

Statistical inference for stationary linear models with tapered data

Mamikon S. Ginovyan*¹ and Artur A. Sahakyan²

¹*Department of Mathematics and Statistics, Boston University, USA*
e-mail: ginovyan@math.bu.edu

²*Department of Mathematics and Mechanics, Yerevan State University, Armenia*
e-mail: sart@ysu.am

Abstract: In this paper, we survey some recent results on statistical inference (parametric and nonparametric statistical estimation, hypotheses testing) about the spectrum of stationary models with tapered data. We also discuss some questions concerning tapered Toeplitz matrices and operators, central limit theorems for tapered Toeplitz type quadratic functionals, and tapered Fejér-type kernels and singular integrals. These are the main tools for obtaining the corresponding results, and also are of interest in themselves. The processes considered will be discrete-time and continuous-time Gaussian, linear or Lévy-driven linear processes with memory.

MSC2020 subject classifications: Primary 62F10, 62F12, 60G10; Secondary 62G05, 62G10, 60G15.

Keywords and phrases: Tapered data, stationary processes, spectral density, parametric and nonparametric estimation, goodness-of-fit test.

Received May 2021.

Contents

1	Introduction	155
2	Preliminaries	157
2.1	The model	157
2.1.1	Second-order (wide-sense) stationary process	157
2.1.2	Linear processes. Existence of spectral density functions	158
2.1.3	Lévy-driven linear process	159
2.1.4	Dependence (memory) structure of the model.	159
3	Tapering, data tapers and the tapered periodogram	163
3.1	Motivation	163
3.2	Data tapers and the tapered periodogram	165
3.3	Benefits of tapering	166
4	Nonparametric estimation problem	168
4.1	Estimation of linear spectral functionals	169
4.2	Asymptotic unbiasedness	170
4.3	Consistency	171

arXiv: [2105.06890](https://arxiv.org/abs/2105.06890)

*Author for correspondence

4.4 Asymptotic normality 172

5 Parametric Estimation Problem 174

5.1 The Whittle estimation procedure 174

5.2 Asymptotic properties of the Whittle estimator 177

6 Goodness-of-fit tests 179

6.1 A Goodness-of-fit test for simple hypothesis 180

6.2 A Goodness-of-fit test for composite hypothesis 181

7 Methods and tools 183

7.1 Approximation of traces of products of Toeplitz matrices and operators. 183

7.2 Central limit theorems for tapered quadratic functionals 185

7.2.1 CLT for Gaussian models 186

7.2.2 CLT for Lévy-driven stationary linear models 187

7.3 Fejér-type kernels and singular integrals 188

Acknowledgments 189

References 189

1. Introduction

Let $\{X(t), t \in \mathbb{U}\}$ be a centered real-valued stationary process with spectral density $f(\lambda)$, $\lambda \in \Lambda$, and covariance function $r(t)$, $t \in \mathbb{U}$. We consider simultaneously the continuous-time (c.t.) case, where $\mathbb{U} = \mathbb{R} := (-\infty, \infty)$, and the discrete-time (d.t.) case, where $\mathbb{U} = \mathbb{Z} := \{0, \pm 1, \pm 2, \dots\}$. The domain Λ of the frequency variable λ is $\Lambda = \mathbb{R}$ in the c.t. case, and $\Lambda := [-\pi, \pi]$ in the d.t. case.

We want to make statistical inferences (parametric and nonparametric estimation, and hypotheses testing) about the spectrum of $X(t)$. In the classical setting, the inferences are based on an observed finite realization \mathbf{X}_T of the process $X(t)$: $\mathbf{X}_T := \{X(t), t \in D_T\}$, where $D_T := [0, T]$ in the c.t. case and $D_T := \{1, \dots, T\}$ in the d.t. case.

A sufficiently developed inferential theory is now available for stationary models based on the standard (non-tapered) data \mathbf{X}_T . We cite merely the following references Anh et al. [4], Avram et al. [6], Casas and Gao [12], Dahlhaus [22], Dahlhaus and Wefelmeyer [27], Dzhaparidze [29, 30], Dzhaparidze and Yaglom [32], Fox and Taquq [33], Gao [34], Gao et al. [35], Ginovyan [36, 37, 40, 42, 41, 44, 46], Giraitis et al. [55], Giraitis and Surgailis [56], Guyon [60], Hannan [61], Has'minskii and Ibragimov [62], Heyde and Dai [63], Heyde and Gay [64], Ibragimov [65, 66], Ibragimov and Khas'minskii [67], Leonenko and Sakhno [70], Millar [72], Osidze [73, 74], Taniguchi [78], Taniguchi and Kakizawa [79], Tsai and Chan [81], Walker [83], Whittle [84], where can also be found additional references.

In the statistical analysis of stationary processes, however, the data are frequently tapered before calculating the statistic of interest, and the statistical inference procedure, instead of the original data \mathbf{X}_T , is based on the *tapered data*: $\mathbf{X}_T^h := \{h_T(t)X(t), t \in D_T\}$, where $D_T := [0, T]$ in the c.t. case and $D_T := \{1, \dots, T\}$ in the d.t. case, and $h_T(t) := h(t/T)$ with $h(t)$, $t \in \mathbb{R}$ being a *taper function*.

The use of data tapers in nonparametric time series was suggested by Tukey [80]. The benefits of tapering the data have been widely reported in the literature (see, e.g., Alomari et al. [1], Brillinger [9], Dahlhaus [18]–[21], [23], Dahlhaus and Künsch [26], Ginovyan [46], Ginovyan and Sahakyan [52], Guyon [60], Janas and von Sachs [69], Ludeña and Lavielle [71], Priestley [76], von Sachs [82], and references therein).

For example, data-tapers are introduced to reduce the so-called ‘leakage effects’, that is, to obtain better estimation of the spectrum of the model in the case where it contains high peaks. Other application of data-tapers is in situations in which some of the data values are missing. Also, the use of tapers leads to bias reduction, which is especially important when dealing with spatial data. In this case, the tapers can be used to fight the so-called ‘edge effects’ (for details see Section 3).

In this paper, we survey some recent results on statistical inference (parametric and nonparametric statistical estimation, and hypotheses testing) about the spectrum of stationary models with tapered data. We also discuss some questions concerning tapered Toeplitz matrices and operators, central limit theorems for tapered Toeplitz type quadratic functionals, and tapered Fejér-type kernels and singular integrals. These are the main tools for obtaining the corresponding results, and also are of interest in themselves. The processes considered will be discrete-time and continuous-time Gaussian, linear or Lévy-driven linear processes with memory.

Some notation and conventions. The following notation and conventions are used throughout the paper.

The symbol ‘:=’ stands for ‘by definition’; c.t. := continuous-time; d.t. := discrete-time; s.d. := spectral density; c.f. := covariance function; CLT := central limit theorem. The symbols ‘ \xrightarrow{P} ’ and ‘ \xrightarrow{d} ’ stand for convergence in probability and in distribution, respectively. The notation $X_T \xrightarrow{d} \eta \sim N(0, \sigma^2)$ as $T \rightarrow \infty$ will mean that the distribution of the random variable X_T tends (as $T \rightarrow \infty$) to the centered normal distribution with variance σ^2 . $\mathbb{E}[\cdot]$:= expectation operator; $\text{tr}[A]$:= trace of an operator (matrix) A ; $\mathbb{I}_A(\cdot)$:= indicator of a set $A \subset \Lambda$; $\text{WN}(0, 1)$:= standard white-noise. The standard symbols \mathbb{N} , \mathbb{Z} and \mathbb{R} denote the sets of natural, integer and real numbers, respectively; $\mathbb{N}_0 := \mathbb{N} \cup 0$. By Λ we denote the frequency domain, that is, $\Lambda := \mathbb{R}$ in the c.t. case, and $\Lambda := [-\pi, \pi]$ in the d.t. case. By $L^p(\mu) := L^p(\Lambda, \mu)$ ($p \geq 1$) we denote the weighted Lebesgue space with respect to the measure μ , and by $\|\cdot\|_{p, \mu}$ we denote the norm in $L^p(\mu)$. In the special case where $d\mu(\lambda) = d\lambda$, we will use the notation L^p and $\|\cdot\|_p$, respectively. The letters C and c with or without indices are used to denote positive constants, the values of which can vary from line to line. Also, in the d.t. case all the considered functions are assumed to be 2π -periodic and periodically extended to \mathbb{R} .

The structure of the paper. The rest of the paper is structured as follows. In Section 2 we specify the model of interest – a stationary process, recall some key notions and results from the theory of stationary processes. In Section 3 we

introduce data tapers and the tapered periodogram, and present some benefits of tapering. In Section 4 we discuss the nonparametric estimation problem. We analyze the asymptotic properties, involving asymptotic unbiasedness, bias rate convergence, consistency, a central limit theorem and asymptotic normality of the empirical spectral functionals. In Section 5 we discuss the parametric estimation problem. We present sufficient conditions for consistency and asymptotic normality of minimum contrast estimator based on the Whittle contrast functional for stationary linear models with tapered data. Section 6 is devoted to the construction of goodness-of-fit tests for testing hypotheses that the hypothetical spectral density of a stationary Gaussian model has the specified form, based on the tapered data. In Section 7 we briefly discuss the methods and tools, used to prove the results stated in Sections 4–6.

2. Preliminaries

In this section we specify the model of interest – a stationary process, recall some key notions and results from the theory of stationary processes.

2.1. The model

2.1.1. Second-order (wide-sense) stationary process

Let $\{X(u), u \in \mathbb{U}\}$ be a centered real-valued second-order (wide-sense) stationary process defined on a probability space (Ω, \mathcal{F}, P) with covariance function $r(t)$, that is, $\mathbb{E}[X(u)] = 0$, $r(u) = \mathbb{E}[X(t+u)X(t)]$, $u, t \in \mathbb{U}$, where $\mathbb{E}[\cdot]$ stands for the expectation operator with respect to measure P . We consider simultaneously the c.t. case, where $\mathbb{U} = \mathbb{R} := (-\infty, \infty)$, and the d.t. case, where $\mathbb{U} = \mathbb{Z} := \{0, \pm 1, \pm 2, \dots\}$. We assume that $X(u)$ is a *non-degenerate process*, that is, $\text{Var}[X(u)] = \mathbb{E}|X(u)|^2 = r(0) > 0$. (Without loss of generality, we assume that $r(0) = 1$). In the c.t. case the process $X(u)$ is also assumed mean-square continuous, that is, $\mathbb{E}[X(t) - X(s)]^2 \rightarrow 0$ as $t \rightarrow s$. This assumption is equivalent to that of the covariance function $r(u)$ be continuous at $u = 0$ (see, e.g., Cramér and Leadbetter [17], Section 5.2).

By the Herglotz theorem in the d.t. case, and the Bochner-Khintchine theorem in the c.t. case (see, e.g., Cramér and Leadbetter [17]), there is a finite measure μ on $(\Lambda, \mathfrak{B}(\Lambda))$, where $\Lambda = \mathbb{R}$ in the c.t. case, and $\Lambda = [-\pi, \pi]$ in the d.t. case, and $\mathfrak{B}(\Lambda)$ is the Borel σ -algebra on Λ , such that for any $u \in \mathbb{U}$ the covariance function $r(u)$ admits the following *spectral representation*:

$$r(u) = \int_{\Lambda} \exp\{i\lambda u\} d\mu(\lambda), \quad u \in \mathbb{U}. \quad (2.1)$$

The measure μ in (2.1) is called the *spectral measure* of the process $X(u)$. The function $F(\lambda) := \mu[-\pi, \lambda]$ in the d.t. case and $F(\lambda) := \mu[-\infty, \lambda]$ in the c.t. case, is called the *spectral function* of the process $X(t)$. If $F(\lambda)$ is absolutely continuous (with respect to Lebesgue measure), then the function $f(\lambda) := dF(\lambda)/d\lambda$

is called the *spectral density* of the process $X(t)$. Notice that if the spectral density $f(\lambda)$ exists, then $f(\lambda) \geq 0$, $f(\lambda) \in L^1(\Lambda)$, and (2.1) becomes

$$r(u) = \int_{\Lambda} \exp\{i\lambda u\} f(\lambda) d\lambda, \quad u \in \mathbb{U}. \quad (2.2)$$

Thus, the covariance function $r(u)$ and the spectral function $F(\lambda)$ (resp. the spectral density function $f(\lambda)$) are equivalent specifications of the second order properties for a stationary process $\{X(u), u \in \mathbb{U}\}$.

2.1.2. Linear processes. Existence of spectral density functions

We will consider here stationary processes possessing spectral density functions. For the following results we refer to Cramér and Leadbetter [17], Doob [28], and Ibragimov and Linnik [68].

- (a) The spectral function $F(\lambda)$ of a d.t. stationary process $\{X(u), u \in \mathbb{Z}\}$ is absolutely continuous (with respect to the Lebesgue measure), $F(\lambda) = \int_{-\pi}^{\lambda} f(x) dx$, if and only if it can be represented as an infinite moving average (MA(∞)):

$$X(u) = \sum_{k=-\infty}^{\infty} a(u-k)\xi(k), \quad \sum_{k=-\infty}^{\infty} |a(k)|^2 < \infty, \quad (2.3)$$

where $\{\xi(k), k \in \mathbb{Z}\} \sim \text{WN}(0,1)$ is a standard white-noise, that is, a sequence of orthonormal random variables.

- (b) The covariance function $r(u)$ and the spectral density $f(\lambda)$ of $X(u)$ are given by formulas:

$$r(u) = \mathbb{E}X(u)X(0) = \sum_{k=-\infty}^{\infty} a(u+k)a(k), \quad (2.4)$$

and

$$f(\lambda) = \frac{1}{2\pi} |\widehat{a}(\lambda)|^2 = \frac{1}{2\pi} \left| \sum_{k=-\infty}^{\infty} a(k) e^{-ik\lambda} \right|^2, \quad \lambda \in [-\pi, \pi]. \quad (2.5)$$

- (c) In the case where $\xi(k)$ is a sequence of Gaussian random variables, the process $X(u)$ is Gaussian.

Similar results hold for c.t. processes. Indeed, the following holds.

- (a) The spectral function $F(\lambda)$ of a c.t. stationary process $\{X(u), u \in \mathbb{R}\}$ is absolutely continuous (with respect to Lebesgue measure), $F(\lambda) = \int_{-\infty}^{\lambda} f(x) dx$, if and only if it can be represented as an infinite continuous moving average:

$$X(u) = \int_{\mathbb{R}} a(u-t) d\xi(t), \quad \int_{\mathbb{R}} |a(t)|^2 dt < \infty, \quad (2.6)$$

where $\{\xi(t), t \in \mathbb{R}\}$ is a process with orthogonal increments and $\mathbb{E}|d\xi(t)|^2 = dt$.

- (b) The covariance function $r(u)$ and the spectral density $f(\lambda)$ of $X(u)$ are given by formulas:

$$r(u) = \mathbb{E}X(u)X(0) = \int_{\mathbb{R}} a(u+x)a(x)dx, \quad (2.7)$$

and

$$f(\lambda) = \frac{1}{2\pi} |\widehat{a}(\lambda)|^2 = \frac{1}{2\pi} \left| \int_{\mathbb{R}} e^{-i\lambda t} a(t) dt \right|^2, \quad \lambda \in \mathbb{R}. \quad (2.8)$$

- (c) In the case where $\xi(t)$ is a Gaussian process, the process $X(u)$ is Gaussian.

2.1.3. Lévy-driven linear process

We first recall that a Lévy process, $\{\xi(t), t \in \mathbb{R}\}$ is a process with independent and stationary increments, continuous in probability, with sample-paths which are right-continuous with left limits (càdlàg) and $\xi(0) = \xi(0-) = 0$. The Wiener process $\{B(t), t \geq 0\}$ and the centered Poisson process $\{N(t) - \mathbb{E}N(t), t \geq 0\}$ are typical examples of centered Lévy processes. A Lévy-driven linear process $\{X(t), t \in \mathbb{R}\}$ is a real-valued c.t. stationary process defined by (2.6), where $\xi(t)$ is a Lévy process satisfying the conditions: $\mathbb{E}\xi(t) = 0$, $\mathbb{E}\xi^2(1) = 1$ and $\mathbb{E}\xi^4(1) < \infty$. In the case where $\xi(t) = B(t)$, $X(t)$ is a Gaussian process (see, e.g., Bai et al. [7]).

The function $a(\cdot)$ in representations (2.3) and (2.6) plays the role of a *time-invariant filter*, and the linear processes defined by (2.3) and (2.6) can be viewed as the output of a linear filter $a(\cdot)$ applied to the process $\xi(t)$, called the innovation or driving process of $X(t)$.

