu')}{\mu'(1+\mu')} d\mu' < \infty.$$

□

Lemma 5.7.4.

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} \frac{1}{1+|\mu+\mu'|} \frac{1}{1+|\mu|} d\mu \right)^2 d\mu' < \infty.$$

Proof of Lemma 5.7.4. Let us set $I = \int_{-\infty}^\infty \left(\int_{-\infty}^\infty p(\mu, \mu') d\mu \right)^2 d\mu'$, with

$$p(\mu, \mu') = \frac{1}{1+|\mu+\mu'|} \frac{1}{1+|\mu|}.$$

Note that $I \leq 2(I_1 + I_2 + I_3 + I_4)$, where $I_1 = \int_0^\infty \left(\int_0^\infty p(\mu, \mu') d\mu \right)^2 d\mu'$, $I_2 = \int_0^\infty \left(\int_{-\infty}^0 p(\mu, \mu') d\mu \right)^2 d\mu'$, $I_3 = \int_{-\infty}^0 \left(\int_0^\infty p(\mu, \mu') d\mu \right)^2 d\mu'$ and $I_4 = \int_{-\infty}^0 \left(\int_{-\infty}^0 p(\mu, \mu') d\mu \right)^2 d\mu'$. It is easy to see that $I_1 = I_4$ and $I_2 = I_3$. Let us now compute I_1 . Using partial fraction decomposition,

$$I_1 = \int_0^\infty \left(\int_0^\infty \frac{1}{1+\mu+\mu'} \frac{1}{1+\mu} d\mu \right)^2 d\mu' = \int_0^\infty \left(\frac{1}{\mu'} \log(1+\mu') \right)^2 d\mu' < \infty$$

since in the neighborhood of 0, $[\log(1+\mu')]^2/\mu'^2 \sim 1$, and in the neighborhood of ∞ , $[\log(1+\mu')]^2/\mu'^2 \sim [\log(\mu')]^2/\mu'^2$. Let us now compute I_2 . Using the same arguments as previously, we get that there exists a positive constant C such that

$$\begin{aligned} I_2 &\leq 2 \int_0^\infty \left(\int_{\mu'}^\infty \frac{1}{1+\mu-\mu'} \frac{1}{1+\mu} d\mu \right)^2 d\mu' + 2 \int_0^\infty \left(\int_0^{\mu'} \frac{1}{1-\mu+\mu'} \frac{1}{1+\mu} d\mu \right)^2 d\mu' \\ &\leq C \int_0^\infty \left[\frac{\log(1+\mu')}{\mu'} \right]^2 d\mu' + C \int_0^\infty \left[\frac{\log(1+\mu')}{(2+\mu')} \right]^2 d\mu' < \infty. \end{aligned}$$

□

Lemma 5.7.5. Let $\mathbf{e}_u(\xi) = 2^{-u/2} [1, e^{-i2^{-u}\xi}, \dots, e^{-i(2^u-1)2^{-u}\xi}]^T$, where $\xi \in \mathbb{R}$. For all $u \geq 0$, each component of the vector

$$\mathbf{D}_{\infty, u}(\lambda; d) = \sum_{l \in \mathbb{Z}} |\lambda + 2l\pi|^{-2d} \mathbf{e}_u(\lambda + 2l\pi) \overline{\widehat{\psi}(\lambda + 2l\pi)} \widehat{\psi}(2^{-u}(\lambda + 2l\pi)),$$

is bounded on $(-\pi, \pi)$, where $\widehat{\psi}$ is defined in (5.2).

Proof of Lemma 5.7.5. We start with the case where $l = 0$. Using (5.3), we obtain that $2^{-u/2}|\lambda|^{-2d}|\widehat{\psi}(\lambda)||\widehat{\psi}(2^{-u}\lambda)| = O(|\lambda|^{2M-2d})$, as $\lambda \rightarrow 0$ hence, (5.5) ensures that $2^{-u/2}|\lambda|^{-2d}|\widehat{\psi}(\lambda)||\widehat{\psi}(2^{-u}\lambda)| = O(1)$. Let $\mathbf{e}_u^{(k)}$ denotes the k -th component of the vector \mathbf{e}_u . For $l \neq 0$, (W-2) ensures that for all λ in $(-\pi, \pi)$ there exists a positive constant C such that $|\widehat{\psi}(\lambda)| \leq C/(1 + |\lambda|)^\alpha$. Then, there exists a positive constant C' such that

$$\sum_{l \in \mathbb{Z}^*} |\lambda + 2\pi l|^{-2d} \overline{\widehat{\psi}(\lambda + 2\pi l)} \widehat{\psi}(2^{-u}(\lambda + 2\pi l)) \mathbf{e}_u^{(k)}(\lambda) \leq C' \sum_{l \in \mathbb{Z}^*} |\lambda + 2\pi l|^{-2d-2\alpha} .$$

