Известия НАН Армении, Математика, том 56, н. 6, 2021, стр. 12 – 38.

STATISTICAL ESTIMATION FOR STATIONARY MODELS WITH TAPERED DATA

M. S. GINOVYAN, A. A. SAHAKYAN

Boston University, Boston, MA, USA
Yerevan State University
E-mails: *ginovyan@math.bu.edu*, *sart@ysu.am*

Abstract. In this paper, we survey some recent results on parametric and nonparametric statistical estimation about the spectrum of stationary models with tapered data, as well as, a question concerning robustness of inferences, carried out on a linear stationary process contaminated by a small trend. We also discuss some questions concerning tapered Toeplitz matrices and operators, central limit theorems for tapered Toeplitz type quadratic functionals, and tapered Fejér-type 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.

MSC2010 numbers: 62F10; 62F12; 62G05; 60G10.
Keywords: tapered data; stationary processes; spectral density; parametric and nonparametric estimation; robustness.

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 $\mathbb{U} = \mathbb{Z} := \{0, \pm 1, \pm 2, \ldots\}$. 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) 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 $D_T := \{1, \ldots, T\}$ 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 Avram et al. [3], Casas and Gao [8], Dahlhaus [12], Dahlhaus and Wefelmeyer [14], Dzhaparidze [15], Dzhaparidze and Yaglom [16], Fox and Taqqu [17], Gao [18], Gao et al. [19], Ginovyan [20, 21, 24, 25], Giraitis et al. [37], Giraitis and Surgailis [38], Guyon [40], Has'minskii and Ibragimov [41], Heyde and Dai [42], Ibragimov [43, 44], Ibragimov and Khas'minskii [45], Leonenko and Sakhno [47], Taniguchi

[49], Taniguchi and Kakizawa [50], Tsai and Chan [52], Walker [53], Whittle [54], 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 $h_T(t) := h(t/T)$ with $h(t)$, $t \in \mathbb{R}$ being a *taper function*.

The use of data tapers in nonparametric time series was suggested by Tukey [51]. The benefits of tapering the data have been widely reported in the literature (see, e.g., Brillinger [6], Dahlhaus [10, 11], Dahlhaus and Künsch [13], Guyon [40], 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'.

In this paper, we survey some recent results on parametric and nonparametric statistical estimation about the spectrum of stationary models with tapered data, as well as, a question concerning robustness of inferences, carried out on a linear stationary process contaminated by a small trend. 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.

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, and introduce the data tapers and tapered periodogram. In Section 3 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 4 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. A question concerning robustness of inferences, carried out on a linear stationary process contaminated by a small trend is discussed in Section 5. In Section 6 we briefly discuss the methods and tools, used to prove the results stated in Sections 3–5.

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, and introduce the data tapers and tapered periodogram.

2.1. **The model.** *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 $(\Omega, \mathcal{F}, P)$ 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 $\mathbb{U} = \mathbb{Z} := \{0, \pm 1, \pm 2, \ldots\}$. We assume that $X(u)$ is a *non-degenerate process*, that is, $\mathrm{Var}[X(u)] = \mathbb{E}|X(u)|^2 = r(0) > 0$. (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, $\mathbb{E}[X(t) - X(s)]^2 \to 0$ as $t \to s$.

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 [9]), 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*:

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

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

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

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

Linear processes. Existence of spectral density functions. We consider here stationary processes possessing spectral densities. For the following results we refer to Ibragimov and Linnik [46].