Processes of the form (2.3) and (2.6) appear in many fields of science (economics, finance, physics, etc.), and cover large classes of popular models in time series modeling. For instance, the classical autoregressive moving average (ARMA) models and their continuous counterparts the c.t. autoregressive moving average (CARMA) models are of the form (2.3) and (2.6), respectively, and play a central role in the representations of stationary time series (see, e.g., Brockwell [10], Brockwell and Davis [11]).

2.1.4. Dependence (memory) structure of the model.

In the frequency domain setting, the statistical and spectral analysis of stationary processes requires *two types of conditions* on the spectral density $f(\lambda)$. The first type controls the *singularities* of $f(\lambda)$, and involves the *dependence (or memory) structure* of the process, while the second type – controls the *smoothness* of $f(\lambda)$. The memory structure of a stationary process is essentially a measure of the dependence between all the variables in the process, considering the effect of all correlations simultaneously. Traditionally memory structure has

been defined in the time domain in terms of decay rates of the autocorrelations, or in the frequency domain in terms of rates of explosion of low frequency spectra (see, e.g., Beran et al. [8], Giraitis et al. [55], Guégan [58]). It is convenient to characterize the memory structure in terms of the spectral density function.

We will distinguish the following types of stationary models:

- (a) short memory (or short-range dependent),
- (b) long memory (or long-range dependent),
- (c) intermediate memory (or anti-persistent).

Short-memory models. Much of statistical inference is concerned with *short-memory* stationary models, where the spectral density $f(\lambda)$ of the model is bounded away from zero and infinity, that is, there are constants C_1 and C_2 such that

$$0 < C_1 \leq f(\lambda) \leq C_2 < \infty \quad \text{for all } \lambda \in \Lambda.$$

A typical d.t. short memory model example is the stationary Autoregressive Moving Average (ARMA)(p, q) process $X(t)$ defined to be a stationary solution of the difference equation:

$$\psi_p(B)X(t) = \theta_q(B)\varepsilon(t), \quad t \in \mathbb{Z},$$

where ψ_p and θ_q are polynomials of degrees p and q , respectively, B is the backshift operator defined by $BX(t) = X(t-1)$, and $\{\varepsilon(t), t \in \mathbb{Z}\}$ is a d.t. white noise, that is, a sequence of zero-mean, uncorrelated random variables with variance σ^2 . The covariance $r(k)$ of (ARMA)(p, q) process is exponentially bounded:

$$|r(k)| \leq Cr^{-k}, \quad k = 1, 2, \dots; \quad 0 < C < \infty; \quad 0 < r < 1,$$

and the spectral density $f(\lambda)$ is a rational function (see, e.g., Brockwell and Davis [11], Section 3.1):

$$f(\lambda) = \frac{\sigma^2}{2\pi} \cdot \frac{|\theta_q(e^{-i\lambda})|^2}{|\psi_p(e^{-i\lambda})|^2}. \quad (2.9)$$

A typical c.t. short-memory model example is the stationary c.t. ARMA(p, q) processes, denoted by CARMA(p, q). The spectral density function $f(\lambda)$ of a CARMA(p, q) process $X(t)$ is given by the following formula (see, e.g., Brockwell [10]):

$$f(\lambda) = \frac{\sigma^2}{2\pi} \cdot \frac{|\beta_q(i\lambda)|^2}{|\alpha_p(i\lambda)|^2}, \quad (2.10)$$

where $\alpha_p(z) = z^p - \alpha_p z^{(p-1)} - \dots - \alpha_1$ and $\beta_q(z) = 1 + \beta_1 z + \dots + \beta_q z^q$ are polynomials of degrees p and q , respectively.

Another important c.t. short-memory model is the *Ornstein-Uhlenbeck* process, which is a Gaussian stationary process with covariance function $r(t) = \sigma^2 e^{-\alpha|t|}$ ($t \in \mathbb{R}$), and spectral density

$$f(\lambda) = \frac{\sigma^2}{\pi} \cdot \frac{\alpha^2}{\lambda^2 + \alpha^2}, \quad \alpha > 0, \lambda \in \mathbb{R}. \quad (2.11)$$

Discrete-time long-memory and anti-persistent models. Data in many fields of science (economics, finance, hydrology, etc.), however, is well modeled by stationary processes whose spectral densities are *unbounded* or *vanishing* at some fixed points (see, e.g., Beran et al. [8], Guégan [58], and references therein).

A *long-memory* model is defined to be a stationary process with *unbounded* spectral density, and an *anti-persistent* model – a stationary process with *vanishing* (at some fixed points) spectral density.

In the discrete context, a basic long-memory model is the Autoregressive Fractionally Integrated Moving Average (ARFIMA)(0, d , 0) process $X(t)$ defined to be a stationary solution of the difference equation (see, e.g., Brockwell and Davis [11], Section 13.2):

$$(1 - B)^d X(t) = \varepsilon(t), \quad 0 < d < 1/2,$$

where B is the backshift operator and $\varepsilon(t)$ is a d.t. white noise defined above. The spectral density $f(\lambda)$ of $X(t)$ is given by

$$f(\lambda) = |1 - e^{-i\lambda}|^{-2d} = (2 \sin(\lambda/2))^{-2d}, \quad 0 < \lambda \leq \pi, \quad 0 < d < 1/2. \quad (2.12)$$

Notice that $f(\lambda) \sim c|\lambda|^{-2d}$ as $\lambda \rightarrow 0$, that is, $f(\lambda)$ blows up at $\lambda = 0$ like a power function, which is the typical behavior of a long memory model.

A typical example of an *anti-persistent* model is the ARFIMA(0, d , 0) process $X(t)$ with spectral density specified by (2.12) with $d < 0$, which vanishes at $\lambda = 0$. Note that the condition $d < 1/2$ ensures that $\int_{-\pi}^{\pi} f(\lambda) d\lambda < \infty$, implying that the process $X(t)$ with spectral density (2.12) is well defined because $E|X(t)|^2 = \int_{-\pi}^{\pi} f(\lambda) d\lambda$.

Data can also occur in the form of a realization of a ‘mixed’ short-long-intermediate-memory stationary process $X(t)$. A well-known example of such a process, which appears in many applied problems, is an ARFIMA(p , d , q) process $X(t)$ defined to be a stationary solution of the difference equation:

$$\psi_p(B)(1 - B)^d X(t) = \theta_q(B)\varepsilon(t), \quad d < 1/2,$$

where B is the backshift operator, $\varepsilon(t)$ is a d.t. white noise, and ψ_p and θ_q are polynomials of degrees p and q , respectively. The spectral density $f_X(\lambda)$ of $X(t)$ is given by

$$f_X(\lambda) = |1 - e^{-i\lambda}|^{-2d} f(\lambda), \quad d < 1/2, \quad (2.13)$$

where $f(\lambda)$ is the spectral density of an ARMA(p , q) process, given by (2.9). Observe that for $0 < d < 1/2$ the model $X(t)$ specified by the spectral density (2.13) displays long-memory, for $d < 0$ – intermediate-memory, and for $d = 0$ – short-memory. For $d \geq 1/2$ the function $f_X(\lambda)$ in (2.13) is not integrable, and thus it cannot represent a spectral density of a stationary process. Also, if $d \leq -1$, then the series $X(t)$ is not invertible in the sense that it cannot be used to recover a white noise $\varepsilon(t)$ by passing $X(t)$ through a linear filter (see, e.g., Brockweel and Davis [11]).

Another important long-memory model is the fractional Gaussian noise (fGn). To define the fGn we first introduce the *fractional Brownian motion* (fBm)

$\{B_H(t), t \in \mathbb{R}\}$ with Hurst index H , $0 < H < 1$, defined to be a centered Gaussian H -self-similar process having stationary increments (see, e.g., Samorodnisky and Taqqu [77]). Then the increment process $\{X(k) := B_H(k+1) - B_H(k), k \in \mathbb{Z}\}$, called *fractional Gaussian noise* (fGn), is a d.t. centered Gaussian stationary process with spectral density function:

$$f(\lambda) = c |1 - e^{-i\lambda}|^2 \sum_{k=-\infty}^{\infty} |\lambda + 2\pi k|^{-(2H+1)}, \quad -\pi \leq \lambda \leq \pi, \quad (2.14)$$

where c is a positive constant.

It follows from (2.14) that $f(\lambda) \sim c |\lambda|^{1-2H}$ as $\lambda \rightarrow 0$, that is, $f(\lambda)$ blows up if $H > 1/2$ and tends to zero if $H < 1/2$. Also, comparing (2.12) and (2.14), we observe that, up to a constant, the spectral density of fGn has the same behavior at the origin as ARFIMA(0, d , 0) with $d = H - 1/2$.

Thus, the fGn $\{X(k), k \in \mathbb{Z}\}$ has long-memory if $1/2 < H < 1$ and is anti-persistent if $0 < H < 1/2$. The variables $X(k)$, $k \in \mathbb{Z}$, are independent if $H = 1/2$. For more details we refer to Samorodnisky and Taqqu [77].

Continuous-time long-memory and anti-persistent models. In the continuous context, a basic process which has commonly been used to model long-range dependence is the fractional Brownian motion (fBm) B_H with Hurst index H , defined above, which can be regarded as a Gaussian process having a ‘spectral density’:

$$f(\lambda) = c |\lambda|^{-(2H+1)}, \quad c > 0, \quad 0 < H < 1, \quad \lambda \in \mathbb{R}. \quad (2.15)$$

The form (2.15) can be understood in a generalized sense (see, e.g., Yaglom [85]), since the fBm B_H is a nonstationary process.

A proper stationary model in lieu of fBm is the *fractional Riesz-Bessel motion* (fRBm), introduced in Anh et al. [2], and defined as a c.t. Gaussian process $X(t)$ with spectral density

$$f(\lambda) = c |\lambda|^{-2\alpha} (1 + \lambda^2)^{-\beta}, \quad \lambda \in \mathbb{R}, \quad 0 < c < \infty, \quad (2.16)$$

where the exponents α and β are such that $0 < \alpha < 1$, $\beta > 0$ and $\alpha + \beta > 1/2$. The exponent α determines the long-range dependence, while the exponent β indicates the second-order intermittency of the process (see, e.g., Anh et al. [3] and Gao et al. [35]).

Notice that the process $X(t)$, specified by the spectral density (2.16), is stationary if $0 < \alpha < 1/2$ and is non-stationary with stationary increments if $1/2 \leq \alpha < 1$. Observe also that the function $f(\lambda)$ in (2.16) behaves as $O(|\lambda|^{-2\alpha})$ as $|\lambda| \rightarrow 0$ and as $O(|\lambda|^{-2(\alpha+\beta)})$ as $|\lambda| \rightarrow \infty$. Thus, under the conditions $0 < \alpha < 1/2$, $\beta > 0$ and $\alpha + \beta > 1/2$, the function $f(\lambda)$ in (2.16) is well-defined for both $|\lambda| \rightarrow 0$ and $|\lambda| \rightarrow \infty$ due to the presence of the component $(1 + \lambda^2)^{-\beta}$, $\beta > 0$, which is the Fourier transform of the Bessel potential.

Comparing (2.15) and (2.16), we observe that the spectral density of fBm is the limiting case as $\beta \rightarrow 0$ that of fRBm with Hurst index $H = \alpha - 1/2$.

Another important c.t. long-memory model is the CARFIMA(p, H, q) process. The spectral density function $f(\lambda)$ of a CARFIMA(p, H, q) process $X(t)$

is given by the following formula (see, e.g., Brockwell [10], and Tsai and Chan [81]):

$$f(\lambda) = \frac{\sigma^2}{2\pi} \Gamma(2H + 1) \sin(\pi H) |\lambda|^{1-2H} \frac{|\beta_q(i\lambda)|^2}{|\alpha_p(i\lambda)|^2}, \quad (2.17)$$

where $\alpha_p(z) = z^p - \alpha_p z^{(p-1)} - \dots - \alpha_1$ and $\beta_q(z) = 1 + \beta_1 z + \dots + \beta_q z^q$ are polynomials of degrees p and q , respectively. Notice that for $H = 1/2$, the spectral density given by (2.17) becomes that of the short-memory CARMA(p, q) process, given by (2.10).

3. Tapering, data tapers and the tapered periodogram

In this section we introduce data tapers and the tapered periodogram, and present some benefits of tapering.

3.1. Motivation

Tapering is a technique which aim to improve the accuracy of spectral estimators by making certain preliminary transformation to the original data before the estimation procedure is applied. The main purpose of tapering is to reduce the bias in the spectral estimators by reducing the bias due to the periodogram. It is well known (see, e.g., Brillinger [9], Priestley [76], Dahlhaus [23]) that the periodogram is not a good statistical estimator for unknown spectral density, and it has two major disadvantages. The first one is that the periodogram is not a consistent estimator for unknown spectral density and the periodogram values at different frequencies are nearly uncorrelated. The second important disadvantage of the periodogram is the so-called leakage effect caused by the strong peaks in the spectrum.

To explain how to avoid these disadvantages, assume for simplicity that the underlying process $X(t)$ is a d.t. process with an unknown spectral density f , and we want to estimate f based on the sample $\mathbf{X}_T := \{X(t), t = 1, \dots, T\}$ (in the c.t. case we have similar arguments).

As an estimator of $f(\lambda)$ consider the periodogram $I_T(\lambda)$ of the process $X(t)$:

$$I_T(\lambda) := \frac{1}{2\pi T} \left| \sum_{t=1}^T X(t) e^{-it\lambda} \right|^2. \quad (3.1)$$

Then for the expected value of $I_T(\lambda)$ we have the convolution:

$$\mathbb{E}[I_T(\lambda)] = \int_{-\pi}^{\pi} f(\lambda - \mu) F_T(\mu) d\mu, \quad (3.2)$$

where $F_T(\mu)$ is the classical Fejér kernel:

$$F_T(\mu) := \frac{1}{2\pi T} \frac{\sin^2(T\mu/2)}{\sin^2(\mu/2)}. \quad (3.3)$$

In order to improve the estimator for $f(\lambda)$, we first could try to replace the periodogram $I_T(\lambda)$ by a better nonparametric estimator $\widehat{f}_T(\lambda)$, the smoothed periodogram estimator with a suitable chosen smoothing spectral window $W(\cdot)$:

$$\widehat{f}_T(\lambda) := \int_{-\pi}^{\pi} I_T(\mu + \lambda) W_N(\mu) d\mu, \quad (3.4)$$

where, e.g., $W_N(\mu) = NW(N\mu)$ with $N/T \rightarrow 0$ as $T \rightarrow \infty$.

Although the usual asymptotic theory leads to satisfactory results for this estimator, $\widehat{f}_T(\lambda)$ behaves in certain situations rather badly. It is well known (see, e.g., Priestley [76], p. 558) that if the underlying model is short memory with continuous spectral density, then the periodogram bias is, in general, negligible compared with the bias due to the smoothing window. However, if the spectral window has discontinuities, this is not the case.

The second disadvantage – the leakage effect – is due to properties of the Fejér kernel $F_T(\mu)$. It is known (see, e.g., Dahlhaus [21]) that $F_T(\mu)$ has side peaks of magnitude $O(T^{-1})$ at frequencies $\mu = 2\pi(s + 1/2)/T$ with $\mu \in (0, \pi]$. If in the convolution (3.2) such a side peak is multiplied with a strong peak of the spectral density f this may result in an expectation that is too large, that is, the spectrum is overestimated. The result is that other lower peaks are superimposed and possibly not discovered. Thus, if no data taper is used the estimator is not able to resolve lower peaks of the spectrum. This effect has been called *nonparametric leakage effect*.

Remark 3.1. Due to the fact that the Whittle approximation of the log-likelihood function can be considered as the information divergence between the periodogram and the spectral density, the leakage effect will also transfer to the parametric estimation procedure. This effect has been called *parametric leakage effect* (see, e.g., Dahlhaus [21]).

To avoid the leakage effects and other “bad” effects, discussed below, we taper the original data before the estimation procedure is applied.

The idea behind tapering is to replace the Fejér kernel $F_T(\mu)$ by the more general ‘*tapered kernel*’, and the objective is to choose the taper function in such a way that the tapered kernel has much smaller side lobes than $F_T(\mu)$.

The use of data tapers in nonparametric time series was suggested by Tukey [80]. Then it was discussed by many authors both for parametric and nonparametric estimation procedures. The benefits of tapering the data have been widely reported in the literature (see, e.g., Alomari et al. [1], Anh et al. [5], Brillinger [9], Dahlhaus [18]–[21], [23], Dahlhaus and Künsch [26], Ginovyan [46], Ginovyan and Sahakyan [52], Guyon [60], Janas and von Sachs [69], Ludeña and Lavielle [71], Priestley [76], von Sachs [82], and references therein).