If $\lambda = 0$, $\sum_{l \in \mathbb{Z}^*} 1/|2\pi l|^{2d+2\alpha} < \infty$ by (5.5). If $\lambda \neq 0$, then, since $-\pi \leq \lambda \leq \pi$,

$$\sum_{l \in \mathbb{Z}^*} 1/|\lambda + 2\pi l|^{2d+2\alpha} \leq \sum_{l \in \mathbb{Z}^*} 1/|\pi(2l - 1)|^{2d+2\alpha} < \infty$$

by (5.5). □

Lemma 5.7.6. *Let f_n and g_n be two sequences of measurable functions on a measure space $(\Omega, \mathcal{F}, \mu)$ such that for all n $|f_n| \leq g_n$. Assume that $\liminf_{n \rightarrow \infty} g_n$ exists and is equal to g . Assume also that $\int g d\mu = \liminf_{n \rightarrow \infty} \int g_n d\mu$ and $\lim_{n \rightarrow \infty} f_n = f$. Then $\int \liminf_{n \rightarrow \infty} f_n d\mu = \liminf_{n \rightarrow \infty} \int f_n d\mu$.*

Proof of Lemma 5.7.6. Since $f_n = f_n^+ - f_n^-$, where $f_n^+, f_n^- \geq 0$, we assume in the sequel that f_n is non negative. By Fatou's Lemma $\int \liminf_{n \rightarrow \infty} (g_n - f_n) d\mu \leq \liminf_{n \rightarrow \infty} \int (g_n - f_n) d\mu$. Using that $\liminf_{n \rightarrow \infty} g_n = g$ and that $\int g d\mu = \liminf_{n \rightarrow \infty} \int g_n d\mu$, we obtain $\limsup_{n \rightarrow \infty} \int f_n d\mu \leq \int \limsup_{n \rightarrow \infty} f_n d\mu$. By applying Fatou's Lemma to f_n , we obtain $\int \liminf_{n \rightarrow \infty} f_n d\mu \leq \liminf_{n \rightarrow \infty} \int f_n d\mu$. Thus,

$$\int f d\mu = \int \liminf_{n \rightarrow \infty} f_n d\mu \leq \liminf_{n \rightarrow \infty} \int f_n d\mu \leq \limsup_{n \rightarrow \infty} \int f_n d\mu \leq \int \limsup_{n \rightarrow \infty} f_n d\mu = \int f d\mu ,$$

which concludes the proof. □

Bibliography

- P. Abry and D. Veitch. Wavelet analysis of long-range-dependent traffic. *IEEE Trans. Inform. Theory*, 44(1):2–15, 1998. ISSN 0018-9448.
- P. Abry, P. Flandrin, M. S. Taqqu, and D. Veitch. Wavelets for the analysis, estimation and synthesis of scaling data. In K. Park and W. Willinger, editors, *Self-Similar Network Traffic and Performance Evaluation*, pages 39–88, New York, 2000. Wiley (Interscience Division).
- P. Abry, P. Flandrin, M. S. Taqqu, and D. Veitch. Self-similarity and long-range dependence through the wavelet lens. In P. Doukhan, G. Oppenheim, and M. S. Taqqu, editors, *Theory and Applications of Long-range Dependence*, pages 527–556. Birkhäuser, 2003.
- L. Agostinelli, C. Bisaglia. Robust estimation of arfima processes. Technical report, Università Ca' Forscari di Venezia, 2003.
- D. W. K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858, 1991. ISSN 0012-9682.
- D. W. K. Andrews and Y. Sun. Adaptive local polynomial Whittle estimation of long-range dependence. *Econometrica*, 72(2):569–614, 2004. ISSN 0012-9682.
- M. Arcones. Limit theorems for nonlinear functionals of a stationary gaussian sequence of vectors. *Annals of probability*, 22(4):2242–2274, 1994.
- R. Azencott and D.D. Castelle. *Séries d'observations irrégulières Modélisation et Prévision*. Masson, Paris New york Barcelone Mexico Sao Paulo, 1984. ISBN 2-225-80138-X.
- Baillie. Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73(1):5–59, 1996.
- J.-M. Bardet. Testing for the presence of self-similarity of Gaussian time series having stationary increments. *Journal of Time Series Analysis*, 21:497–515, 2000.