(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) if and only if it can be represented as an infinite moving average:

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

where $\{\xi(k), k \in \mathbb{Z}\} \sim$ 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:

$$r(u) = \sum_{k=-\infty}^{\infty} a(u+k)a(k), \quad f(\lambda) = \frac{1}{2\pi}\left|\sum_{k=-\infty}^{\infty} a(k)e^{-ik\lambda}\right|^2, \quad \lambda \in [-\pi, \pi]. \tag{2.4}$$

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) if and only if it can be represented as an infinite continuous moving average:

$$X(u) = \int_{\mathbb{R}} a(u-t)d\xi(t),, \qquad \int_{\mathbb{R}} |a(t)|^2 dt < \infty, \tag{2.5}$$

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:

$$r(u) = \int_{\mathbb{R}} a(u+x)a(x)dx, \quad f(\lambda) = \frac{1}{2\pi}\left|\int_{\mathbb{R}} e^{-i\lambda t}a(t)dt\right|^2, \quad \lambda \in \mathbb{R}. \tag{2.6}$$

The function $a(\cdot)$ in representations (2.3) and (2.5) plays the role of a *time-invariant filter*, and the linear processes defined by (2.3) and (2.5) 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 (2.5) 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 models and their continuous counterparts the c.t. autoregressive moving average models are of the form (2.3) and (2.5), respectively, and play a central role in the representations of stationary time series (see, e.g., Brockwell and Davis [7]).

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

(2.5), 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 Bai et al. [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. [5], Giraitis et al. [37], Guégan [39]). 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 $0 < C_1 \le f(\lambda) \le C_2 < \infty$.

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 spectral density $f(\lambda)$ of (ARMA)(p, q) process is a rational function (see, e.g., Brockwell and Davis [7], Section 3.1):

$$f(\lambda) = \frac{\sigma^2}{2\pi} \cdot \frac{|\theta_q(e^{-i\lambda})|^2}{|\psi_p(e^{-i\lambda})|^2}. \tag{2.7}$$

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., Tsai and Chan [52]):

$$f(\lambda) = \frac{\sigma^2}{2\pi} \cdot \frac{|\beta_q(i\lambda)|^2}{|\alpha_p(i\lambda)|^2}, \tag{2.8}$$

where $\alpha_p(z)$ and $\beta_q(z)$ are polynomials of degrees p and q, respectively.

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. [5], Guégan [39], 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 model that displays long-memory or is anti-persistent is the Autoregressive Fractionally Integrated Moving Average (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

$$f_X(\lambda) = |1 - e^{-i\lambda}|^{-2d} f(\lambda) = (2\sin(\lambda/2))^{-2d} f(\lambda), \quad d < 1/2, \tag{2.9}$$

where $f(\lambda)$ is the spectral density of an ARMA(p, q) process, given by (2.7). Observe that for $0 < d < 1/2$ the model $X(t)$ specified by the spectral density (2.9) 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 (2.9) is not integrable, and thus it cannot represent a spectral density of a stationary process.

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) $\{B_H(t), t \in \mathbb{R}\}$ with Hurst index H, $0 < H < 1$, defined to be a centered Gaussian H-self-similar process having stationary increments. The fBm B_H can be regarded as a Gaussian process having a 'spectral density':

$$f(\lambda) = c|\lambda|^{-(2H+1)}, \quad c > 0, \quad 0 < H < 1, \quad \lambda \in \mathbb{R}. \tag{2.10}$$

The form (2.10) can be understood in a generalized sense (see, e.g., Yaglom [55]), 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. [1], and defined as a c.t. Gaussian process $X(t)$

with spectral density

$$f(\lambda) = c\,|\lambda|^{-2\alpha}(1+\lambda^2)^{-\beta}, \quad \lambda \in \mathbb{R},\, 0 < c < \infty,\, 0 < \alpha < 1,\, \beta > 0. \tag{2.11}$$

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. [2] and Gao et al. [19]).

Notice that the process $X(t)$, specified by the spectral density (2.11), is stationary if $0 < \alpha < 1/2$ and is non-stationary with stationary increments if $1/2 \le \alpha < 1$.

Comparing (2.10) and (2.11), we observe that the spectral density of fBm is the limiting case as $\beta \to 0$ 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 $f(\lambda)$ of a CARFIMA(p, H, q) process is given by formula (see, e.g., Tsai and Chan [52]):

$$f(\lambda) = \frac{\sigma^2}{2\pi}\Gamma(2H+1)\sin(\pi H)|\lambda|^{1-2H}\frac{|\beta_q(i\lambda)|^2}{|\alpha_p(i\lambda)|^2}, \tag{2.12}$$

where $\alpha_p(z)$ and $\beta_q(z)$ are polynomials of degrees p and q, respectively. Notice that for $H = 1/2$, the spectral density given by (2.12) becomes that of the short-memory CARMA(p, q) process, given by (2.8).