For example, data-tapers are introduced to reduce the above described ‘leakage effects’, that is, to obtain better estimator of the spectrum of the model in the case where it contains strong peaks. Tapering also can be used to reduce the so-called ‘trough effects’, that is, to obtain better estimator of the spectrum in the case where it contains strong troughs. The trough effect was first observed

and discussed by Dahlhaus [21]. Other application of data-tapers is in situations in which some of the data values are missing. Also, the use of tapers leads to bias reduction, which is especially important when dealing with spatial data. In this case, the tapers can be used to fight the so-called ‘edge effects’ (for details see Section 3.3).

3.2. Data tapers and the tapered periodogram

Our inference procedures will be based on the tapered data \mathbf{X}_T^h :

$$\mathbf{X}_T^h := \{h_T(t)X(t), t \in D_T\}, \quad (3.5)$$

where $D_T := [0, T]$ in the c.t. case and $D_T := \{1, \dots, T\}$ in the d.t. case, and

$$h_T(t) := h(t/T) \quad (3.6)$$

with $h(t)$, $t \in \mathbb{R}$ being a *taper function* to be specified below.

Note. The case $h(t) = \mathbb{I}_{[0,1]}(t)$, where $\mathbb{I}_{[0,1]}(\cdot)$ denotes the indicator of the segment $[0, 1]$, will be referred to as the *non-tapered* case.

For $k \in \mathbb{N} := \{1, 2, \dots\}$, denote by $H_{k,T}(\lambda)$ the *tapered Dirichlet type kernel*, defined by

$$H_{k,T}(\lambda) := \begin{cases} \sum_{t=1}^T h_T^k(t) e^{-i\lambda t} & \text{in the d.t. case,} \\ \int_0^T h_T^k(t) e^{-i\lambda t} dt & \text{in the c.t. case,} \end{cases} \quad (3.7)$$

and put

$$H_{k,T} := H_{k,T}(0). \quad (3.8)$$

Define the finite Fourier transform of the tapered data (3.5):

$$d_T^h(\lambda) := \begin{cases} \sum_{t=1}^T h_T(t)X(t)e^{-i\lambda t} & \text{in the d.t. case,} \\ \int_0^T h_T(t)X(t)e^{-i\lambda t} dt & \text{in the c.t. case.} \end{cases} \quad (3.9)$$

and the tapered periodogram $I_T^h(\lambda)$ of the process $X(t)$:

$$\begin{aligned} I_T^h(\lambda) &:= \frac{1}{C_T} d_T^h(\lambda) d_T^h(-\lambda) \\ &= \begin{cases} \frac{1}{C_T} \left| \sum_{t=1}^T h_T(t)X(t)e^{-i\lambda t} \right|^2 & \text{in the d.t. case,} \\ \frac{1}{C_T} \left| \int_0^T h_T(t)X(t)e^{-i\lambda t} dt \right|^2 & \text{in the c.t. case.} \end{cases} \end{aligned} \quad (3.10)$$

where

$$C_T := 2\pi H_{2,T}(0) \neq 0. \quad (3.11)$$

Notice that for non-tapered case ($h(t) = \mathbb{I}_{[0,1]}(t)$), we have $C_T = 2\pi T$.

Throughout the paper, we will assume that the taper function $h(\cdot)$ satisfies the following assumption.

Assumption 3.1. The taper $h : \mathbb{R} \rightarrow \mathbb{R}$ is a continuous nonnegative function of bounded variation and of bounded support $[0, 1]$, such that $H_k \neq 0$, where

$$H_k := \lim_{T \rightarrow \infty} (1/T)H_{k,T}, \text{ and } H_{k,T} \text{ is as in (3.8)}. \quad (3.12)$$

Observe that in the c.t. case we have $H_k = \int_0^1 h^k(t)dt$.

Remark 3.2. The data taper $h(t)$ normally has a maximum at $t = 1/2$ and decreases smoothly to zero as t tends to 0 or 1. For the d.t. case, an example of a taper function $h(t)$ satisfying Assumption 3.1 is the Tukey-Hanning taper function $h(t) = 0.5(1 - \cos(\pi t))$ for $t \in [0, 1]$. For the c.t. case, a simple example of a taper function $h(t)$ satisfying Assumption 3.1 is the function $h(t) = 1 - t$ for $t \in [0, 1]$. See also Example 3.1 below.

3.3. Benefits of tapering

The leakage and trough effects. Dahlhaus' workshop. Considering the parametric estimation problem, Dahlhaus [19] proved that the tapered Whittle estimators are asymptotically normal with the same increase of the asymptotic variance as in the nonparametric case. Thus, asymptotic results do not describe the effects observed in the small sample situation. On the other hand, due to extreme technical complexity, exact calculations for finite sample size T are only possible in very special situations, e.g., for AR(1) processes.

To handle these difficulties, in two important papers Dahlhaus [21, 23], the author introduced and developed special asymptotic models that cover in some sense the small sample situation both for parametric and nonparametric cases. In this asymptotic setting, Dahlhaus demonstrated the benefits of tapering in a rigorous way. In particular, using the proposed models, Dahlhaus explained theoretically the leakage effect for the parametric Whittle estimator and for nonparametric spectral estimator and showed that this effect can be removed by applying a suitable data taper function. Also, he proved, contrary to widespread conjectures (see, e.g., Brillinger [9], p.151 and Priestley [76], p.272), that data tapers do not only reduce the bias, but may also reduce the variance of the corresponding estimators (see Dahlhaus [23], p.161).

The idea behind Dahlhaus' approach is the following. He required that the estimators behave well for an increasing number of processes if the sample size increases. For example, considering the equation (3.2), and taking into account that the side peaks of the Fejér kernel $F_T(\mu)$ for $\mu \in (0, \pi]$ is of order $O(T^{-1})$, he observed that the leakage effect disappears asymptotically. To preserve this small sample effect, Dahlhaus suggested to let the magnitude of the main peak of the spectrum increase with sample size T and studied the properties of the estimators concerning the lower peaks of the spectrum. For instance, considering the spectral density of an AR(p) process:

$$f(\lambda) = \frac{\sigma^2}{2\pi} \cdot \prod_{k=1}^p |1 - z_k e^{-i\lambda}|^{-2}, \quad (3.13)$$

where z_k are the roots of the process, and using simulation results, Dahlhaus observed that it is quite easy to estimate the strongest peak, that is, the roots closest to the unit circle, and concluded that it should be reasonable to assume that these roots are known, whereas the other roots should be estimated.

Based on these arguments, Dahlhaus [21] introduced the following parametric estimation model: write the spectral density $f_\theta(\lambda)$ in the form of a product:

$$f_\theta(\lambda) = g_T f_\theta^*(\lambda) \quad (3.14)$$

with a ‘strong peak’ part g_T (whose peaks may increase with sample size T) and a part $f_\theta^*(\lambda)$, which is independent of T and whose parameter θ has to be estimated.

In order to investigate the properties of the tapered Whittle estimators Dahlhaus [21] introduced an increasing class of stochastic processes \mathfrak{X} (see Definition 2.2 in Dahlhaus [21]), and imposed additional smoothness conditions on the taper function $h(\cdot)$. Specifically, he introduced the class of taper functions $h_T(t) := h_T(t/T)$ of degree (k, κ) , where $k \in \mathbb{N}_0 := \{0, 1, 2, \dots\}$ and $\kappa \in [0, 1/2)$ (see Definition 5.1 in Dahlhaus [21]). Denote this class by $D(k, \kappa)$. Observe that in the non-tapered case $h_T(t) := \mathbb{I}_{[0,1]}(t/T)$ the degree is $(0, 0)$. The following commonly used taper function belongs to the class $D(k, \kappa)$.

Example 3.1 (Polynomial taper). Consider the function:

$$h_\rho(x) := \begin{cases} 4^k (x/\rho)^k (1 - x/\rho)^k & x \in [0, \rho/2), \\ 1 & x \in [\rho/2, 1/2], \\ h_\rho(1 - x) & x \in (1/2, 1]. \end{cases} \quad (3.15)$$

In Dahlhaus [21] (see also Dahlhaus [23]) it was shown that if ρ is fixed, then the taper $h_T(t)$ is of degree $(k, 0)$, and if $\rho = \rho_T = T^{-\kappa/(k+1)}$, then the taper $h_T(t)$ is of degree (k, κ) .

In Section 6 of Dahlhaus [21], the behavior of the tapered Whittle estimators was studied for double indexed sequences of processes $X(t, T) \in \mathfrak{X}$ with spectral density of the form (3.14), and taper functions $h_T(t)$ from the class $D(k, \kappa)$. Specifically, it was proved that if $X(t, T)$ is a sequence of stationary Gaussian processes and $h_T(t)$ is a taper function from the class $D(k, \kappa)$ satisfying $\lim_{T \rightarrow \infty} (TH_{4,T})/H_{2,T}^2 = 1$ (e.g., the polynomial taper in Example 3.1), then the tapered Whittle estimator is Fisher efficient, and thus, it is equivalent to the exact maximum likelihood estimator (for details see Theorem 6.3 in Dahlhaus [21]). Moreover, it was proved that the classical non-tapered Whittle estimator may even be inconsistent in the considered model (Dahlhaus [21], Theorem 7.1). The same holds if strong troughs (values close to zero caused, e.g., by MA roots close to the unit circle) are present in the spectrum (Dahlhaus [21], Theorem 7.2). This effect has been called the *trough effect*.

Remark 3.3. Observe that the Dahlhaus’ model is also of importance for the study of long-range dependence. Recall that (see Section 2.1.4) long-range dependence (or long memory), is normally assumed that the spectral density $f(\lambda)$

has a singularity of the form $|\lambda|^{-1+\varepsilon}$ in the neighborhood of zero. Dahlhaus' model includes spectral densities of the form (cf. (2.12)):

$$f(\lambda) = |1 - (1 - 1/T)e^{-i\lambda}|^{-2s} f_{\theta}^*(\lambda), \quad (3.16)$$

which for large T is approximately of the form $|\lambda|^{-2s} f_{\theta}^*(\lambda)$ for small λ (see Dahlhaus [21], p. 816).

The bias and the edge effect. Edge effects are a serious problem in spatial statistics because the number of boundary points increases with the dimension. Assume that a d.t. d -dimensional centered random field $\{X(t), t \in \mathbb{Z}^d\}$ is observed on a rectangle $P_T := \{1, \dots, T_1\} \times \dots \times \{1, \dots, T_d\}$ of sample size $T := \prod_{k=1}^d T_k$. Considering the usual parametric Whittle estimator, Guyon [59] observed that if the estimator of the underlying unknown parameter is based on the sample covariances $c_T(k)$, then there is a bias due to boundary (edge effect). If $T_k \rightarrow \infty$ with the same speed in all directions, this bias is of order $O(T^{-1/d})$, which is for $d \geq 2$ of the same order or a higher order as the standard deviation of $c_T(k)$ which is usually $O(T^{-1/2})$. To remedy this edge effect, Guyon [59] proposed the use of an unbiased modification of sample covariances, $c_T^*(k)$, and gave a \sqrt{T} -consistent and asymptotically normal estimator of the underlying parameter.

Dahlhaus and Künsch [26] observed that the Guyon estimator $c_N^*(k)$ has some unpleasant properties. First, it is not always positive-definite so that the spectral estimates based on $c_T^*(k)$ may be negative. The second disadvantage of $c_T^*(k)$ is its large variance, particularly for larger lags. They showed that these disadvantages of $c_N^*(k)$ can be avoided and nevertheless get an asymptotically negligible bias by using the data tapers (for details see Dahlhaus and Künsch [26], Theorem 1). In Anh et al. [5] was discussed the bias and edge effect for estimators of spectral functionals both for d.t. and c.t. cases. It was shown that under some smoothness conditions imposed on the taper function h the bias is of order $O(T^{-2})$ (for details see Anh et al. [5], Theorem 2.1). Notice that the edge effect for Whittle estimator for a class of long-memory stationary Gaussian fields was discussed in Ludeña and Lavielle [71].

4. Nonparametric estimation problem

Suppose we observe a finite realization $\mathbf{X}_T := \{X(u), 0 \leq u \leq T$ (or $u = 1, \dots, T$ in the d.t. case) of a centered stationary process $X(u)$ with an *unknown* spectral density function $f(\lambda)$, $\lambda \in \Lambda$. We assume that $f(\lambda)$ belongs to a given (infinite-dimensional) class $\mathcal{F} \subset L^p := L^p(\Lambda)$ ($p \geq 1$) of spectral densities possessing some specified smoothness properties. The problem is to estimate the value $J(f)$ of a given functional $J(\cdot)$ at an *unknown* 'point' $f \in \mathcal{F}$ on the basis of an observation \mathbf{X}_T , and investigate the asymptotic (as $T \rightarrow \infty$) properties of the suggested estimators, depending on the dependence structure of the model $X(u)$ and the smoothness structure of the 'parametric' set $\mathcal{F} \subset L^p(\Lambda)$ ($p \geq 1$).

Linear and non-linear functionals of the periodogram play a key role in the parametric estimation of the spectrum of stationary processes, when using the

minimum contrast estimation method with various contrast functionals (see, e.g., Anh et al. [4], Dzharidze [30], Ginovyan [36, 38, 39, 42, 43], Guyon [60], Leonenko and Sakhno [70], Taniguchi and Kakizawa [79], and references therein). In this section, we review the asymptotic properties, involving asymptotic unbiasedness, bias rate convergence, consistency, a central limit theorem and asymptotic normality of the empirical spectral functionals based on the tapered data. Some of these properties were discussed and proved in Ginovyan and Sahakyan [52, 53]. For non-tapered case, these properties were established in the papers Ginovyan [38, 42]. The results stated in this section are used to prove consistency and asymptotic normality of the minimum contrast estimator based on the Whittle contrast functional for stationary linear models with tapered data (see Section 5). Here we follow the papers Ginovyan [39, 42, 43], and Ginovyan and Sahakyan [52, 53].

4.1. Estimation of linear spectral functionals

We are interested in the nonparametric estimation problem, based on the tapered data (3.5), of the following linear spectral functional:

$$J = J(f, g) := \int_{\Lambda} f(\lambda)g(\lambda)d\lambda, \tag{4.1}$$

where $g(\lambda) \in L^q(\Lambda)$, $1/p + 1/q = 1$.

As an estimator J_T^h for functional $J(f)$, given by (4.1), based on the tapered data (3.5), we consider the averaged tapered periodogram (or a simple ‘plug-in’ statistic), defined by

$$J_T^h = J(I_T^h, g) := \int_{\Lambda} I_T^h(\lambda)g(\lambda)d\lambda, \tag{4.2}$$

where $I_T^h(\lambda)$ is the tapered periodogram of the process $X(t)$ given by (3.10). Denote

$$Q_T^h := \begin{cases} \sum_{t=1}^T \sum_{s=1}^T \widehat{g}(t-s)h_T(t)h_T(s)X(t)X(s) & \text{in the d.t. case,} \\ \int_0^T \int_0^T \widehat{g}(t-s)h_T(t)h_T(s)X(t)X(s) dt ds & \text{in the c.t. case,} \end{cases} \tag{4.3}$$

where $\widehat{g}(t)$ is the Fourier transform of function $g(\lambda)$:

$$\widehat{g}(t) := \int_{\Lambda} e^{i\lambda t}g(\lambda)d\lambda, \quad t \in \Lambda. \tag{4.4}$$

In view of (3.10) and (4.2) – (4.4) we have

$$J_T^h = C_T^{-1}Q_T^h, \tag{4.5}$$

where C_T is as in (3.11). We will refer to $g(\lambda)$ and to its Fourier transform $\widehat{g}(t)$ as a *generating function* and *generating kernel* for the functional J_T^h , respectively.

Thus, to study the asymptotic properties of the estimator J_T^h , we have to study the asymptotic distribution (as $T \rightarrow \infty$) of the tapered Toeplitz type quadratic functional Q_T^h given by (4.3) (for details see Section 7.2).

4.2. Asymptotic unbiasedness

We begin with the following assumption.

Assumption 4.1. The function

$$\Psi(u) = \int_{\Lambda} f(v)g(u+v) dv \quad (4.6)$$

belongs to $L^1(\Lambda) \cap L^2(\Lambda)$ and is continuous at $u = 0$.