- J.-M. Bardet. Statistical study of the wavelet analysis of fractional Brownian motion. *IEEE Trans. Inform. Theory*, 48(4):991–999, 2002. ISSN 0018-9448.
- J.-M. Bardet, G. Lang, E. Moulines, and P. Soulier. Wavelet estimator of long-range dependent processes. *Stat. Inference Stoch. Process.*, 3(1-2):85–99, 2000. ISSN 1387-0874. 19th “Rencontres Franco-Belges de Statisticiens” (Marseille, 1998).
- J. Beran. Statistical methods for data with long-range dependence. *Statistical science.*, Vol, 7(4):404–427, 1992.
- J. Beran. Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *Journal of Royal Statistical Society.*, 57:659–672, 1995.
- Jan Beran. *Statistics for long-memory processes*, volume 61 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York, 1994. ISBN 0-412-04901-5.
- I. Berkes, L. Horvatz, Kokoszka P., and Qi-Man Shao. On discriminating between long-range dependence and change in mean,. *The annals of statistics*, 34(3):1140–1165, 2006.
- R. N. Bhattacharya, Vijay K. Gupta, and Ed Waymire. The hurst effect under trends. *Journal of Applied Probability*, 20(3):649–662, 1983. ISSN 00219002. URL <http://www.jstor.org/stable/3213900>.
- P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. ISBN 0-471-19745-9. A Wiley-Interscience Publication.
- P. Bloomfield. On series representations for linear predictors. 13:226–233, 1985.
- D. C. Boes and J. D. Salas. Nonstationarity of the Mean and the Hurst Phenomenon. *Water Resources Research*, 14:135–143, 1978. doi: 10.1029/WR014i001p00135.
- P. Bondon and W. Palma. A class of antipersistent processes. *Journal of Time Series Analysis*, in press, 2007.
- G.E.P. Box and G.M. Jenkins. Some statistical aspects of adaptative optimization and control. *Journal of Royal Statistic Society*, Ser.B 24:297., 1962.
- G.E.P. Box and G.M. Jenkins. *Time Series Analysis : Forecasting and Control*. Holden day, San Francisco, 1976.
- P. Breuer and P. Major. Central limit theorems for nonlinear functionals of Gaussian fields. *J. Multivariate Anal.*, 13(3):425–441, 1983. ISSN 0047-259X.

- P. J. Brockwell and R. A. Davis. *Time series: theory and methods*. Springer Series in Statistics. Springer-Verlag, New York, 1987. ISBN 0-387-96406-1.
- P. J. Brockwell and R. A. Davis. *Introduction to time series analysis*. Texts in Statistics. Springer-Verlag N.Y, 1996.
- B. E. Brodsky and B. S. Darkhovsky. *Non-parametric statistical diagnosis*, volume 509 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht, 2000. ISBN 0-7923-6328-0. Problems and methods.
- P. Carmona, F. Petit, J. Pitman, and M. Yor. On the laws of homogeneous functionals of the Brownian bridge. *Studia Sci. Math. Hungar.*, 35(3-4):445–455, 1999. ISSN 0081-6906.
- W. Chan. A note on time series model specification in the presence of outliers. *Journal of Applied Statistics*, 19:117–124, 1992.
- W. Chan. Outliers and financial time series modelling: a cautionary note. *Mathematics and Computers in Simulation*, 39:425–430, 1995.
- P. Chareka, F. Matarise, and R. Turner. A test for additive outliers applicable to long-memory time series. *Journal of Economic Dynamics & Control*, 30:595–621, 2006.
- J.-F. Coeurjolly. Bahadur representation of sample quantiles for functional of Gaussian dependent sequences under a minimal assumption. *Statist. Probab. Lett.*, 78(15):2485–2489, 2008a. ISSN 0167-7152. doi: 10.1016/j.spl.2008.02.037. URL <http://dx.doi.org/10.1016/j.spl.2008.02.037>.
- J.-F. Coeurjolly. Hurst exponent estimation of locally self-similar Gaussian processes using sample quantiles. *Ann. Statist.*, 36(3):1404–1434, 2008b. ISSN 0090-5364. doi: 10.1214/009053607000000587. URL <http://dx.doi.org/10.1214/009053607000000587>.
- Jean-François Coeurjolly. Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. *Stat. Inference Stoch. Process.*, 4(2):199–227, 2001. ISSN 1387-0874. doi: 10.1023/A:1017507306245. URL <http://dx.doi.org/10.1023/A:1017507306245>.
- Jean-François Coeurjolly and Jacques Istas. Cramèr-Rao bounds for fractional Brownian motions. *Statist. Probab. Lett.*, 53(4):435–447, 2001. ISSN 0167-7152. doi: 10.1016/S0167-7152(00)00197-8. URL [http://dx.doi.org/10.1016/S0167-7152\(00\)00197-8](http://dx.doi.org/10.1016/S0167-7152(00)00197-8).