2.2. **Data tapers and tapered periodogram.** Our inference procedures will be based on the tapered data $\mathbf{X}_T^h$:

$$\mathbf{X}_T^h := \begin{cases} \{h_T(t)X(t),\, t = 1, \ldots, T\} & \text{in the d.t. case,} \\ \{h_T(t)X(t),\, 0 \le t \le T\} & \text{in the c.t. case,} \end{cases} \tag{2.13}$$

where

$$h_T(t) := h(t/T) \tag{2.14}$$

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

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

Assumption 2.1. The taper $h : \mathbb{R} \to \mathbb{R}$ is a continuous nonnegative function of bounded variation and of bounded support $[0, 1]$, such that $H_k \neq 0$, where

$$H_k := \int_0^1 h^k(t)dt, \quad k \in \mathbb{N} := \{1, 2, \ldots\}. \tag{2.15}$$

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.

Remark 2.1. For the d.t. case, an example of a taper function $h(t)$ satisfying Assumption 2.1 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 2.1 is the function $h(t) = 1 - t$ for $t \in [0, 1]$.

Denote by $H_{k,T}(\lambda)$ the *tapered Dirichlet type kernel*, defined by

$$H_{k,T}(\lambda) := \begin{cases} \sum_{t=1}^{T} h_T^k(t) e^{-i\lambda t} & \text{in the d.t. case,} \\ \int_0^T h_T^k(t) e^{-i\lambda t} dt & \text{in the c.t. case.} \end{cases} \tag{2.16}$$

Define the finite Fourier transform of the tapered data (2.13):

$$d_T^h(\lambda) := \begin{cases} \sum_{t=1}^{T} h_T(t) X(t) e^{-i\lambda t} & \text{in the d.t. case,} \\ \int_0^T h_T(t) X(t) e^{-i\lambda t} dt & \text{in the c.t. case.} \end{cases} \tag{2.17}$$

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

$$I_T^h(\lambda) := \frac{1}{C_T} d_T^h(\lambda) d_T^h(-\lambda) = \tag{2.18}$$

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

where

$$C_T := 2\pi H_{2,T}(0) \neq 0. \tag{2.19}$$

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

3. Nonparametric estimation problem

Suppose we observe a finite realization $\mathbf{X}_T := \{X(u), 0 \leq u \leq T$ (or $u = 1, \ldots, T$ 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 $T \to \infty$) 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., Dzhaparidze [15], Guyon [40], Leonenko and Sakhno [47], Taniguchi and Kakizawa [50], 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 [34, 35]. For non-tapered case, these properties were established in the papers Ginovyan [22, 25]. 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 4). Here we follow the papers Ginovyan [23, 25, 26], and Ginovyan and Sahakyan [34, 35].

3.1. **Estimation of linear spectral functionals.** We are interested in the nonparametric estimation problem, based on the tapered data (2.13), of the following linear spectral functional:

$$J = J(f,g) := \int_\Lambda f(\lambda)g(\lambda)d\lambda, \tag{3.1}$$

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

As an estimator J_T^h for functional $J(f)$, given by (3.1), based on the tapered data (2.13), we consider the averaged tapered periodogram (or a simple 'plug-in' statistic), defined by

$$J_T^h \quad = \quad J(I_T^h) := \int_\Lambda I_T^h(\lambda)g(\lambda)d\lambda, \tag{3.2}$$

where $I_T^h(\lambda)$ is the tapered periodogram of the process $X(t)$ given by (2.18). Denote

$$Q_T^h := \begin{cases} \sum_{t=1}^T \sum_{s=1}^T \widehat{g}(t-s)h_T(t)h_T(s)X(t)X(s) & \text{in the d.t. case,} \\ \\ \int_0^T \int_0^T \widehat{g}(t-s)h_T(t)h_T(s)X(t)X(s)\,dt\,ds & \text{in the c.t. case,} \end{cases} \tag{3.3}$$

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

$$\widehat{g}(t) := \int_\Lambda e^{i\lambda t}g(\lambda)d\lambda, \quad t \in \Lambda. \tag{3.4}$$

In view of (2.18) and (3.2) – (3.4) we have

$$J_T^h = C_T^{-1}Q_T^h, \tag{3.5}$$

where C_T is as in (2.19). 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 $T \to \infty$) of the tapered Toeplitz type quadratic functional Q_T^h given by (3.3) (for details see Section 6.2).