Theorem 4.1. Let the functionals $J := J(f, g)$ and $J_T^h := J(I_T^h, g)$ be defined by (4.1) and (4.2), respectively. Then under Assumptions 3.1 and 4.1 the statistic J_T^h is an asymptotically unbiased estimator for $J(f)$, that is, the following relation holds:

$$\lim_{T \rightarrow \infty} [\mathbb{E}(J_T^h) - J] = 0. \quad (4.7)$$

Remark 4.1. Using Hölder inequality, it can easily be shown that if $f \in L^1(\Lambda) \cap L^{p_1}(\Lambda)$ and $g \in L^1(\Lambda) \cap L^{p_2}(\Lambda)$ with $1 \leq p_1, p_2 \leq \infty$, $1/p_1 + 1/p_2 \leq 1$, then the relation (4.7) is satisfied.

Under additional smoothness conditions on functions $f(\lambda)$ and $g(\lambda)$ we can estimate the rate of convergence in (4.7). To state the corresponding result, we first introduce some notation and assumptions.

Given numbers $p \geq 1$, $0 < \alpha < 1$, $r \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$, where \mathbb{N} is the set of natural numbers, we set $\beta = \alpha + r$ and denote by $H_p(\beta)$ the L^p -Hölder class, that is, the class of those functions $\psi(\lambda) \in L^p(\Lambda)$, which have r -th derivatives in $L^p(\Lambda)$ and with some positive constant C satisfy

$$\|\psi^{(r)}(\cdot + h) - \psi^{(r)}(\cdot)\|_p \leq C|h|^\alpha.$$

Assumption 4.2. We say that a pair of integrable functions $(f(\lambda), g(\lambda))$, $\lambda \in \Lambda$, satisfies condition (\mathcal{H}) , and write $(f, g) \in (\mathcal{H})$, if $f \in H_p(\beta_1)$ for $\beta_1 > 0$, $p > 1$ and $g \in H_q(\beta_2)$ for $\beta_2 > 0$, $q > 1$ with $1/p + 1/q = 1$, and one of the conditions a) – d) is satisfied:

- a) $\beta_1 > 1/p$, $\beta_2 > 1/q$,
- b) $\beta_1 \leq 1/p$, $\beta_2 \leq 1/q$ and $\beta_1 + \beta_2 > 1/2$,
- c) $\beta_1 > 1/p$, $1/q - 1/2 < \beta_2 \leq 1/q$,
- d) $\beta_2 > 1/q$, $1/p - 1/2 < \beta_1 \leq 1/p$.

Remark 4.2. In Ginovian [38] it was proved that if $(f, g) \in (\mathcal{H})$, then there exist numbers p_1 ($p_1 > p$) and q_1 ($q_1 > q$), such that $H_p(\beta_1) \subset L_{p_1}$, $H_q(\beta_2) \subset L_{q_1}$ and $1/p_1 + 1/q_1 \leq 1/2$.

Assumption 4.3. The spectral density f and the generating function g are such that $f, g \in L^1(\Lambda) \cap L^2(\Lambda)$ ($f, g \in L^2(\Lambda)$ in the d.t. case) and g is of bounded variation.

The following theorem controls the bias $E(J_T^h) - J$ and provides sufficient conditions assuring the proper rate of convergence of bias to zero, necessary for asymptotic normality of the estimator J_T^h . Specifically, we have the following result.

Theorem 4.2. *Let the functionals $J := J(f, g)$ and $J_T^h := J(I_T^h, g)$ be defined by (4.1) and (4.2), respectively. Then under Assumptions 3.1 and 4.2 (or 4.3), the following asymptotic relation holds:*

$$T^{1/2} [\mathbb{E}(J_T^h) - J] \rightarrow 0 \quad \text{as } T \rightarrow \infty. \quad (4.8)$$

Remark 4.3. We call an estimator J_T^h of J *asymptotically unbiased of the order of T^β* , $\beta > 0$ if $\lim_{T \rightarrow \infty} T^\beta [E(J_T^h) - J] = 0$. Thus, Theorem 4.2 states that the statistic J_T^h is an asymptotically unbiased estimator for J of the order of $T^{1/2}$.

4.3. Consistency

Recall that an estimator J_T^h of J is said to be (a) consistent if $J_T^h \rightarrow J$ in probability as $T \rightarrow \infty$, (b) mean square consistent if $\mathbb{E}(J_T^h - J)^2 \rightarrow 0$ as $T \rightarrow \infty$, (c) \sqrt{T} -consistent in the mean square sense if $\mathbb{E}([\sqrt{T}(J_T^h - J)]^2) = O(1)$ as $T \rightarrow \infty$,

To state the corresponding results we first introduce the following assumption.

Assumption 4.4. (A) (d.t. case). The spectral density $f(\cdot)$ and the generating function $g(\cdot)$ are such that $f(\cdot) \in L^p(\Lambda)$ ($p \geq 1$) and $g(\cdot) \in L^q(\Lambda)$ ($q \geq 1$) with $1/p + 1/q \leq 1/2$

(B) (c.t. case). The spectral density $f(\cdot)$ and the generating function $g(\cdot)$ are such that $f(\cdot) \in L^1(\Lambda) \cap L^p(\Lambda)$ ($p \geq 1$) and $g(\cdot) \in L^1(\Lambda) \cap L^q(\Lambda)$ ($q \geq 1$) with $1/p + 1/q \leq 1/2$.

(C) (c.t. Lévy-driven case). The filter $a(\cdot)$, that is, the $\text{MA}(\infty)$ coefficients (see (2.6)), and the generating kernel $\hat{g}(\cdot)$ are such that

$$a(\cdot) \in L^2(\Lambda) \cap L^p(\Lambda), \quad \hat{g}(\cdot) \in L^q(\Lambda) \quad \text{with } 1 \leq p, q \leq 2, \quad 2/p + 1/q \geq 5/2.$$

We begin with a result on the asymptotic behavior of the variance $\text{Var}(J_T^h) = \mathbb{E}(J_T^h - \mathbb{E}(J_T^h))^2$. The proof of the next theorem can be found in Ginovyan and Sahakyan [52].

Theorem 4.3. *Let the functionals $J := J(f, g)$ and $J_T^h := J(I_T^h, g)$ be defined by (4.1) and (4.2), respectively. Then under Assumptions 3.1 and 4.4 the following asymptotic relation holds:*

$$\lim_{T \rightarrow \infty} T \text{Var}(J_T^h) = \sigma_h^2(J), \quad (4.9)$$

where

$$\sigma_h^2(J) := 4\pi e(h) \int_{\Lambda} f^2(\lambda) g^2(\lambda) d\lambda + \kappa_4 e(h) \left[\int_{\Lambda} f(\lambda) g(\lambda) d\lambda \right]^2. \quad (4.10)$$

Here κ_4 is the fourth cumulant of $\xi(1)$, and

$$e(h) := \lim_{T \rightarrow \infty} \frac{TH_{4,T}}{H_{2,T}^2}. \quad (4.11)$$

Observe that for c.t. case we have

$$e(h) = \lim_{T \rightarrow \infty} \frac{TH_{4,T}}{H_{2,T}^2} = \frac{H_4}{H_2^2} = \int_0^1 h^4(t) dt \left(\int_0^1 h^2(t) dt \right)^{-2}.$$

Remark 4.4. In the case where the underlying process $X(t)$ is nonlinear for the second term on the right-side of formula (4.10) we have an expression in terms of the tri-spectrum (the fourth order spectrum) of the process $X(t)$ (see, e.g., Dahlhaus [18], and Guyon [60], p.142).

From Theorems 4.1–4.3 we infer the following result.

Theorem 4.4. *The following assertions hold.*

- (a) *Under Assumptions 3.1, 4.1 and 4.4 the statistic J_T^h is a mean square consistent estimator for J .*
- (b) *Under Assumptions 3.1, 4.2 (or 4.3) and 4.4 the statistic J_T^h is a \sqrt{T} -consistent in the mean square sense estimator for J .*

4.4. Asymptotic normality

The next result contains sufficient conditions for functional J_T^h to obey the central limit theorem (CLT), and was proved in Ginovyan and Sahakyan [52].

Theorem 4.5 (CLT). *Let $J := J(f, g)$ and $J_T^h := J(I_T^h, g)$ be defined by (4.1) and (4.2), respectively. Then under Assumptions 3.1 and 4.4 the functional J_T^h obeys the central limit theorem. More precisely, we have*

$$T^{1/2} [J_T^h - \mathbb{E}(J_T^h)] \xrightarrow{d} \eta \quad \text{as } T \rightarrow \infty, \quad (4.12)$$

where the symbol \xrightarrow{d} stands for convergence in distribution, and η is a normally distributed random variable with mean zero and variance $\sigma_h^2(J)$ given by (4.10) and (4.11).

Taking into account the equality

$$T^{1/2} [J_T^h - J] = T^{1/2} [\mathbb{E}(J_T^h) - J] + T^{1/2} [J_T^h - \mathbb{E}(J_T^h)], \quad (4.13)$$

as an immediate consequence of Theorems 4.2 and 4.5, we obtain the next result that contains sufficient conditions for a simple ‘plug-in’ statistic $J(I_T^h)$ to be an asymptotically normal estimator for a linear spectral functional J .

Theorem 4.6. *Let the functionals $J := J(f, g)$ and $J_T^h := J(I_T^h, g)$ be defined by (4.1) and (4.2), respectively. Then under Assumptions 3.1, 4.2 (or 4.3) and*

4.4 the statistic J_T^h is an asymptotically normal estimator for functional J . More precisely, we have

$$T^{1/2} [J_T^h - J] \xrightarrow{d} \eta \quad \text{as } T \rightarrow \infty, \tag{4.14}$$

where η is as in Theorem 4.5, that is, η is a normally distributed random variable with mean zero and variance $\sigma_h^2(J)$ given by (4.10) and (4.11).

Remark 4.5. Notice that if the underlying process $X(u)$ is Gaussian (in this case, $\kappa_4 = 0$), then in formula (4.10) we have only the first term. Using the results from Ginovyan [38] and Ginovyan and Sahakyan [47, 48], it can be shown that in this case Theorem 4.6 is true under Assumptions 3.1 and 4.4.

Remark 4.6. In von Sachs [82] (see also Janas and von Sachs [69]) was considered nonparametric estimation problem for some non-linear integral functionals of spectral density, using data-tapers. Consistency and asymptotic normality for the corresponding non-linear functionals of the tapered periodogram was proved. The obtained results were applied to the use of data-tapers in nonparametric peak-insensitive spectrum estimation.

Example 4.1 (Estimation of covariance function). Assume that $X(t)$ is a d.t. process with spectral density f satisfying $f \in L^2(\Lambda)$, and let $g(\lambda) = e^{iu\lambda}$, then

$$J(f) = \int_{-\pi}^{\pi} e^{iu\lambda} f(\lambda) d\lambda := r(u).$$

Thus, in this special case our problem becomes to the estimation of the covariance function $r(u) = \mathbb{E}[X(t+u)X(t)]$ of the process $X(t)$. It is clear that Assumptions 4.3 and 4.4(A) are satisfied, and we can apply Theorem 4.6 to conclude that the simple “plug-in” statistic

$$J_T^h = \widehat{r}_T(u) = \int_{-\pi}^{\pi} e^{iu\lambda} I_T^h(\lambda) d\lambda,$$

where h is a taper function satisfying Assumptions 3.1, is asymptotically normal estimator for $r(u)$ with asymptotic variance

$$\sigma_u^2 = 4\pi e(h) \int_{-\pi}^{\pi} \cos^2(u\lambda) f^2(\lambda) d\lambda,$$

where $e(h)$ is given by (4.11).

Example 4.2 (Estimation of spectral function). Assume that $X(t)$ is as in Example 4.1, that is, it is a d.t. process with spectral density f satisfying $f \in L^2(\Lambda)$, and let $g(\lambda) = \chi_{[0,\mu]}(\lambda)$ be the indicator of an interval $[0, \mu]$, then

$$J(f) = \int_{-\pi}^{\pi} \chi_{[0,\mu]}(\lambda) f(\lambda) d\lambda = \int_0^{\mu} f(\lambda) d\lambda := F(\mu).$$

Thus, in this case the estimand functional is the spectral function $F(\mu)$ of the process $X(u)$. It is clear that the assumptions of Theorem 4.6 are satisfied, and hence we can conclude that the simple “plug-in” statistic

$$J_T^h = \widehat{F}_T(\mu) = \int_0^\mu I_T^h(\lambda) d\lambda \quad (4.15)$$

is asymptotically normal estimator for $F(\mu)$ with asymptotic variance

$$\sigma^2(\mu) = 4\pi e(h) \int_0^\mu f^2(s) ds,$$

where $e(h)$ is given by (4.11).

5. Parametric Estimation Problem

We assume here that the spectral density $f(\lambda)$ belongs to a given parametric family of spectral densities $\mathcal{F} := \{f(\lambda, \theta) : \theta \in \Theta\}$, where $\theta := (\theta_1, \dots, \theta_p)$ is an unknown parameter and Θ is a subset in the Euclidean space \mathbb{R}^p . The problem of interest is to estimate the unknown parameter θ on the basis of the tapered data (3.5), and investigate the asymptotic (as $T \rightarrow \infty$) properties of the suggested estimators, depending on the dependence (memory) structure of the model $X(t)$ and the smoothness of its spectral density f .

There are different methods of estimation: maximum likelihood, Whittle, minimum contrast, etc. Here we focus on the Whittle method.

5.1. The Whittle estimation procedure

The Whittle estimation procedure, originally devised for d.t. short memory stationary processes, is based on the smoothed periodogram analysis on a frequency domain, involving approximation of the likelihood function and asymptotic properties of empirical spectral functionals (see Whittle [84]). The Whittle estimation method since its discovery has played a major role in the asymptotic theory of parametric estimation in the frequency domain, and was the focus of interest of many statisticians. Their aim was to weaken the conditions needed to guarantee the validity of the Whittle approximation for d.t. short memory models, to find analogues for long and intermediate memory models, to find conditions under which the Whittle estimator is asymptotically equivalent to the exact maximum likelihood estimator, and to extend the procedure to the c.t. models and random fields.

For the d.t. case, it was shown that for Gaussian and linear stationary models the Whittle approach leads to consistent and asymptotically normal estimators under short, intermediate and long memory assumptions. Moreover, it was shown that in the Gaussian case the Whittle estimator is also asymptotically efficient in the sense of Fisher (see, e. g., Dahlhaus [22], Dzhaparidze [29], Fox

and Taqqu [33], Giraitis and Surgailis [56], Guyon [60], Heyde and Gay [64], Taniguchi and Kakizawa [79], Walker [83], and references therein).

For c.t. models, the Whittle estimation procedure has been considered, for example, in Anh et al. [4], Avram et al. [6], Casas and Gao [12], Dzhaparidze [29], Dzhaparidze and Yaglom [32], Gao [34], Gao et al. [35], Leonenko and Sakhno [70], Tsai and Chan [81], where can also be found additional references. In this case, it was proved that the Whittle estimator is consistent and asymptotically normal.

The Whittle estimation procedure based on the d.t. tapered data has been studied in Alomari et al. [1], Dahlhaus [18, 19, 21], Dahlhaus and Künsch [26], Guyon [60], Ludeña and Lavielle [71]. In the case where the underlying model is a Lévy-driven c.t. linear process with possibly unbounded or vanishing spectral density function, consistency and asymptotic normality of the Whittle estimator was established in Ginovyan [45].

To explain the idea behind the Whittle estimation procedure, assume for simplicity that the underlying process $X(t)$ is a d.t. Gaussian process, and we want to estimate the parameter θ based on the sample $\mathbf{X}_T := \{X(t), t = 1, \dots, T\}$. A natural approach is to find the maximum likelihood estimator (MLE) $\hat{\theta}_{T,MLE}$ of θ , that is, to maximize the likelihood function, or to minimize the $-1/T \times \log$ -likelihood function $L_T(\theta)$, which in this case takes the form:

$$L_T(\theta) := \frac{1}{2} \ln 2\pi + \frac{1}{2T} \ln \det B_T(f_\theta) + \frac{1}{2T} \mathbf{X}'_T [B_T(f_\theta)]^{-1} \mathbf{X}_T,$$

where $B_T(f_\theta)$ is the Toeplitz matrix generated by f_θ (i.e. the covariance matrix of $X(t)$). Unfortunately, the above function is difficult to handle, and no explicit expression for the estimator $\hat{\theta}_{T,MLE}$ is known (even in the case of simple models). An approach, suggested by P. Whittle, called the Whittle estimation procedure, is to approximate the term $\ln \det B_T(f_\theta)$ by $\frac{T}{2} \int_{-\pi}^{\pi} \ln f_\theta(\lambda) d\lambda$ and the inverse matrix $[B_T(f_\theta)]^{-1}$ by the Toeplitz matrix $B_T(1/f_\theta)$. This leads to the following approximation of the log-likelihood function $L_T(\theta)$, introduced by Whittle [84], and called Whittle functional:

$$L_{T,W}(\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[\ln f_\theta(\lambda) + \frac{I_T(\lambda)}{f_\theta(\lambda)} \right] d\lambda, \quad (5.1)$$

where $I_T(\lambda)$ is the ordinary periodogram of the process $X(t)$ defined by (3.1).