- A. Cohen. *Numerical analysis of wavelet methods*, volume 32 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 2003. ISBN 0-444-51124-5.
- P. Craigmile, D. Percival, and P. Guttorp. Wavelet-based parameter estimation for polynomial trend contaminated fractionally differenced processes. To appear, *IEEE Trans. Signal Process.*, 2005.
- S. Csörgő and J. Mielniczuk. The empirical process of a short-range dependent stationary sequence under Gaussian subordination. *Probab. Theory Related Fields*, 104(1):15–25, 1996. ISSN 0178-8051.
- R. Dahlhaus. Efficient parameter estimation for self-similar processes. *Ann. Statist.*, 17(4):1749–1766, 1989. ISSN 0090-5364.
- S.J. Deutsch, J.E. Richards, and J. Swain. Effects of a single outlier on arma identification. *Communications in Statistics: Theory and Methods*, 19:2207–2227, 1990.
- F. X. Diebold and A. Inoue. Long memory and regime switching. *J. Econometrics*, 105(1):131–159, 2001. ISSN 0304-4076.
- Z. Ding, C. W. J. Granger, and R. F. Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1):83–106, 1993.
- David L. Donoho and Iain M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994. ISSN 0006-3444. doi: 10.1093/biomet/81.3.425. URL <http://dx.doi.org/10.1093/biomet/81.3.425>.
- P. Doukhan, G. Oppenheim, and Taqqu M. S. *Theory and Applications of long-range dependence*. Birkhauser Boston., 2003. ISBN 0-8176-4168-8.
- G. Faÿ, E. Moulines, F. Roueff, and M. Taqqu. Estimation of long-memory : Fourier versus wavelets. *Journal of Econometrics*, 151(2):159–177, 2009.
- P. Flandrin. Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. Inform. Theory*, 38(2, part 2):910–917, 1992. ISSN 0018-9448.
- R. Fox and M. S. Taqqu. Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.*, 14(2):517–532, 1986. ISSN 0090-5364.
- R. Fox and M. S. Taqqu. Central limit theorems for quadratic forms in random variables having long-range dependence. *Probab. Theory Related Fields*, 74(2):213–240, 1987. ISSN 0178-8051.

- R. Gençay, F. Selçuk, and B. Whitcher. *An introduction to wavelets and other filtering methods in finance and economics*. Academic Press Inc., San Diego, CA, 2002.
- J. Geweke and S. Porter-Hudak. The estimation and application of long memory time series models. *J. Time Ser. Anal.*, 4:221–238, 1983.
- L. Giraitis, P. Kokoszka, R. Leipus, and G. Teyssière. Rescaled variance and related tests for long memory in volatility and levels. *Journal of econometrics*, 112(4):265–294, 2003. ISSN 0090-5364.
- I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series and Products*. Academic, San Diego, CA, 2000.
- C. W. Granger and N. Hyung. Occasional structural breaks and long memory. *Journal of Empirical Finance*, 11:99–114, 1999.
- C. W. J. Granger and Z. Ding. Varieties of long memory models. *J. Econometrics*, 73(1):61–77, 1996. ISSN 0304-4076.
- C. W. J. Granger and R. Joyeux. An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.*, 1(1):15–29, 1980.
- J. Hidalgo and P. M. Robinson. Testing for structural change in a long-memory environment. *J. Econometrics*, 70(1):159–174, 1996. ISSN 0304-4076.
- J. R. Hosking. Fractional differencing. *Biometrika*, 68:165–176, 1981.
- Peter J. Huber. *Robust statistics*. John Wiley & Sons Inc., New York, 1981. ISBN 0-471-41805-6. Wiley Series in Probability and Mathematical Statistics.
- H. E. Hurst. Long term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116:770–779, 1951.
- C. M. Hurvich and B. K. Ray. Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *J. Time Ser. Anal.*, 16(1):17–41, 1995.
- C. M. Hurvich, E. Moulines, and P. Soulier. The FEXP estimator for potentially nonstationary linear time series. *Stoch. Proc. App.*, 97(2):307–340, 2002.
- C. M. Hurvich, G. Lang, and P. Soulier. Estimation of long memory in the presence of a smooth nonparametric trend. *J. Amer. Statist. Assoc.*, 100(471):853–871, 2005. ISSN 0162-1459.
- C. Inclan and G. C. Tiao. Use of cumulative sums of squares for retrospective detection of changes of variance. *American Statistics*, 89(427):913–923, 1994. ISSN 0012-9682.