3.2. **Asymptotic unbiasedness.** We begin with the following assumption.

Assumption 3.1. The function

$$\Psi(u) = \int_\Lambda f(v)g(u+v)\,dv \tag{3.6}$$

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

Theorem 3.1. *Let the functionals $J := J(f, g)$ and $J_T^h := J(I_T^h, g)$ be defined by (3.1) and (3.2), respectively. Then under Assumptions 2.1 and 3.1 the statistic J_T^h is an asymptotically unbiased estimator for $J(f)$, that is, the following relation holds:*

$$\lim_{T\to\infty} [E(J_T^h) - J] = 0. \tag{3.7}$$

Remark 3.1. 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 $1 \le p_1, p_2 \le \infty$, $1/p_1 + 1/p_2 \le 1$, then the relation (3.7) is satisfied.

Under additional smoothness conditions on functions $f(\lambda)$ and $g(\lambda)$ we can estimate the rate of convergence in (3.7). To state the corresponding result, we first introduce some notation and assumptions.

Given numbers $p \ge 1$, $0 < \alpha < 1$, $r \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$, 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

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

Assumption 3.2. 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) $\beta_1 \le 1/p$, $\beta_2 \le 1/q$ and $\beta_1 + \beta_2 > 1/2$,

c) $\beta_1 > 1/p$, $1/q - 1/2 < \beta_2 \le 1/q$,

d) $\beta_2 > 1/q$, $1/p - 1/2 < \beta_1 \le 1/p$.

Remark 3.2. In Ginovian [22] 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 $1/p_1 + 1/q_1 \le 1/2$.

Assumption 3.3. The spectral density f and the generating function g are such that $f, g \in L^1(\Lambda) \cap L^2(\Lambda)$ 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 3.2. *Let the functionals $J := J(f, g)$ and $J_T^h := J(I_T^h, g)$ be defined by (3.1) and (3.2), respectively. Then under Assumptions 2.1 and 3.2 (or 3.3), the* *following asymptotic relation holds:*

$$T^{1/2}\left[\mathbb{E}(J_T^h) - J\right] \to 0 \quad \text{as} \quad T \to \infty. \tag{3.8}$$

Remark 3.3. We call an estimator J_T^h of J *asymptotically unbiased of the order of* T^β, $\beta > 0$ if $\lim_{T\to\infty} T^\beta[E(J_T^h) - J] = 0$. Thus, Theorem 3.2 states that the statistic J_T^h is an asymptotically unbiased estimator for J of the order of $T^{1/2}$.

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

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

Assumption 3.4. The filter $a(\cdot)$ and the generating kernel $\widehat{g}(\cdot)$ are such that

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

We begin with a result on the asymptotic behavior of the variance $\mathrm{Var}(J_T^h) = \mathbb{E}(J_T^h - \mathbb{E}(J_T^h))^2$. The proof of the next theorem can be found in Ginovyan and Sahakyan [34].

Theorem 3.3. *Let the functionals $J := J(f, g)$ and $J_T^h := J(I_T^h, g)$ be defined by (3.1) and (3.2), respectively. Then under Assumptions 2.1 and 3.4 the following asymptotic relation holds:*

$$\lim_{T\to\infty} T\mathrm{Var}(J_T^h) = \sigma_h^2(J), \tag{3.9}$$

where

$$\sigma_h^2(J) := 4\pi e(h) \int_\Lambda f^2(\lambda) g^2(\lambda) d\lambda + \kappa_4 e(h) \left[\int_\Lambda f(\lambda) g(\lambda) d\lambda\right]^2. \tag{3.10}$$

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

$$e(h) := \frac{H_4}{H_2^2} = \int_0^1 h^4(t)dt \left(\int_0^1 h^2(t)dt\right)^{-2}. \tag{3.11}$$

From Theorems 3.1–3.3 we infer the following result.