Now minimizing the Whittle functional $L_{T,W}(\theta)$ with respect to θ , we get the Whittle estimator $\hat{\theta}_T$ for θ . It can be shown that if

$$T^{1/2}(L_T(\theta) - L_{T,W}(\theta)) \rightarrow 0 \quad \text{as } n \rightarrow \infty \quad \text{in probability,}$$

then the MLE $\hat{\theta}_{T,MLE}$ and the Whittle estimator $\hat{\theta}_T$ are asymptotically equivalent in the sense that $\hat{\theta}_T$ also is consistent, asymptotically normal and asymptotically Fisher-efficient (see, e.g., Coursol and Dacunha-Castelle [15], Dzhaparidze [29], Dzhaparidze [30] (p.114), and Dzhaparidze and Yaglom [32]).

In the continuous context, the Whittle procedure of estimation of a spectral parameter θ based on the sample $\mathbf{X}_T := \{X(t), 0 \leq t \leq T\}$ is to choose the estimator $\hat{\theta}_T$ to minimize the weighted Whittle functional:

$$U_T(\theta) := \frac{1}{4\pi} \int_{\mathbb{R}} \left[\ln f(\lambda, \theta) + \frac{I_T(\lambda)}{f(\lambda, \theta)} \right] \cdot w(\lambda) d\lambda, \quad (5.2)$$

where $I_T(\lambda)$ is the continuous periodogram of $X(t)$, and $w(\lambda)$ is a weight function ($w(-\lambda) = w(\lambda)$, $w(\lambda) \geq 0$, $w(\lambda) \in L^1(\mathbb{R})$) for which the integral in (5.2) is well defined. An example of common used weight function is $w(\lambda) = 1/(1 + \lambda^2)$.

The Whittle procedure of estimation of a spectral parameter θ based on the tapered sample (3.5) is to choose the estimator $\hat{\theta}_{T,h}$ to minimize the weighted tapered Whittle functional:

$$U_{T,h}(\theta) := \frac{1}{4\pi} \int_{\Lambda} \left[\log f(\lambda, \theta) + \frac{I_T^h(\lambda)}{f(\lambda, \theta)} \right] \cdot w(\lambda) d\lambda, \quad (5.3)$$

where $I_T^h(\lambda)$ is the tapered periodogram of $X(t)$, given by (3.10), and $w(\lambda)$ is a weight function for which the integral in (5.3) is well defined (in the d.t. case as a weight function we take $w(\lambda) \equiv 1$). Thus, the Whittle estimator $\hat{\theta}_{T,h}$ of θ based on the tapered sample (3.5) is defined by

$$\hat{\theta}_{T,h} := \underset{\theta \in \Theta}{\text{Arg min}} U_{T,h}(\theta), \quad (5.4)$$

where $U_{T,h}(\theta)$ is given by (5.3).

Remark 5.1. In connection with the Whittle approximation to the log-likelihood function, Dahlhaus [24] has discussed several approximations for the inverse of the Toeplitz matrix $[B_T(f_\theta)]^{-1}$ (i.e. the covariance matrix). In particular, he proved that for his increasing class of stochastic processes \mathfrak{X} and the class of taper functions $D(k, \kappa)$ (see Section 3.3), a taper function $h \in D(k, \kappa)$ can be chosen so that the tapered Toeplitz matrix $B_T^h(1/f_\theta)$ is a better approximation for the inverse $[B_T(f_\theta)]^{-1}$ than the nontapered Toeplitz matrix $B_T(1/f_\theta)$.

Remark 5.2. In the c.t. case approximation of the likelihood function, similar to the Whittle approximation (5.1), becomes a complicated problem. The issue is that unlike the d.t. case, in the c.t. case there is no a measure analogous to the Lebesgue measure with respect to which the probability distributions of the observed process are absolutely continuous. We thus have to compute the density of the probability distribution of the sample with respect to some fixed dominating measure. Specifically, let $\{X(t), t \in \mathbb{R}\}$ be a c.t. Gaussian process, $\mathbf{X}_T := \{X(t), 0 \leq t \leq T\}$, and let $P_T = P_T(\theta)$ be the related Gaussian measure defined on the space of functions of t ($0 \leq t \leq T$). Then the likelihood $p_T(\mathbf{X}_T)$ is defined here as the Radon-Nikodym derivative $dP_T/dP_T^{(0)}$ of the measure P_T with respect to some *standard* dominating measure $P_T^{(0)}$ on the same space. Hence both the likelihood $p_T(\mathbf{X}_T)$ and the log-likelihood $L_T(\mathbf{X}_T) := \ln p_T(\mathbf{X}_T)$ are functionals depending on \mathbf{X}_T .

To the best of our knowledge, the equation (similar to the Whittle equation (5.1)) for the *main part* $L_{T,W}$ of the log-likelihood L_T , has been derived rigorously so far only for a very special class of c.t. Gaussian stationary processes with everywhere positive rational spectral densities of the form (2.10) (i.e. for the class of Gaussian CARMA processes). For details, see Pisarenko [75] (see also Dzhaparidze [29], and Dzhaparidze and Yaglom [32]).

5.2. Asymptotic properties of the Whittle estimator

To state results involving properties of the Whittle estimator, we first introduce the following set of assumptions.

Assumption 5.1. The true value θ_0 of the parameter θ belongs to a compact set Θ in the p -dimensional Euclidean space \mathbb{R}^p , and $f(\lambda, \theta_1) \neq f(\lambda, \theta_2)$ whenever $\theta_1 \neq \theta_2$ almost everywhere in Λ with respect to the Lebesgue measure.

Assumption 5.2. The functions $f(\lambda, \theta)$, $f^{-1}(\lambda, \theta)$ and $\frac{\partial}{\partial \theta_k} f^{-1}(\lambda, \theta)$, $k = 1, \dots, p$, are continuous in (λ, θ) .

Assumption 5.3. The functions $f := f(\lambda, \theta)$ and $g := w(\lambda) \frac{\partial}{\partial \theta_k} f^{-1}(\lambda, \theta)$ satisfy Assumptions 4.2 or 4.3 for all $k = 1, \dots, p$ and $\theta \in \Theta$.

Assumption 5.4. The functions f , g , $a := a(\lambda, \theta)$ (the MA(∞) coefficients) and $b := \hat{g}$, where g is as in Assumption 5.3, satisfy Assumption 4.4.

Assumption 5.5. The functions $\frac{\partial^2}{\partial \theta_k \partial \theta_j} f^{-1}(\lambda, \theta)$ and $\frac{\partial^3}{\partial \theta_k \partial \theta_j \partial \theta_l} f^{-1}(\lambda, \theta)$, $k, j, l = 1, \dots, p$, are continuous in (λ, θ) for $\lambda \in \Lambda$, $\theta \in N_\delta(\theta_0)$, where $N_\delta(\theta_0) := \{\theta : |\theta - \theta_0| < \delta\}$ is some neighborhood of θ_0 .

Assumption 5.6. The matrices

$$W(\theta) := \|w_{ij}(\theta)\|, \quad A(\theta) := \|a_{ij}(\theta)\|, \quad B(\theta) := \|b_{ij}(\theta)\|, \quad i, j = 1, \dots, p \quad (5.5)$$

are positive definite, where

$$w_{ij}(\theta) = \frac{1}{4\pi} \int_{\Lambda} \frac{\partial}{\partial \theta_i} \ln f(\lambda, \theta) \frac{\partial}{\partial \theta_j} \ln f(\lambda, \theta) w(\lambda) d\lambda, \quad (5.6)$$

$$a_{ij}(\theta) = \frac{1}{4\pi} \int_{\Lambda} \frac{\partial}{\partial \theta_i} \ln f(\lambda, \theta) \frac{\partial}{\partial \theta_j} \ln f(\lambda, \theta) w^2(\lambda) d\lambda, \quad (5.7)$$

$$b_{ij}(\theta) = \frac{\kappa_4}{16\pi^2} \int_{\Lambda} \frac{\partial}{\partial \theta_i} \ln f(\lambda, \theta) w(\lambda) d\lambda \int_{\Lambda} \frac{\partial}{\partial \theta_j} \ln f(\lambda, \theta) w(\lambda) d\lambda, \quad (5.8)$$

and κ_4 is the fourth cumulant of $\xi(1)$.

The next theorem contains sufficient conditions for the tapered Whittle estimator to be consistent (see Ginovyan [45]).

Theorem 5.1. Let $\hat{\theta}_{T,h}$ be the tapered Whittle estimator defined by (5.4) and let θ_0 be the true value of parameter θ . Then, under Assumptions 5.1–5.4 and 3.1, the statistic $\hat{\theta}_{T,h}$ is a consistent estimator for θ , that is, $\hat{\theta}_{T,h} \rightarrow \theta_0$ in probability as $T \rightarrow \infty$.

Having established the consistency of the Whittle estimator $\widehat{\theta}_{T,h}$, we can go on to obtain the limiting distribution of $T^{1/2}(\widehat{\theta}_{T,h} - \theta_0)$ in the usual way by applying the Taylor's formula, the mean value theorem, and Slutsky's arguments. Specifically we have the following result, showing that under the above assumptions, the tapered Whittle estimator $\widehat{\theta}_{T,h}$ is asymptotically normal (see Ginovyan [45]).

Theorem 5.2. *Suppose that Assumptions 5.1–5.6 and 3.1 are satisfied. Then the Whittle estimator $\widehat{\theta}_{T,h}$ of an unknown spectral parameter θ based on the tapered data (3.5) is asymptotically normal. More precisely, we have*

$$T^{1/2}(\widehat{\theta}_{T,h} - \theta_0) \xrightarrow{d} N_p(0, e(h)\Gamma(\theta_0)) \quad \text{as } T \rightarrow \infty, \quad (5.9)$$

where $N_p(\cdot, \cdot)$ denotes the p -dimensional normal law, \xrightarrow{d} stands for convergence in distribution,

$$\Gamma(\theta_0) = W^{-1}(\theta_0)(A(\theta_0) + B(\theta_0))W^{-1}(\theta_0), \quad (5.10)$$

where the matrices W , A and B are defined in (5.5)–(5.8), and the tapering factor $e(h)$ is given by formula (4.11).

Remark 5.3 (The variance effect). Since tapering of the data, roughly speaking, reduces the effective length of the data, it is not surprising that in the classical asymptotic setting, the corresponding tapered estimators, generally, will have larger variances than their non-tapered counterparts. Specifically, using the Cauchy-Schwartz inequality for the tapering factor $e(h)$ (defined by formula (4.11)) we have $e(h) \geq 1$, and the equality is attained in the non-tapered case, that is, for $h(t) = \mathbb{I}_{[0,1]}(t)$. Thus, the use of tapers, generally, will result in an efficiency loss. However, as it was observed by Dahlhaus [23] (p.156, 161), ‘it is not correct to conclude from this that tapering always increases the variance of the estimators’, because a taper function h can be chosen to satisfy $e(h) = 1$.

In the classical asymptotic setting, for d.t. processes it is possible to choose the taper function $h_T(t) := h_T(t/T)$ so that the corresponding tapered estimator will be asymptotically efficient. To clarify this in the next corollary we consider an important subclass of tapers (cf. Dahlhaus [19], Corollary 2.3).

Corollary 5.1. *Suppose in addition to the assumptions of Theorem 5.2 that the process $X(t)$ ($t \in \mathbb{Z}$) is Gaussian and that $h_T(t) := h_T(t/T)$ with continuous functions $h_T : [0, 1] \rightarrow [0, \infty)$ of uniformly bounded variation $BV(h_T)$ that converge almost everywhere to a function $h(x)$. Then the tapered Whittle estimator $\widehat{\theta}_{T,h}$ is asymptotically efficient if and only if $h(x) = 1$ almost everywhere.*

Remark 5.4. In Dzaparidze and Yaglom [32] (p.68) it was shown that in the non-tapered d.t. case the covariance structure (5.10) in Theorem 5.2 is the smallest obtainable for a large class of estimators even in the non-Gaussian case. Using this result, Corollary 5.1 can also be extended to the case of non-Gaussian processes (see also Dahlhaus [19]).

Remark 5.5. In the d.t. case as a weight function we take $w(\lambda) \equiv 1$, and the matrices $A(\theta_0)$ and $W(\theta_0)$ coincide (see (5.5) – (5.7)). So, in this case, formula (5.10) becomes $\Gamma(\theta_0) = W^{-1}(\theta_0)(W(\theta_0) + B(\theta_0))W^{-1}(\theta_0)$. If, in addition, $B(\theta_0) = 0$ and the taper is as in Corollary 5.1, then we have $e(h) = 1$ and $\Gamma(\theta_0) = W^{-1}(\theta_0)$, that is, the tapered Whittle estimator $\hat{\theta}_{T,h}$ is asymptotically efficient (cf. Theorem 6.3 of Dahlhaus [20]). For example, for the polynomial taper in Example 3.1, we have $h_{\rho_T}(x) \rightarrow \mathbb{I}_{[0,1]}(x)$ and $e(h) = \lim_{T \rightarrow \infty} (TH_{4,T})/H_{2,T}^2 = 1$ (see Dahlhaus [23]), and thus the corresponding tapered Whittle estimator $\hat{\theta}_{T,h}$ is asymptotically efficient.

Remark 5.6. In view of the definition of the matrix $B(\theta)$ (see (5.5), (5.8)) for d.t. processes ($w(\lambda) \equiv 1$), the equality $B(\theta) = 0$ holds in the following two cases (cf. Dahlhaus and Janas [25]).

Case 1. $\kappa_4 = 0$. This condition is fulfilled, for example, if the innovations are assumed to be Gaussian.

Case 2. $\int_{\Lambda} \frac{\partial}{\partial \theta} \ln f(\lambda, \theta) d\lambda = 0$. This condition holds for several parametrizations, which can be deduced from Kolmogorov's formula (cf. Brockwell and Davis [11], Section 5.8).

6. Goodness-of-fit tests

In this section we consider the following problem of hypotheses testing.

Based on the tapered sample \mathbf{X}_T^h given by (3.5), we want to construct goodness-of-fit tests for testing a hypothesis H_0 that the spectral density of the process $X(t)$ has the specified form $f(\lambda)$. We will distinguish the following two cases.

- a) The hypothesis H_0 is simple, that is, the hypothetical spectral density $f(\lambda)$ of $X(t)$ does not depend on unknown parameters.
- b) The hypothesis H_0 is composite, that is, the hypothetical spectral density $f(\lambda)$ of $X(t)$ depends on an unknown p -dimensional vector parameter $\theta = (\theta_1, \dots, \theta_p)$, that is, $f(\lambda) = f(\lambda, \theta)$, $\lambda \in \Lambda$, $\theta \in \Theta \subset \mathbb{R}^p$.

The above stated problem in the non-tapered case has been considered by many authors for different models. For instance, for independent observations, the problem was considered in Chernov and Lehman [13], Chibisov [14], Cramer [16], and Dzhaparidze and Nikulin [31]. For observations generated by d.t. Gaussian stationary processes it was considered in Dzhaparidze [30], Ginovyan [41], Hannan [61], and Osidze [73, 74]. For c.t. Gaussian stationary observations, the problem was discussed in Ginovyan [44] and Osidze [73, 74]. For tapered case the problem has been considered in Ginovyan [46].

To test the hypothesis H_0 , similar to the non-tapered case, it is natural to introduce a measure of divergence (disparity) of the hypothetical and empirical spectral densities, and construct a goodness-of-fit test based on the distribution of the chosen measure (see, e.g., Dzhaparidze [30], Ginovyan [41, 44, 46], Hannan [61], and Osidze [73, 74]).