- I. M. Johnstone and B. W. Silverman. Wavelet threshold estimators for data with correlated noise. *J. Roy. Statist. Soc. Ser. B*, 59(2):319–351, 1997. ISSN 0035-9246.
- L. M. Kaplan and C.-C. J. Kuo. Fractal estimation from noisy data via discrete fractional Gaussian noise (DFGN) and the Haar basis. *IEEE Trans. Signal Process.*, 41(12):3554–3562, 1993.
- J. Kiefer. K -sample analogues of the Kolmogorov-Smirnov and Cramér-V. Mises tests. *Ann. Math. Statist.*, 30:420–447, 1959. ISSN 0003-4851.
- V. Klemes. The Hurst Phenomenon: A Puzzle? *Water Resources Research*, 10:675–688, 1974. doi: 10.1029/WR010i004p00675.
- P.S. Kokoszka and M.S. Taqqu. Fractional arima with stable innovations. *Stochastic process and their Applications*, 6:19–47, 1995.
- O. Kouamo, E. Moulines, and F. Roueff. Test for homogeneity of variance in wavelet domain. *Lecture Notes in Statistics, Springer Verlag*, pages 175–205, 2010.
- H. Künsch. Statistical aspects of self-similar processes. In *Proceedings of the 1st World Congress of the Bernoulli Society, Vol. 1 (Tashkent, 1986)*, pages 67–74, Utrecht, 1987. VNU Sci. Press.
- A.J. Lawrence and N.T. Kottegoda. Stochastic modelling of riverflow times series. *Journal of Royal Statistic Society*, A(140):1–41, 1977.
- C. Lévy-Leduc, H. Boistard, E. Moulines, S. Taqqu, M, and R. Valderio. Robust estimation of the scale and of the autocovariance function of gaussian short and long range dependent processes. To appear in *Journal of Time Series Analysis*, 2009.
- Y. Ma and M. Genton. Highly robust estimation of the auto-covariance function. *Journal of Time Series Analysis*, 21(6):663–684, 2000.
- S. Mallat. *A wavelet tour of signal processing*. Academic Press Inc., San Diego, CA, 1998. ISBN 0-12-466605-1.
- B. Mandelbrot. Une classe processus stochastiques homothétiques à soi; application à la loi climatologique H. E. Hurst. *C. R. Acad. Sci. Paris*, 260:3274–3277, 1965.
- B. B. Mandelbrot and John W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, 10:422–437, 1968. ISSN 0036-1445.
- B. B. Mandelbrot and J. R. Wallis. Noah joseph and operational hydrology. *Water Resources Research*, 4:909–918, 1968.