Theorem 3.4. *The following assertions hold.*

(a) *Under Assumptions 2.1, 3.1 and 3.4 the statistic J_T^h is a mean square consistent estimator for J.*
(b) *Under Assumptions 2.1, 3.2 (or 3.3) and 3.4 the statistic J_T^h is a $\sqrt{T}$-consistent in the mean square sense estimator for J.*

3.4. **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 [34].

Theorem 3.5 (CLT). *Let $J := J(f,g)$ and $J_T^h := J(I_T^h, g)$ be defined by (3.1) and (3.2), respectively. Then under Assumptions 2.1 and 3.4 the functional J_T^h obeys the central limit theorem. More precisely, we have*

$$T^{1/2}\left[J_T^h - \mathbb{E}(J_T^h)\right] \xrightarrow{d} \eta \quad \text{as} \quad T\to\infty, \tag{3.12}$$

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 (3.10) *and* (3.11).

Taking into account the equality

$$T^{1/2}\left[J_T^h - J\right] = T^{1/2}\left[\mathbb{E}(J_T^h) - J\right] + T^{1/2}\left[J_T^h - \mathbb{E}(J_T^h)\right], \tag{3.13}$$

as an immediate consequence of Theorems 3.2 and 3.5, 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 3.6. *Let the functionals $J := J(f,g)$ and $J_T^h := J(I_T^h, g)$ be defined by (3.1) and (3.2), respectively. Then under Assumptions 2.1, 3.2 (or 3.3) and 3.4 the statistic J_T^h is an asymptotically normal estimator for functional J. More precisely, we have*

$$T^{1/2}\left[J_T^h - J\right] \xrightarrow{d} \eta \quad \text{as} \quad T\to\infty, \tag{3.14}$$

where η is as in Theorem 3.5, that is, η is a normally distributed random variable with mean zero and variance $\sigma_h^2(J)$ given by (3.10) *and* (3.11).

Remark 3.4. Notice that if the underlying process $X(u)$ is Gaussian, then in formula (3.10) we have only the first term. Using the results from Ginovyan [22] and Ginovyan and Sahakyan [29, 30], it can be shown that in this case Theorem 3.6 is true under Assumptions 2.1 and 3.4.

4. 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 (2.13), and investigate the asymptotic (as $T\to\infty$) 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.

4.1. **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 [54]). 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 [12], Dzhaparidze [15], Fox and Taqqu [17], Giraitis and Surgailis [38], Guyon [40], Taniguchi and Kakizawa [50], Walker [53], and references therein).

For c.t. models, the Whittle estimation procedure has been considered, for example, in Avram et al. [3], Casas and Gao [8], Dzhaparidze and Yaglom [16], Gao [18], Gao et al. [19], Leonenko and Sakhno [47], Tsai and Chan [52], 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 Dahlhaus [10], Dahlhaus and Künsch [13], Guyon [40], Ludeña and Lavielle [48]. 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 [27].

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 $X_T := \{X(t), t = 1, \dots, T\}$. A natural approach is to find the maximum likelihood estimator (MLE) $\widehat{\theta}_{T,MLE}$ 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:

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

where $B_T(f_\theta)$ is the Toeplitz matrix generated by f_θ. Unfortunately, the above function is difficult to handle, and no explicit expression for the estimator $\widehat{\theta}_{T,MLE}$ 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 [54], and called Whittle functional:

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

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

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

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

then the MLE $\widehat{\theta}_{T,MLE}$ and the Whittle estimator $\widehat{\theta}_T$ are asymptotically equivalent in the sense that $\widehat{\theta}_T$ also is consistent, asymptotically normal and asymptotically Fisher-efficient (see, e.g., Dzhaparidze and Yaglom [16]).