6.1. A Goodness-of-fit test for simple hypothesis

We first consider the relatively easy case a) of a simple hypothesis H_0 . As a measure of divergence of the hypothetical spectral density $f(\lambda)$ and the tapered empirical spectral density $I_T^h(\lambda)$, we consider the m -dimensional random vector

$$\Phi_T^h := (\Phi_{1T}^h, \dots, \Phi_{mT}^h) \quad (6.1)$$

with elements

$$\Phi_{jT}^h := \Phi_{jT}(\mathbf{X}_T^h) = \frac{\sqrt{T}}{\sqrt{4\pi e(h)}} \int_{\Lambda} \left[\frac{I_T^h(\lambda)}{f(\lambda)} - 1 \right] \varphi_j(\lambda) d\lambda, \quad j = 1, 2, \dots, m, \quad (6.2)$$

where $e(h)$ is as in (4.11) and $\{\varphi_j(\lambda), j = 1, 2, \dots, m\}$ is some orthonormal system on Λ :

$$\int_{\Lambda} \varphi_k(\lambda) \varphi_j(\lambda) d\lambda = \delta_{kj} = \begin{cases} 1 & \text{for } k = j, \\ 0 & \text{for } k \neq j. \end{cases} \quad (6.3)$$

In Ginovyan [46] it was shown that under wide conditions on $f(\lambda)$ and $\varphi_j(\lambda)$, the random vector Φ_T^h in (6.1) – (6.2) has asymptotically $N(0, I_m)$ -normal distribution as $T \rightarrow \infty$, where I_m is the $m \times m$ identity matrix. Therefore in the case of simple hypothesis H_0 , we can use the statistic

$$S_T^h = S_T(\mathbf{X}_T^h) := \Phi_T^h(\mathbf{X}_T^h) \Phi_T^h(\mathbf{X}_T^h) = \sum_{j=1}^m \Phi_{jT}^2(\mathbf{X}_T^h), \quad (6.4)$$

which for $T \rightarrow \infty$ will have a χ^2 -distribution with m degrees of freedom.

Thus, fixing an asymptotic level of significance α we can consider the class of goodness-of-fit tests for testing the simple hypothesis H_0 about the form of the spectral density f with asymptotic level of significance α determined by critical regions of the form:

$$\{\mathbf{X}_T^h : S_T(\mathbf{X}_T^h) > d_\alpha\},$$

where $S_T(\mathbf{X}_T^h)$ is given by (6.4), and d_α is the α -quantile of χ^2 -distribution with m degrees of freedom, that is, d_α is determined from the condition:

$$P(\chi^2 > d_\alpha) = \int_{d_\alpha}^{\infty} k_m(x) dx = \alpha,$$

where $k_m(x)$ is the density of χ^2 -distribution with m degrees of freedom.

The next theorem contains sufficient conditions for statistic S_T^h , given by (6.4), to have a limiting (as $T \rightarrow \infty$) χ^2 -distribution with m degrees of freedom (see Ginovyan [46]).

Theorem 6.1. *Let the spectral density $f(\lambda)$ and the orthonormal functions $\{\varphi_j(\lambda), j = 1, 2, \dots, m\}$ be such that $(f, g_j) \in (\mathcal{H})$ for all $j = 1, 2, \dots, m$, where $g_j = \varphi_j/f$ (see Assumption 4.2). Then under Assumption 3.1 the limiting (as $T \rightarrow \infty$) distribution of the statistic $S_T^h = S_T(\mathbf{X}_T^h)$ given by (6.4) is a χ^2 -distribution with m degrees of freedom.*

Remark 6.1. For the non-tapered case, for observations generated by d.t. short-memory Gaussian stationary processes the result of Theorem 6.1 was first proved in Hannan [61] (p. 94) (see, also, Dzhaparidze [30] and Osidze [73, 74]). In the case where the spectral density has singularities (zeros and/or poles), the result for d.t. processes was proved in Ginovyan [41]. The non-tapered counterpart of Theorem 6.1 for c.t. processes was proved in Ginovyan [44].

6.2. A Goodness-of-fit test for composite hypothesis

Now we consider the case of composite hypothesis H_0 , and assume that the hypothetical spectral density $f = f(\lambda, \theta)$ is known with the exception of a vector parameter $\theta := (\theta_1, \dots, \theta_p) \in \Theta \subset \mathbb{R}^p$. In this case, the problem of construction of goodness-of-fit tests becomes more complex, because we first have to choose an appropriate statistical estimator $\hat{\theta}_T$ for the unknown parameter θ , constructed on the basis of the tapered sample (3.5). It is important to remark that in this case the limiting distribution of the test statistic will change in accordance with properties of an estimator of θ , and generally will not be a χ^2 -distribution.

For testing a composite hypothesis H_0 , we again can use a statistic of type (6.4), but with a statistical estimator $\hat{\theta}_T$ instead of unknown θ . The corresponding statistic can be written as follows:

$$S_T^h(\hat{\theta}_T) = S_T(\mathbf{X}_T^h, \hat{\theta}_T) := \Phi_T'(\mathbf{X}_T^h, \hat{\theta}_T)\Phi_T(\mathbf{X}_T^h, \hat{\theta}_T) = \sum_{j=1}^m \Phi_{jT}^2(\mathbf{X}_T^h, \hat{\theta}_T), \quad (6.5)$$

where now

$$\Phi_T^h(\mathbf{X}_T^h, \hat{\theta}_T) := (\Phi_{1T}(\mathbf{X}_T^h, \hat{\theta}_T), \dots, \Phi_{mT}(\mathbf{X}_T^h, \hat{\theta}_T))$$

with elements

$$\Phi_{jT}(\mathbf{X}_T^h, \hat{\theta}_T) := \frac{\sqrt{T}}{\sqrt{4\pi e(h)}} \int_{\Lambda} \left[\frac{I_T^h(\lambda)}{f(\lambda, \hat{\theta}_T)} - 1 \right] \varphi_j(\lambda) d\lambda, \quad j = 1, 2, \dots, m. \quad (6.6)$$

So, we must choose an appropriate statistical estimator $\hat{\theta}_T$ for unknown θ , and determine the limiting distribution of the statistic (6.5). Then, having the limiting distribution of the statistic (6.5), for given level of significance α we can consider the class of goodness-of-fit tests for testing the composite hypothesis H_0 about the form of the spectral density f with asymptotic level of significance α determined by critical regions of the form:

$$\{\mathbf{X}_T^h : S_T(\mathbf{X}_T^h, \hat{\theta}_T) > d_\alpha\},$$

where d_α is the α -quantile of the limiting distribution of the statistic (6.5), that is, d_α is determined from the condition:

$$\int_{d_\alpha}^{\infty} \hat{k}_m(x) dx = \alpha,$$

where $\widehat{k}_m(x)$ is the density of the limiting distribution of $S_T^h(\widehat{\theta}_T)$ defined by (6.5).

To state the corresponding result we first introduce the following set of assumptions:

Assumption 6.1. For $\theta \in \Theta$, $(f, g_j) \in (\mathcal{H})$ for all $j = 1, 2, \dots, m$, where $f := f(\lambda, \theta)$ and $g_j := \varphi_j(\lambda)/f(\lambda, \theta)$.

Assumption 6.2. For $\theta \in \Theta$, $(f, h_{kj}) \in (\mathcal{H})$ for all $k = 1, 2, \dots, p$ and $j = 1, 2, \dots, m$, where $f := f(\lambda, \theta)$ and $h_{kj} := \frac{\varphi_j(\lambda)}{f(\lambda, \theta)} \frac{\partial}{\partial \theta_k} \ln f(\lambda, \theta)$.

Assumption 6.3. The $(p \times p)$ -matrix $\Gamma(\theta_0) = \|\gamma_{kj}(\theta_0)\|_{k,j=1,p}$ with elements

$$\gamma_{kj}(\theta_0) := \frac{1}{4\pi} \int_{\Lambda} \left[\frac{\partial}{\partial \theta_k} \ln f(\lambda, \theta) \right]_{\theta=\theta_0} \left[\frac{\partial}{\partial \theta_j} \ln f(\lambda, \theta) \right]_{\theta=\theta_0} d\lambda \quad (6.7)$$

is non-singular.

Assumption 6.4. There exists a \sqrt{T} -consistent estimator $\widehat{\theta}_T$ for the parameter θ such that the following asymptotic relation holds:

$$\sqrt{T}(\widehat{\theta}_T - \theta_0) - \Gamma^{-1}(\theta_0)\Delta_T(\theta_0) = o_P(1), \quad (6.8)$$

where $\Gamma^{-1}(\theta_0)$ is the inverse of the matrix $\Gamma(\theta_0)$ defined in Assumption 6.3, and

$$\Delta_T(\theta) := \Delta_T^h(\theta) = (\Delta_{1T}(\theta), \dots, \Delta_{pT}(\theta)) \quad (6.9)$$

is a p -dimensional random vector with components

$$\Delta_{kT}(\theta) := \frac{\sqrt{T}}{\sqrt{4\pi e(h)}} \int_{\Lambda} \left[\frac{I_T^h(\lambda)}{f(\lambda, \theta)} - 1 \right] \frac{\partial}{\partial \theta_k} \ln f(\lambda, \theta) d\lambda, \quad k = 1, \dots, p. \quad (6.10)$$

The term $o_P(1)$ in (6.8) tends to zero in probability as $T \rightarrow \infty$. (Recall that an estimator $\widehat{\theta}_T$ for θ is called \sqrt{T} -consistent if $\sqrt{T}(\widehat{\theta}_T - \theta)$ is bounded in probability).

Remark 6.2. As an estimator $\widehat{\theta}_T$ for θ satisfying (6.8) can be considered minimum contrast estimators (in particular, the Whittle estimator) based on the tapered data. Minimum contrast estimators based on the tapered data for d.t. processes have been studied in Dahlhaus [18, 20, 19], for Gaussian c.t. processes in Ginovyan [45], and for some classes of c.t. non-Gaussian processes in Alomari et al [1].

Let $B(\theta) := B^h(\theta) = \|b_{jk}(\theta)\|$, $j = 1, \dots, m$, $k = 1, \dots, p$, be a $(m \times p)$ -matrix with elements

$$b_{jk}(\theta) := \frac{1}{\sqrt{4\pi e(h)}} \int_{\Lambda} \varphi_j(\lambda) \frac{\partial}{\partial \theta_k} \ln f(\lambda, \theta) d\lambda, \quad (6.11)$$

where $\varphi_j(\lambda)$ ($j = 1, 2, \dots, m$) are the functions from (6.2), and $e(h)$ is as in (4.11).

The following theorem was proved in Ginovyan [46].

Theorem 6.2. *Under Assumptions 3.1, 5.1 and 6.1–6.4 the limiting distribution (as $T \rightarrow \infty$) of the statistic $S_T(\mathbf{X}_T^h, \hat{\theta}_T)$ given by (6.5), coincides with the distribution of the random variable*

$$\sum_{j=1}^{m-p} \xi_j^2 + \sum_{j=1}^p \nu_j \xi_{m-p+j}^2,$$

where ξ_j , $j = 1, 2, \dots, m$, are iid $N(0, 1)$ random variables, while the numbers ν_k ($0 \leq \nu_k < 1$), $k = 1, 2, \dots, p$, are the roots relative to ν of the following equation:

$$\det [(1 - \nu)\Gamma(\theta_0) - B'(\theta_0)B(\theta_0)] = 0. \tag{6.12}$$

Remark 6.3. For the non-tapered case, for independent observations the result of Theorem 6.2 was first obtained by Chernov and Lehman [13] (see, also, Chibisov [14]). For observations generated by d.t. short-memory Gaussian stationary processes the result was stated in Osidze [73] (see, also, Dzhaparidze [30]). In the case where the spectral density has singularities, the result for d.t. processes was proved in Ginovyan [41]. The non-tapered counterpart of Theorem 6.2 for c.t. processes was proved in Ginovyan [44].

Example 6.1. Let $X(t)$ be a d.t. Autoregressive Process of order p ($\text{AR}(p)$), that is, $X(t)$ is a stationary process with the spectral density $f(\lambda) = \frac{1}{2\pi} \cdot |\alpha_p(e^{-i\lambda})|^{-2}$, where $\alpha_p(z) = 1 - \theta_1 z - \dots - \theta_p z^p$. Consider the functions $\varphi_j(\lambda)$ ($j = 1, \dots, m$) defined by $\varphi_j(\lambda) := c e^{ij\lambda} \alpha_p(e^{-i\lambda}) |\alpha_p(e^{-i\lambda})|^{-1}$ for $j > p$ and $\varphi_j(\lambda) := 0$ for $j \leq p$, where c is a normalizing constant. As an estimator of $\theta := (\theta_1, \dots, \theta_p)$ consider the Whittle estimator, and as a taper the Tukey-Hanning taper function $h(t)$ (see Remark 3.2). Then it is easy to check that the conditions of Theorem 6.2 are satisfied and $B(\theta) = 0$. Therefore, the limiting (as $T \rightarrow \infty$) distribution of the statistic in (6.5) is a χ^2 -distribution with $m - p$ degrees of freedom.

7. Methods and tools

In this section we briefly discuss the methods and tools, used to prove the results stated in Sections 4–6.

7.1. Approximation of traces of products of Toeplitz matrices and operators.

The trace approximation problem for truncated Toeplitz operators and matrices has been discussed in detail in the survey paper Ginovyan et al. [54] in the non-tapered case. Here we present some important results in the tapered case, which were used to prove the results stated in Sections 4–6.

Let $\psi(\lambda)$ be an integrable real symmetric function defined on $[-\pi, \pi]$, and let $h(t)$, $t \in [0, 1]$ be a taper function. For $T = 1, 2, \dots$, the $(T \times T)$ -truncated

tapered Toeplitz matrix generated by ψ and h , denoted by $B_T^h(\psi)$, is defined by the following equation:

$$B_T^h(\psi) := \|\widehat{\psi}(t-s)h_T(t)h_T(s)\|, \quad t, s = 1, 2, \dots, T, \quad (7.1)$$

where $\widehat{\psi}(t)$ ($t \in \mathbb{Z}$) are the Fourier coefficients of ψ .

Given a real number $T > 0$ and an integrable real symmetric function $\psi(\lambda)$ defined on \mathbb{R} , the T -truncated tapered Toeplitz operator (also called tapered Wiener-Hopf operator) generated by ψ and a taper function h , denoted by $W_T^h(\psi)$ is defined as follows:

$$[W_T^h(\psi)u](t) = \int_0^T \widehat{\psi}(t-s)u(s)h_T(s)ds, \quad u(s) \in L^2([0, T]; h_T), \quad (7.2)$$

where $\widehat{\psi}(\cdot)$ is the Fourier transform of $\psi(\cdot)$, and $L^2([0, T]; h_T)$ denotes the weighted L^2 -space with respect to the measure $h_T(t)dt$.

Let h be a taper function satisfying Assumption 3.1, and let $A_T^h(\psi)$ be either the $T \times T$ tapered Toeplitz matrix $B_T^h(\psi)$, or the T -truncated tapered Toeplitz operator $W_T^h(\psi)$ generated by a function ψ (see (7.1) and (7.2)).

Observe that, in view of (3.7), (3.8) (3.11), (3.12), (7.1) and (7.2), we have

$$\frac{1}{T} \operatorname{tr} [A_T^h(\psi)] = \frac{1}{T} \cdot 2\pi H_{2,T}(0) \cdot \widehat{\psi}(0) = 2\pi H_2 \int_{\Lambda} \psi(\lambda) d\lambda. \quad (7.3)$$

What happens to the relation (7.3) when $A_T^h(\psi)$ is replaced by a product of Toeplitz matrices (or operators)? Observe that the product of Toeplitz matrices (resp. operators) is not a Toeplitz matrix (resp. operator).

The idea is to approximate the trace of the product of Toeplitz matrices (resp. operators) by the trace of a Toeplitz matrix (resp. operator) generated by the product of the generating functions. More precisely, let $\{\psi_1, \psi_2, \dots, \psi_m\}$ be a collection of integrable real symmetric functions defined on Λ . Let $A_T^h(\psi_i)$ be either the $T \times T$ tapered Toeplitz matrix $B_T^h(\psi_i)$, or the T -truncated tapered Toeplitz operator $W_T^h(\psi_i)$ generated by a function ψ_i and a taper function h . Define

$$S_{A, \mathcal{H}, h}(T) := \frac{1}{T} \operatorname{tr} \left[\prod_{i=1}^m A_T^h(\psi_i) \right], \quad M_{\Lambda, \mathcal{H}, h} := (2\pi)^{m-1} H_m \int_{\Lambda} \left[\prod_{i=1}^m \psi_i(\lambda) \right] d\lambda,$$

where H_m is as in (3.12), and let

$$\Delta(T) := \Delta_{A, \Lambda, \mathcal{H}, h}(T) = |S_{A, \mathcal{H}, h}(T) - M_{\Lambda, \mathcal{H}, h}|. \quad (7.4)$$

Proposition 7.1. *Let $\Delta(T)$ be as in (7.4). Each of the following conditions is sufficient for*

$$\Delta(T) = o(1) \quad \text{as } T \rightarrow \infty. \quad (7.5)$$

(C1) $\psi_i \in L^1(\Lambda) \cap L^{p_i}(\Lambda)$, $p_i > 1$, $i = 1, 2, \dots, m$, with $1/p_1 + \dots + 1/p_m \leq 1$.