- B. B. Mandelbrot and J. R. Wallis. Robustness of the rescaled range r/s in the measurement of noncyclic long-run statistical dependence. *Water Resour. Res.*, 5:967–988, 1969.
- Ricardo A. Maronna, R. Douglas Martin, and Victor J. Yohai. *Robust statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 2006. ISBN 978-0-470-01092-1; 0-470-01092-4. Theory and methods.
- A.I. McLeod and K.W. Hipel. Preservation of the rescaled adjusted range, i. a reassessment of the hurst phenomenon. *Water Resources Res.*, 14:269–273, 1978.
- E. J. McCoy and A. T. Walden. Wavelet analysis and synthesis of stationary long-memory processes. *J. Comput. Graph. Statist.*, 5(1):26–56, 1996. ISSN 1061-8600.
- Y. Meyer. *Wavelet*, Ed. J.M. Combes et al., Springer Verlag, Berlin., 1989.
- T. Mikosch and C. Stărică. Changes of structure in financial time series and the Garch model. *REVSTAT*, 2(1):41–73, 2004. ISSN 1645-6726.
- Fabio Fajardo Molinares, Valdério Anselmo Reisen, and Francisco Cribari-Neto. Robust estimation in long-memory processes under additive outliers. *J. Statist. Plann. Inference*, 139(8):2511–2525, 2009. ISSN 0378-3758. doi: 10.1016/j.jspi.2008.12.014. URL <http://dx.doi.org/10.1016/j.jspi.2008.12.014>.
- E. Moulines and P. Soulier. Broadband log-periodogram regression of time series with long-range dependence. *Ann. Statist.*, 27(4):1415–1439, 1999.
- E. Moulines and P. Soulier. Semiparametric spectral estimation for fractional processes. In *Theory and applications of long-range dependence*, pages 251–301. Birkhäuser Boston, Boston, MA, 2003.
- E. Moulines, F. Roueff, and M. S. Taqqu. Central Limit Theorem for the log-regression wavelet estimation of the memory parameter in the Gaussian semi-parametric context. *Fractals*, 15(4):301–313, 2007a.
- E. Moulines, F. Roueff, and M. S. Taqqu. On the spectral density of the wavelet coefficients of long memory time series with application to the log-regression estimation of the memory parameter. *J. Time Ser. Anal.*, 28(2):157–187, 2007b.
- E. Moulines, F. Roueff, and M.S. Taqqu. A wavelet Whittle estimator of the memory parameter of a non-stationary Gaussian time series. *Ann. Statist.*, 36(4):1925–1956, 2008.

- Whitney K. Newey and Kenneth D. West. A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987. ISSN 0012-9682.
- Whitney K. Newey and Kenneth D. West. Automatic lag selection in covariance matrix estimation. *Rev. Econom. Stud.*, 61(4):631–653, 1994. ISSN 0034-6527.
- W. Palma. *Long Memory-times series Theory and Methods*. John Wiley and Sons Inc., Pontificia Universidad Católica de Chile, 2007.
- C. Park, F. Godtlielsen, M. Taqqu, S. Stoev, and J. S. Marron. Visualization and inference based on wavelet coefficients, SiZer and SiNos. *Comput. Statist. Data Anal.*, 51(12):5994–6012, 2007. ISSN 0167-9473. doi: 10.1016/j.csda.2006.11.037. URL <http://dx.doi.org/10.1016/j.csda.2006.11.037>.
- J. Park and C. Park. Robust estimation of the Hurst parameter and selection of an onset scaling. *Statistica Sinica*, 19:1531–1555, 2009.
- D. B. Percival and A. T. Walden. *Wavelet methods for time series analysis*, volume 4 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2000. ISBN 0-521-64068-7.
- D. B. Percival and A. T. Walden. *Wavelet methods for time series analysis*, volume 4 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2006. ISBN 978-0-521-68508-5; 0-521-68508-7.
- Peter C. B. Phillips and Katsumi Shimotsu. Local Whittle estimation in nonstationary and unit root cases. *Ann. Statist.*, 32(2):656–692, 2004. ISSN 0090-5364. doi: 10.1214/009053604000000139. URL <http://dx.doi.org/10.1214/009053604000000139>.
- J. Pitman and M. Yor. The law of the maximum of a Bessel bridge. *Electron. J. Probab.*, 4:no. 15, 35 pp. (electronic), 1999. ISSN 1083-6489.
- K.W. Potter. Evidence of nonstationarity as a physical explanation of the hurst phenomenon. *Water Resources Research*, 12:1047–1052, 1976.
- A. Ramachandra Rao and G. H. Yu. Detection of nonstationarity in hydrologic time series. *Management Science*, 32(9):1206–1217, 1986. ISSN 00251909. URL <http://www.jstor.org/stable/2631546>.
- P. M. Robinson. Efficient tests of nonstationary hypotheses. *J. Amer. Statist. Assoc.*, 89(428):1420–1437, 1994. ISSN 0162-1459.