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

$$U_T(\theta) := \frac{1}{4\pi}\int_{\mathbb{R}}\left[\ln f(\lambda,\theta) + \frac{I_T(\lambda)}{f(\lambda,\theta)}\right]\cdot w(\lambda)\,d\lambda, \tag{4.1}$$

where $I_T(\lambda)$ is the continuous periodogram of $X(t)$, and $w(\lambda)$ is a weight function ($w(-\lambda) = w(\lambda)$, $w(\lambda)\ge 0$, $w(\lambda)\in L^1(\mathbb{R})$) for which the integral in (4.1) 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 (2.13) is to choose the estimator $\widehat{\theta}_{T,h}$ to minimize the weighted tapered Whittle functional:

$$U_T^h(\theta) := \frac{1}{4\pi}\int_{\Lambda}\left[\log f(\lambda,\theta) + \frac{I_T^h(\lambda)}{f(\lambda,\theta)}\right]\cdot w(\lambda)\,d\lambda, \tag{4.2}$$

where $I_T^h(\lambda)$ is the tapered periodogram of $X(t)$, given by (2.18), and $w(\lambda)$ is a weight function for which the integral in (4.2) is well defined. Thus, the Whittle estimator $\widehat{\theta}_{T,h}$ of θ based on the tapered sample (2.13) is defined by

$$\widehat{\theta}_{T,h} := \operatorname*{Arg\,min}_{\theta\in\Theta} U_T^h(\theta). \tag{4.3}$$

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

Assumption 4.1. The true value θ_0 of the parameter θ belongs to a compact set Θ, which is contained in an open set S 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 4.2. 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 4.3. The functions $f := f(\lambda, \theta)$ and $g := w(\lambda) \frac{\partial}{\partial \theta_k} f^{-1}(\lambda, \theta)$ satisfy Assumptions 3.3 or 3.4 for all $k = 1, \dots, p$ and $\theta \in \Theta$.

Assumption 4.4. The functions $a := a(\lambda, \theta)$ and $b := \widehat{g}$, where g is as in Assumption 4.3, satisfy Assumption 3.1.

Assumption 4.5. The functions $\frac{\partial^2}{\partial \theta_k \partial \theta_j} f^{-1}(\lambda, \theta)$ and $\frac{\partial^3}{\partial \theta_k \partial \theta_j \partial \theta_j} f^{-1}(\lambda, \theta)$, $k, j, l = 1, \dots, p$, 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 4.6. The matrices

$$W(\theta) := \|w_{ij}(\theta)\|, \ A(\theta) := \|a_{ij}(\theta)\|, \ B(\theta) := \|b_{ij}(\theta)\|, \ i, j = 1, \dots, p \tag{4.4}$$

are positive definite, where

$$w_{ij}(\theta) = \frac{1}{4\pi} \int_\Lambda \frac{\partial}{\partial \theta_i} \ln f(\lambda, \theta) \frac{\partial}{\partial \theta_j} \ln f(\lambda, \theta) w(\lambda) d\lambda, \tag{4.5}$$

$$a_{ij}(\theta) = \frac{1}{4\pi} \int_\Lambda \frac{\partial}{\partial \theta_i} \ln f(\lambda, \theta) \frac{\partial}{\partial \theta_j} \ln f(\lambda, \theta) w^2(\lambda) d\lambda, \tag{4.6}$$

$$b_{ij}(\theta) = \frac{\kappa_4}{16\pi^2} \int_\Lambda \frac{\partial}{\partial \theta_i} \ln f(\lambda, \theta) w(\lambda) d\lambda \int_{\mathbb{R}} \frac{\partial}{\partial \theta_j} \ln f(\lambda, \theta) w(\lambda) d\lambda, \tag{4.7}$$

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

The next theorem contains sufficient conditions for Whittle estimator to be consistent (see Ginovyan [27]).