(C2) The function $\varphi(\mathbf{u})$ defined by

$$\varphi(\mathbf{u}) := \int_{\Lambda} \psi_1(\lambda)\psi_2(\lambda - u_1)\psi_3(\lambda - u_2) \cdots \psi_m(\lambda - u_{m-1}) d\lambda, \quad (7.6)$$

where $\mathbf{u} = (u_1, u_2, \dots, u_{m-1}) \in \Lambda^{m-1}$, belongs to $L^{m-2}(\Lambda^{m-1})$ and is continuous at $\mathbf{0} = (0, 0, \dots, 0) \in \Lambda^{m-1}$.

Remark 7.1. In the non-tapered case, Proposition 7.1 was proved in Ginovyan et al. [54], while in the tapered case, it was proved in Ginovyan [46]. Proposition 7.1 was used to prove Theorems 4.5, 4.6, 5.2 and 6.2.

Remark 7.2. More results concerning the trace approximation problem for truncated Toeplitz operators and matrices can be found in Ginovyan and Sahakyan [49, 50], and in Ginovyan et al. [54].

7.2. Central limit theorems for tapered quadratic functionals

In this subsection we state central limit theorems for tapered quadratic functional Q_T^h given by (4.3), which were used to prove the results stated in Sections 4–6 (see, e.g., Ginovyan [38], Ginovyan and Sahakyan [48, 49, 50, 52]).

Let $A_T^h(f)$ be either the $T \times T$ tapered Toeplitz matrix $B_T^h(f)$, or the T -truncated tapered Toeplitz operator $W_T^h(f)$ generated by the spectral density f and taper h , and let $A_T^h(g)$ denote either the $T \times T$ tapered Toeplitz matrix, or the T -truncated tapered Toeplitz operator generated by the functions g and h (for definitions see formulas (7.1) and (7.2)).

Remark 7.3. Observe that $A_T^h(f)$ is the covariance matrix in the d.t. case and the covariance operator in the c.t. case, and so is positive definite, while $A_T^h(g)$ need not be positive definite. Also, in view of (7.1) and (7.2), the quadratic functional Q_T^h in (4.3) can be represented in terms of $A_T^h(g)$ as follows:

$$Q_T^h = (A_T^h(g)\mathbf{X}_T, \mathbf{X}_T), \quad (7.7)$$

where $\mathbf{X}_T := \{X(t), t \in D_T\}$ with $D_T := [0, T]$ in the c.t. case and $D_T := \{1, \dots, T\}$ in the d.t. case.

Similar to the non-tapered case, for the distribution of the quadratic functional Q_T^h we have the following result (cf. Ginovyan et al. [54], Grenander and Szegő [57], Ibragimov [65]).

1. The quadratic functional Q_T^h in (7.7) has the same distribution as the sum $\sum_{j=1}^{\infty} \lambda_{j,T} \xi_j^2$, where $\{\xi_j, j \geq 1\}$ are independent $N(0, 1)$ Gaussian random variables and $\{\lambda_{j,T}, j \geq 1\}$ are the eigenvalues of the operator $A_T^h(f) A_T^h(g)$.
2. The characteristic function $\varphi(t)$ of Q_T^h is given by formula:

$$\varphi(t) = \prod_{j=1}^{\infty} |1 - 2it\lambda_{j,T}|^{-1/2}. \quad (7.8)$$

3. The k -th order cumulant $\chi_k(Q_T^h)$ of Q_T^h is given by formula:

$$\chi_k(Q_T) = 2^{k-1}(k-1)! \sum_{j=1}^{\infty} \lambda_{j,T}^k = 2^{k-1}(k-1)! \operatorname{tr} [A_T^h(f) A_T^h(g)]^k. \quad (7.9)$$

Thus, to describe the asymptotic distribution of the quadratic functional Q_T^h , we have to control the traces and eigenvalues of the products of truncated tapered Toeplitz operators and matrices.

7.2.1. CLT for Gaussian models

We assume that the model process $X(t)$ is Gaussian, and with no loss of generality, that $g \geq 0$. We will use the following notation. By \tilde{Q}_T^h we denote the standard normalized quadratic functional:

$$\tilde{Q}_T^h = T^{-1/2} (Q_T^h - \mathbb{E}[Q_T^h]). \quad (7.10)$$

Also, we set

$$\sigma_h^2 := 16\pi^3 H_4 \int_{\Lambda} f^2(\lambda) g^2(\lambda) d\lambda, \quad (7.11)$$

where H_4 is as in (3.12). The notation

$$\tilde{Q}_T^h \xrightarrow{d} \eta \sim N(0, \sigma_h^2) \quad \text{as } T \rightarrow \infty \quad (7.12)$$

will mean that the distribution of the random variable \tilde{Q}_T^h tends (as $T \rightarrow \infty$) to the centered normal distribution with variance σ_h^2 given by (7.11).

The following theorems were proved in Ginovyan and Sahakyan [53].

Theorem 7.1. *Assume that $f \cdot g \in L^1(\Lambda) \cap L^2(\Lambda)$, the taper function h satisfies Assumption 3.1, and for $T \rightarrow \infty$*

$$\chi_2(\tilde{Q}_T^h) = \frac{2}{T} \operatorname{tr} [A_T^h(f) A_T^h(g)]^2 \longrightarrow \sigma_h^2, \quad (7.13)$$

where σ_h^2 is as in (7.11). Then $\tilde{Q}_T^h \xrightarrow{d} \eta \sim N(0, \sigma_h^2)$ as $T \rightarrow \infty$.

Theorem 7.2. *Assume that the function*

$$\varphi(x_1, x_2, x_3) = \int_{\Lambda} f(u) g(u - x_1) f(u - x_2) g(u - x_3) du \quad (7.14)$$

belongs to $L^2(\Lambda^3)$ and is continuous at $(0, 0, 0)$, and the taper function h satisfies Assumption 3.1. Then $\tilde{Q}_T^h \xrightarrow{d} \eta \sim N(0, \sigma_h^2)$ as $T \rightarrow \infty$.

Theorem 7.3. *Assume that $f(\lambda) \in L^p(\Lambda)$ ($p \geq 1$) and $g(\lambda) \in L^q(\Lambda)$ ($q \geq 1$) with $1/p + 1/q \leq 1/2$, and the taper function h satisfies Assumption 3.1. Then $\tilde{Q}_T^h \xrightarrow{d} \eta \sim N(0, \sigma_h^2)$ as $T \rightarrow \infty$.*

Theorem 7.4. Let $f \in L^2(\Lambda)$, $g \in L^2(\Lambda)$, $fg \in L^2(\Lambda)$,

$$\int_{\Lambda} f^2(\lambda)g^2(\lambda - \mu) d\lambda \longrightarrow \int_{\Lambda} f^2(\lambda)g^2(\lambda) d\lambda \quad \text{as } \mu \rightarrow 0, \quad (7.15)$$

and let the taper function h satisfy Assumption 3.1. Then $\tilde{Q}_T^h \xrightarrow{d} \eta \sim N(0, \sigma_h^2)$ as $T \rightarrow \infty$.

To state the next theorem, we recall the class $SV_0(\mathbb{R})$ of slowly varying functions at zero $u(\lambda)$, $\lambda \in \mathbb{R}$, satisfying the following conditions: for some $a > 0$, $u(\lambda)$ is bounded on $[-a, a]$, $\lim_{\lambda \rightarrow 0} u(\lambda) = 0$, $u(\lambda) = u(-\lambda)$ and $0 < u(\lambda) < u(\mu)$ for $0 < \lambda < \mu < a$.

Theorem 7.5. Assume that the functions f and g are integrable on \mathbb{R} and bounded outside any neighborhood of the origin, and satisfy for some $a > 0$

$$f(\lambda) \leq |\lambda|^{-\alpha} L_1(\lambda), \quad |g(\lambda)| \leq |\lambda|^{-\beta} L_2(\lambda), \quad \lambda \in [-a, a], \quad (7.16)$$

for some $\alpha < 1$, $\beta < 1$ with $\alpha + \beta \leq 1/2$, where $L_1(x)$ and $L_2(x)$ are slowly varying functions at zero satisfying

$$L_i \in SV_0(\mathbb{R}), \quad \lambda^{-(\alpha+\beta)} L_i(\lambda) \in L^2[-a, a], \quad i = 1, 2. \quad (7.17)$$

Also, let the taper function h satisfy Assumption 3.1. Then $\tilde{Q}_T^h \xrightarrow{d} \eta \sim N(0, \sigma_h^2)$ as $T \rightarrow \infty$.

The conditions $\alpha < 1$ and $\beta < 1$ in Theorem 7.5 ensure that the Fourier transforms of f and g are well defined. Observe that when $\alpha > 0$ the process $X(t)$ may exhibit long-range dependence. We also allow here $\alpha + \beta$ to assume the critical value $1/2$. The assumptions $f \cdot g \in L^1(\Lambda)$, $f, g \in L^\infty(\Lambda \setminus [-a, a])$ and (7.17) imply that $f \cdot g \in L^2(\Lambda)$, so that the variance σ_h^2 in (7.11) is finite.

7.2.2. CLT for Lévy-driven stationary linear models

Now we assume that the underlying model $X(t)$ is a Lévy-driven stationary linear process defined by (2.6), where $a(\cdot)$ is a filter from $L^2(\mathbb{R})$, and $\xi(t)$ is a Lévy process satisfying the conditions: $\mathbb{E}\xi(t) = 0$, $\mathbb{E}\xi^2(1) = 1$ and $\mathbb{E}\xi^4(1) < \infty$.

The central limit theorem that follows was proved in Ginovyan and Sahakyan [52].

Theorem 7.6. Assume that the filter $a(\cdot)$ and the generating kernel $\hat{g}(\cdot)$ are such that

$$a(\cdot) \in L^p(\mathbb{R}) \cap L^2(\mathbb{R}), \quad \hat{g}(\cdot) \in L^q(\mathbb{R}), \quad 1 \leq p, q \leq 2, \quad 2/p + 1/q \geq 5/2, \quad (7.18)$$

and the taper h satisfies Assumption 3.1. Then $\tilde{Q}_T^h \xrightarrow{d} \eta \sim N(0, \sigma_{L,h}^2)$ as $T \rightarrow \infty$, where

$$\sigma_{L,h}^2 = 16\pi^3 H_4 \int_{\mathbb{R}} f^2(\lambda)g^2(\lambda) d\lambda + \kappa_4 4\pi^2 H_4 \left[\int_{\mathbb{R}} f(\lambda)g(\lambda) d\lambda \right]^2, \quad (7.19)$$

where H_4 is as in (3.12).

Remark 7.4. Notice that if the underlying process $X(t)$ is Gaussian, then in formula (7.19) we have only the first term and so $\sigma_{L,h}^2 = \sigma_h^2$ (see (7.11)), because in this case $\kappa_4 = 0$. On the other hand, the condition (7.18) is more restrictive than the conditions in Theorems 7.1–7.5. Thus, for Gaussian processes Theorems 7.1–7.5 improve Theorem 7.6. For non-tapered case Theorem 7.6 was proved in Bai et al. [7].

7.3. Fejér-type kernels and singular integrals

We define Fejér-type tapered kernels and singular integrals, and state some of their properties.

For a number k ($k = 2, 3, \dots$) and a taper function h satisfying Assumption 3.1 consider the following Fejér-type tapered kernel function:

$$F_{k,T}^h(\mathbf{u}) := \frac{H_T(\mathbf{u})}{(2\pi)^{k-1} H_{k,T}(0)}, \quad \mathbf{u} = (u_1, \dots, u_{k-1}) \in \mathbb{R}^{k-1}, \quad (7.20)$$

where

$$H_T(\mathbf{u}) := H_{1,T}(u_1) \cdots H_{1,T}(u_{k-1}) H_{1,T} \left(-\sum_{j=1}^{k-1} u_j \right), \quad (7.21)$$

and the function $H_{k,T}(\cdot)$ is defined by (3.7).

The next result shows that, similar to the classical Fejér kernel, the tapered kernel $F_{k,T}^h(\mathbf{u})$ is an approximation identity. Here we state the result in the c.t. case (see Ginovyan and Sahakyan [52], Lemma 3.4). In the d.t. case the result was stated in Dahlhaus [18].

Proposition 7.2. *For any $k = 2, 3, \dots$ and a taper function h satisfying Assumption 3.1 the kernel $F_{k,T}^h(\mathbf{u})$, $\mathbf{u} = (u_1, \dots, u_{k-1}) \in \mathbb{R}^{k-1}$, possesses the following properties:*

- a) $\sup_{T>0} \int_{\mathbb{R}^{k-1}} |F_{k,T}^h(\mathbf{u})| d\mathbf{u} = C_1 < \infty$;
- b) $\int_{\mathbb{R}^{k-1}} F_{k,T}^h(\mathbf{u}) d\mathbf{u} = 1$;
- c) $\lim_{T \rightarrow \infty} \int_{\mathbb{E}_\delta^c} |F_{k,T}^h(\mathbf{u})| d\mathbf{u} = 0$ for any $\delta > 0$;
- d) If $k > 2$ for any $\delta > 0$ there exists a constant $M_\delta > 0$ such that for $T > 0$

$$\|F_{k,T}^h\|_{L^{p_k}(\mathbb{E}_\delta^c)} \leq M_\delta, \quad (7.22)$$

where $p_k = \frac{k-2}{k-3}$ for $k > 3$, $p_3 = \infty$, $\mathbb{E}_\delta^c = \mathbb{R}^{k-1} \setminus \mathbb{E}_\delta$ and

$$\mathbb{E}_\delta = \{\mathbf{u} = (u_1, \dots, u_{k-1}) \in \mathbb{R}^{k-1} : |u_i| \leq \delta, i = 1, \dots, k-1\}.$$

- e) If the function $Q \in L^1(\mathbb{R}^{k-1}) \cap L^{k-2}(\mathbb{R}^{k-1})$ and is continuous at $\mathbf{v} = (v_1, \dots, v_{k-1})$ (L^0 is the space of measurable functions), then

$$\lim_{T \rightarrow \infty} \int_{\mathbb{R}^{k-1}} Q(\mathbf{u} + \mathbf{v}) F_{k,T}^h(\mathbf{u}) d\mathbf{u} = Q(\mathbf{v}). \quad (7.23)$$

Denote

$$\Delta_{2,T}^h := \int_{\mathbb{R}^2} f(\lambda)g(\lambda + \mu)F_{2,T}^h(\mu)d\lambda d\mu - \int_{\mathbb{R}} f(\lambda)g(\lambda)d\lambda, \quad (7.24)$$

where $F_{2,T}^h(\mu)$ is given by (7.20) and (7.21).

The next two propositions give information on the rate of convergence to zero of $\Delta_{2,T}^h$ as $T \rightarrow \infty$ (see Ginovyan and Sahakyan [52], Lemmas 4.1 and 4.2).

Proposition 7.3. *Assume that Assumptions 3.1 and 4.3 are satisfied. Then the following asymptotic relation holds:*

$$\Delta_{2,T}^h = o\left(T^{-1/2}\right) \quad \text{as } T \rightarrow \infty. \quad (7.25)$$

Proposition 7.4. *Assume that Assumptions 3.1 and 4.2 are satisfied. Then the following inequality holds:*

$$|\Delta_{2,T}^h| \leq C_h \begin{cases} T^{-(\beta_1+\beta_2)}, & \text{if } \beta_1 + \beta_2 < 1 \\ T^{-1} \ln T, & \text{if } \beta_1 + \beta_2 = 1 \\ T^{-1}, & \text{if } \beta_1 + \beta_2 > 1, \end{cases} \quad T > 0, \quad (7.26)$$

where C_h is a constant depending on h .

Notice that for non-tapered case ($h(t) = \mathbb{I}_{[0,1]}(t)$), the above stated results were proved in Ginovyan and Sahakyan [48] (see also Ginovyan and Sahakyan [49, 50]). In the d.t. tapered case, Proposition 7.3 under different conditions was proved in Dahlhaus [18].

Acknowledgments

The authors would like to thank the associate editor and two anonymous referees for their careful review of the manuscript and valuable remarks, comments and suggestions that led to an improved exposition of this work.