- P. M. Robinson. Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, 23:1048–1072, 1995a.
- P. M. Robinson. Gaussian semiparametric estimation of long range dependence. *Ann. Statist.*, 23:1630–1661, 1995b.
- M. Rosenblatt. *Stationary sequences and random fields*. Birkhäuser Boston Inc., Boston, MA, 1985. ISBN 0-8176-3264-6.
- F. Roueff and M. S. Taqqu. Asymptotic normality of wavelet estimators of the memory parameter for linear processes. *Journal of Times Series Analysis*, 30(5):534–558, 2009.
- F. Roueff and R. von Sachs. Locally stationary long memory estimation. Technical report, Institut Telecom, 2009.
- P. Rousseeuw and C. Croux. Alternatives to the median absolute deviation. *Journal of the American Statistical association*, 88(424):1273–1283, 1993.
- G. Samorodnitsky and M. S. Taqqu. *Stable non-Gaussian processes: stochastic models with infinite variance*. Chapman and Hall, 1994.
- H. Shen, Z. Zhu, and T. Lee. Robust estimation of the self-similarity parameter in network traffic using wavelet transform. *Signal Process.*, 87(9):2111–2124, 2007. ISSN 0165-1684. doi: <http://dx.doi.org/10.1016/j.sigpro.2007.02.010>.
- K. Shimotsu and P. C. B. Phillips. Exact local Whittle estimation of fractional integration. *Ann. Statist.*, 33(4):1890–1933, 2005. ISSN 0090-5364.
- K. Shimotsu and P. C. B. Phillips. Local Whittle estimation of fractional integration and some of its variants. *J. Econometrics*, 130(2):209–233, 2006. ISSN 0304-4076.
- R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications With R Examples*, volume 4 of *Springer Text in Statistics*. Springer, Second Edition, 1999. ISBN 978-0-387-29317-2.
- S. Stoev and M. S. Taqqu. Asymptotic self-similarity and wavelet estimation for long-range dependent fractional autoregressive integrated moving average time series with stable innovations. *J. Time Ser. Anal.*, 26(2):211–249, 2005. ISSN 0143-9782. doi: 10.1111/j.1467-9892.2005.00399.x. URL <http://dx.doi.org/10.1111/j.1467-9892.2005.00399.x>.
- S. Stoev, M. Taqqu, C. Park, and J. S. Marron. On the wavelet spectrum diagnostic for hurst parameter estimation in the analysis of internet traffic. *Comput. Netw.*, 48(3):423–445, 2005. ISSN 1389-1286. doi: <http://dx.doi.org/10.1016/j.comnet.2004.11.017>.

- S. Stoev, M. S. Taqqu, C. Park, G. Michailidis, and J. S. Marron. LASS: a tool for the local analysis of self-similarity. *Comput. Statist. Data Anal.*, 50(9):2447–2471, 2006. ISSN 0167-9473.
- G. Strang and T. Nguyen. *Wavelet and filter banks*, volume 4 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1997. ISBN 0-9614088-7-1.
- K. Tanaka. The nonstationary fractional unit root. *Econometric Theory*, 15(4):549–582, 1999. ISSN 0266-4666.
- M. S. Taqqu. Law of iterated logarithm for sums of non-linear functions of gaussian variable that exhibit a long range dependence. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 40:203–238, 1977.
- A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.
- D. Veitch and P. Abry. A wavelet-based joint estimator of the parameters of long-range dependence. *IEEE Trans. Inform. Theory*, 45(3):878–897, 1999. ISSN 0018-9448.
- C. Velasco. Gaussian semiparametric estimation of non-stationary time series. *J. Time Ser. Anal.*, 20(1):87–127, 1999.
- G. Walker. On periodicity in series of related terms . *Proc. R.Soc. Lond.*, Ser. A, 131: 518–532, 1931.
- B. Whitcher, S. D. Byers, P. Guttorp, and Percival D. Testing for homogeneity of variance in time series : Long memory, wavelets and the Nile river. *Water Resources Research*, 38(5), 2002.
- G. W. Wornell and A. V. Oppenheim. Estimation of fractal signals from noisy measurements using wavelets. *IEEE Trans. Signal Process.*, 40(3):611 – 623, March 1992.
- G.W Wornell. Wavelet-based representations for the 1/f family of fractal processes. *Proceedings of the IEEE*, 81:1428–1450, 1993.
- A. M. Yaglom. Correlation theory of processes with random stationary n th increments. *Amer. Math. Soc. Transl. (2)*, 8:87–141, 1958. ISSN 0065-9290.
- G. U. Yule. On a method of investigating periodicities in disturbed series with special reference to Wolfer’s sunspot numbers. *Phil. Trans. R.Soc. Lond.*, A226:267–298, 1927.