References

- [1] ALOMARI, H. M., FRÍAS, M. P., LEONENKO, N. N., RUIZ-MEDINA, M. D., SAKHNO, L. and TORRES, A. (2017). Asymptotic properties of parameter estimates for random fields with tapered data. *Electronic Journal of Statistics* **11**, 3332–3367. [MR3708540](#)
- [2] ANH, V. V., ANGULO, J. M. and RUIZ-MEDINA, M.D. (1999). Possible long-range dependence in fractional random fields. *J. Statist. Plann. Inference* **80**, 95–110. [MR1713795](#)
- [3] ANH, V. V., LEONENKO, N. N. and MCVINISH, R. (2001). Models for fractional Riesz-Bessel motion and related processes. *Fractals* **9**, 329–346.
- [4] ANH, V. V., LEONENKO, N. N. and SAKHNO, L. (2007). Minimum contrast estimation of random processes based on information of second and third orders. *J. Statist. Planning Inference* **137**, 1302–1331. [MR2301481](#)

- [5] ANH, V. V., LEONENKO, N. N. and SAKHNO, L. (2009). Evaluation of bias in higher-order spectral estimation. *J. Theor. Probab. and Math. Statist.* **80**, 1–14. [MR2541947](#)
- [6] AVRAM, F., LEONENKO, N. N. and SAKHNO, L. (2010). On a Szegő type limit theorem, the Hölder-Young-Brascamp-Lieb inequality, and the asymptotic theory of integrals and quadratic forms of stationary fields. *ESAIM: Probability and Statistics* **14**, 210–255. [MR2741966](#)
- [7] BAI, S., GINOVYAN, M. S. and TAQQU, M. S. (2016). Limit theorems for quadratic forms of Levy-driven continuous-time linear processes. *Stochast. Process. Appl.* **126**, 1036–1065. [MR3461190](#)
- [8] BERAN, J., FENG, Y., GHOSH, S. and KULIK, R. (2013). *Long-Memory Processes: Probabilistic Properties and Statistical Methods*. Springer-Verlag, Berlin. [MR3075595](#)
- [9] BRILLINGER, D. R. (1981). *Time Series: Data Analysis and Theory*. Holden Day, San Francisco. [MR0595684](#)
- [10] BROCKWELL, P. J. (2014). Recent results in the theory and applications of CARMA processes. *Annals of the Institute of Statistical Mathematics* **66**(4), 647–685. [MR3224604](#)
- [11] BROCKWELL, P. J. and DAVIS, R. A. (1991) *Time Series: Theory and Methods*, 2nd ed. Springer-Verlag, New York. [MR1093459](#)
- [12] CASAS, I. and GAO, J. (2008). Econometric estimation in long-range dependent volatility models: Theory and practice. *Journal of Econometrics* **147**, 72–83. [MR2472982](#)
- [13] CHERNOV, H. and LEHMAN, E. L. (1954). The use of maximum likelihood estimates in χ^2 -tests for goodness of fit. *Ann. Math. Stat.*, **25**(3), 579–586. [MR0065109](#)
- [14] CHIBISOV, D. M. (1971). Some tests of chi-square type for continuous distributions. *Theory Probab. Appl.*, **16**(1), 1–20. [MR0283914](#)
- [15] COURSOL, J. and DACUNHA-CASTELLE, D. (1982). Remarks on the approximation of the likelihood function of a stationary Gaussian process. *Theory Probab. Appl.* **27**, 162–167. [MR0645138](#)
- [16] CRAMÉR, H. (1946). *Mathematical Methods of Statistics*, Princeton University Press, Princeton. [MR0016588](#)
- [17] CRAMÉR, H. and LEADBETTER, M. R. (1967) *Stationary and Related Stochastic Processes: Sample Function Properties and Their Applications*. John Wiley & Sons, New York. [MR0217860](#)
- [18] DAHLHAUS, R. (1983). Spectral analysis with tapered data. *J. Time Ser. Anal.* **4**, 163–174. [MR0732895](#)
- [19] DAHLHAUS, R. (1984). Parameter estimation of stationary processes with spectra containing strong peaks. In *Robust and Nonlinear Time Series Analysis* (Franke, Hardle and Martin, eds.) Lecture Notes in Statistics, no. 26, 50–67. [MR0786303](#)
- [20] DAHLHAUS, R. (1985). A functional limit theorem for tapered empirical spectral functions, *Stoch. Process. Appl.* **19**, 135–149. [MR0780726](#)
- [21] DAHLHAUS, R. (1988). Small sample effects in time series analysis: a new asymptotic theory and a new estimate. *Ann. Stat.* **16**, 808–841. [MR0947580](#)

- [22] DAHLHAUS, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.* **17**, 1749–1766. [MR1026311](#)
- [23] DAHLHAUS, R. (1990). Nonparametric High Resolution Spectral Estimation *Probab. Th. Rel. Fields* **85**, 147-180. [MR1050742](#)
- [24] DAHLHAUS, R. (1990). Approximations for the inverse of Toeplitz matrices with applications to stationary processes. *Linear Algebra and its Applications.* **127**, 27–40. [MR1048792](#)
- [25] DAHLHAUS, R. and JANAS, D. (1996). A frequency domain bootstrap for ratio statistics in time series analysis. *Ann. Statist.* **24**(5), 1934–1963. [MR1421155](#)
- [26] DAHLHAUS, R. and KÜNSCH, H. (1987). Edge effects and efficient parameter estimation for stationary random fields. *Biometrika* **74**(4), 877–882. [MR0919857](#)
- [27] DAHLHAUS, R. and WEFELMEYER, W. (1996). Asymptotically optimal estimation in misspecified time series models. *Ann. Statist.* **24**, 952–974. [MR1401832](#)
- [28] DOOB, J. L. (1953). *Stochastic Processes*, Wiley, New York. [MR0058896](#)
- [29] DZHAPARIDZE, K. (1970). On the estimation of spectral parameters of a Gaussian stationary process with rational spectral density. *Theory Probab. Appl.* **15**(3), 531–538. [MR0281308](#)
- [30] DZHAPARIDZE, K. (1986). *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*. Springer, New York. [MR0812272](#)
- [31] DZHAPARIDZE, K. O. and NIKULIN, M. S. (1975). On a Modification of the Standard Statistics of Pearson. *Theory Probab. Appl.*, **19**(4), 851–853. [MR0362633](#)
- [32] DZHAPARIDZE, K. O. and YAGLOM, A. M. (1983). Spectrum parameter estimation in time series analysis. In *Developments in Statistics* (P. R. Krishnaiah, ed.) **4** 1-181. Academic Press, New York. [MR0702676](#)
- [33] FOX, R. and TAQQU, M. S. (1986). Large-sample properties of parameter estimation for strongly dependent stationary Gaussian time series. *Ann. Statist.* **14**, 517–532. [MR0840512](#)
- [34] GAO, J. (2004). Modelling long-range dependent Gaussian processes with application in continuous-time financial models. *J. Appl. Probab.* **41**, 467–482. [MR2052585](#)
- [35] GAO, J., ANH, V. V., HEYDE, C. and TIENG, Q. (2001). Parameter Estimation of Stochastic Processes with Long-range Dependence and Intermittency. *J. Time Ser. Anal.* **22**, 517–535. [MR1859563](#)
- [36] GINOVYAN, M. S. (1988). Asymptotically efficient nonparametric estimation of functionals on spectral density with zeros. *Theory Probab. Appl.* **33** 315–322. [MR0954578](#)
- [37] GINOVYAN, M. S. (1988). On estimate of the value of the linear functional in a spectral density of stationary Gaussian process. *Theory Probab. Appl.* **33**, 777–781. [MR0979749](#)

- [38] GINOVYAN, M. S. (1994). On Toeplitz type quadratic functionals in Gaussian stationary process. *Probab. Theory Relat. Fields* **100**, 395–406. [MR1305588](#)
- [39] GINOVYAN, M. S. (1995). Asymptotic properties of spectrum estimate of stationary Gaussian process. *J. Cont. Math. Anal.* **30** 1–17. [MR1643528](#)
- [40] GINOVYAN, M. S. (2003). Asymptotically efficient nonparametric estimation of nonlinear spectral functionals. *Acta Appl. Math.* **78**, 145–154. [MR2024019](#)
- [41] GINOVYAN, M. S. (2003). Chi-square type goodness-of-fit tests for stationary Gaussian processes. *J. of Cont. Math. Anal.* **38**(2), 1–13. [MR2136319](#)
- [42] GINOVYAN, M. S. (2011). Efficient Estimation of Spectral Functionals for Continuous-time Stationary Models. *Acta Appl. Math.* **115**(2), 233–254. [MR2818916](#)
- [43] GINOVYAN, M. S. (2011). Efficient Estimation of Spectral Functionals for Gaussian Stationary Models. *Comm. Stochast. Anal.* **5**(1) (2011), 211–232. [MR2808543](#)
- [44] GINOVYAN, M. S. (2018). Goodness-of-fit tests for continuous-time stationary processes. *J. of Cont. Math. Anal.* **53**(3), 121–128. [MR3855809](#)
- [45] GINOVYAN, M. S. (2020). Parameter estimation for Lévy-driven continuous-time linear models with tapered data. *Acta Appl. Math.* **169**, 79–97. [MR4146891](#)
- [46] GINOVYAN, M. S. (2021). Goodness-of-fit tests for stationary Gaussian processes with tapered data. *Acta Appl. Math.* **171**(1), 1–12. [MR4187307](#)
- [47] GINOVYAN, M. S. and SAHAKYAN, A. A. (2005). On the Central Limit Theorem for Toeplitz Quadratic Forms of Stationary Sequences. *Theory Probab. and Appl.* **49**, 612–628. [MR2142560](#)
- [48] GINOVYAN, M. S. and SAHAKYAN, A. A. (2007). Limit Theorems for Toeplitz quadratic functionals of c.t. stationary process, *Probab. Theory Relat. Fields* **138**, 551–579. [MR2299719](#)
- [49] GINOVYAN, M. S. and SAHAKYAN, A. A. (2012). Trace approximations of products of truncated Toeplitz operators. *Theory Probab. Appl.* **56**(1), 57–71. [MR2848419](#)
- [50] GINOVYAN, M. S. and SAHAKYAN, A. A. (2013). On the trace approximations of products of Toeplitz matrices. *Statist. Probab. Lett.* **83**(3), 753–760. [MR3040300](#)
- [51] GINOVYAN, M. S. and SAHAKYAN, A. A. (2016). Robust estimation for continuous-time linear models with memory. *Theory Probability and Mathematical Statistics.* **95**, 75–91. [MR3631645](#)
- [52] GINOVYAN, M. S. and SAHAKYAN, A. A. (2019). Estimation of spectral functionals for Levy-driven continuous-time linear models with tapered data, *Electronic Journal of Statistics*, **13**, 255–283. [MR3905127](#)
- [53] GINOVYAN, M. S. and SAHAKYAN, A. A. (2019). Limit theorems for tapered Toeplitz quadratic functionals of continuous-time Gaussian stationary processes, *J. Cont. Math. Anal.* **54**(4), 222–239. [MR4019542](#)
- [54] GINOVYAN, M. S., SAHAKYAN, A. A. and TAQQU, M. S. (2014). The trace problem for Toeplitz matrices and operators and its impact in probability,

- Probability Surveys* **11**, 393–440. [MR3290440](#)
- [55] GIRAITIS, L., KOUL, H. and SURGAILIS, D. (2012). *Large Sample Inference for Long Memory Processes*. Imperial College Press, London. [MR2977317](#)
- [56] GIRAITIS and SURGAILIS, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle’s estimate. *Probab. Theory Relat. Fields* **86**, 87–104. [MR1061950](#)
- [57] GRENDER, U. and SZEGŐ, G. (1958). *Toeplitz Forms and Their Applications*. University of California Press, Berkeley and Los Angeles. [MR0094840](#)
- [58] GUÉGAN, D. (2005). How can we define the concept of long memory? An econometric survey. *Econometric Reviews* **24**, 113–149. [MR2190313](#)
- [59] GUYON, X. (1982). Parameter estimation for a stationary process on a d -dimensional lattice. *Biometrika* **69**, 95–105. [MR0655674](#)
- [60] GUYON, X. (1995). *Random Fields on a Network: Modelling, Statistics and Applications*. Springer, New York. [MR1344683](#)
- [61] HANNAN, E. J. (1960). *Time Series Analysis*, John Wiley, New York. [MR0114281](#)
- [62] HASMINSKII, R. Z. and IBRAGIMOV, I. A. (1986). Asymptotically efficient nonparametric estimation of functionals of a spectral density function. *Probab. Theory Related Fields*. **73**, 447–461. [MR0859842](#)
- [63] HEYDE, C. C. and DAI, W. (1996). On the robustness to small trends of estimation based on the smoothed periodogram. *J. Time Ser. Anal.* **17**(2), 141–150. [MR1381169](#)
- [64] HEYDE, C. C. and GAY, R. (1993). Smoothed periodogram asymptotics and estimation for processes and fields with possible long-range dependence. *Stochastic Processes and their Applications*. **45**, 169–182. [MR1204868](#)
- [65] IBRAGIMOV, I. A. (1963). On estimation of the spectral function of a stationary Gaussian process. *Theory Probab. Appl.* **8**, 366–401. [MR0160274](#)
- [66] IBRAGIMOV, I. A. (1967). On maximum likelihood estimation of parameters of the spectrum of stationary time series. *Theory Probab. Appl.* **12**, 115–119. [MR0228095](#)
- [67] IBRAGIMOV, I. A. and KHASHMINSKII, R. Z. (1991). Asymptotically normal families of distributions and efficient estimation. *Ann. Statist.*, **19**, 1681–1724. [MR1135145](#)
- [68] IBRAGIMOV, I. A. and LINNIK, YU. V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff Publishing Groningen, The Netherlands. [MR0322926](#)
- [69] JANAS, D. and VON SACHS, R. (1994). Consistency for non-linear functions of the periodogram of tapered data. *J. Time Ser. Anal.*, **16**(6), 585–606. [MR1366801](#)
- [70] LEONENKO, N. N. and SAKHNO, L. M. (2006). On the Whittle estimators for some classes of continuous-parameter random processes and fields. *Stat. Probab. Lett.* **76**, 781–795. [MR2266092](#)
- [71] LUDEÑA, C. and LAVIELLE, M. (1999). The Whittle Estimator for Strongly Dependent Stationary Gaussian Fields. *Scand. J. Stat.* **26**, 433–450. [MR1712039](#)

- [72] MILLAR, P. W. (1985). Non-parametric applications of an infinite dimensional convolution theorem. *Z. Wahrsch. verw. Gebiete*, **68**, 545–556. [MR0772198](#)
- [73] OSIDZE, A. G. (1974). On a goodness of fit test in the case of dependence of spectral density of Gaussian processes on unknown parameters. *Reports of AN Georgian SSR*, **74**(2), 273–276. [MR0356416](#)
- [74] OSIDZE, A. G. (1975). On a statistic for testing the composite hypothesis regarding the form of a spectral density of a stationary Gaussian random process. *Reports of AN Georgian SSR*, **77**(2), 313–315. [MR0378320](#)
- [75] PISARENKO, V. F. (1965). On the computation of the relation of likelihood for Gaussian processes with rational spectrum. *Theory Prob. Applications*, **10**, 299–303. [MR0183025](#)
- [76] PRIESTLEY, M. (1981). *Spectral Analysis and Time Series*, Vol. I. Academic Press, 1st Edition. [MR0628735](#)
- [77] SAMORODNISKY, G. and TAQQU, M.S. (1994). *Stable Non-Gaussian Random Processes*. Chapman & Hall/CRC, New York. [MR1280932](#)
- [78] TANIGUCHI, M. (1987). Minimum contrast estimation for spectral densities of stationary processes. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* **49**, 315–325. [MR0928940](#)
- [79] TANIGUCHI, M. and KAKIZAWA, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. Academic Press, New York. [MR1785484](#)
- [80] TUKEY, J. W. (1967). An introduction to the calculations of numerical spectrum analysis. In *Advanced seminar on spectral analysis of time series* (Harris, B. (ed.), 25–46, Wiley, New York. [MR0212977](#)
- [81] TSAI, H. and CHAN, K. S. (2005). Quasi-maximum likelihood estimation for a class of continuous-time long memory processes. *J. Time Ser. Anal.* **26**(5), 691–713. [MR2188305](#)
- [82] VON SACHS, R. (1994). Estimating non-linear functions of spectral density, using a data-taper. *Ann. Inst. Statist. Math.*, **46**(3), 453–474. [MR1309717](#)
- [83] WALKER, A. M. (1964). Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series. *J. Austr. Math. Soc.* **4**, 363–384. [MR0171345](#)
- [84] WHITTLE, P. (1951). *Hypothesis Testing in Time Series*. Hafner, New York. [MR0040634](#)
- [85] YAGLOM, A. M. (1986). *The Correlation Theory of Stationary and Related Random Processes, Vol. 1*. Springer, New York. [MR0915557](#)