Theorem 4.1. *Let $\widehat{\theta}_{T,h}$ be the Whittle estimator defined by* (4.3) *and let θ_0 be the true value of parameter θ. Then, under Assumptions 4.1–4.4 and 2.1, the statistic $\widehat{\theta}_{T,h}$ is a consistent estimator for θ, that is, $\widehat{\theta}_{T,h} \to \theta_0$ in probability as $T \to \infty$.*

Having established the consistency of the Whittle estimator $\widehat{\theta}_{T,h}$, we can go on to obtain the limiting distribution of $T^{1/2}\left(\widehat{\theta}_{T,h} - \theta_0\right)$ 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 Whittle estimator $\widehat{\theta}_{T,h}$ is asymptotically normal (see Ginovyan [27]).

Theorem 4.2. *Suppose that Assumptions 4.1–4.6 and 2.1 are satisfied. Then the Whittle estimator $\widehat{\theta}_{T,h}$ of an unknown spectral parameter θ based on the tapered data* (2.13) *is asymptotically normal. More precisely, we have*

$$T^{1/2}\left(\widehat{\theta}_{T,h}-\theta_0\right) \xrightarrow{d} N_p\left(0, e(h)\Gamma(\theta_0)\right) \quad \text{as} \quad T\to\infty, \tag{4.8}$$

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

$$\Gamma(\theta_0) = W^{-1}(\theta_0)\left(A(\theta_0)+B(\theta_0)\right)W^{-1}(\theta_0), \tag{4.9}$$

where the matrices W, A and B are defined in (4.4)-(4.7), and the tapering factor $e(h)$ is given by formula (3.11).

Remark 4.1. 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 (4.4) – (4.6)). So, in this case, formula (4.9) becomes $\Gamma(\theta_0) = W^{-1}(\theta_0)\left(W(\theta_0)+B(\theta_0)\right)W^{-1}(\theta_0)$. If, in addition, the underlying process is Gaussian ($\kappa_4 = 0$, and hence $B(\theta_0) = 0$), and the taper h is chosen so that the tapering factor $e(h)$ is equal to one, then we have $\Gamma(\theta_0) = W^{-1}(\theta_0)$, that is, the Whittle estimator $\widehat{\theta}_{T,h}$ is Fisher-efficient.

5. Robustness to small trends of estimation

In time series analysis, much of statistical inferences about unknown spectral parameters or spectral functionals are concerned with the stationary models, in which case it is assumed that the models are centered, or have constant means. In this section, we are concerned with the robustness of inferences, carried out on a stationary models, possibly exhibiting long memory, contaminated by a small trend. Specifically, let $\{X(t),\, t\in\mathbb{U}\}$ be a centered stationary process possessing a spectral density $f_X(\lambda)$, $\lambda\in\Lambda$. Assuming that either f_X is known with the exception of a vector parameter $\theta\in\Theta\subset\mathbb{R}^p$, or f_X is completely unknown and belongs to a given class $\mathcal{F}$, we want to make inferences about θ or the value $J(f_X)$ of a given functional $J(\cdot)$ at an unknown point $f_X\in\mathcal{F}$ in the case where the actual observed data are in the contaminated form:

$$Y(t) = X(t) + M(t), \quad t\in D_T, \tag{5.1}$$

where $M(t)$ is a deterministic trend, and $D_T = [0,T]$ in the c.t. case and $D_T = \{1,\dots,T\}$ in the d.t. case.

The process $X(t)$ is what we believe is being observed but in reality the data are in the contaminated form $Y(t)$. In this case standard inferences can be carried on the basis of the stationary model $X(t)$, and we are interested in question whether the conclusions are robust against this kind of departure from the stationarity.

In the non-tapered case, this problem for d.t. models was considered in Heyde and Dai [42] (see also Taniguchi and Kakizawa [50], Theorems 6.4.1 and 6.4.2). For c.t. models it was studied in Ginovyan and Sahakyan [33].

The results stated below show that if the trend $M(t)$ is 'small', then the asymptotic properties of estimators of the parameter θ and the functional $J(f)$, stated in Sections 3 and 4 for a stationary model $X(t)$, remain valid for the contaminated model $Y(t)$, that is, both the parametric and nonparametric estimating procedures are robust against replacing the stationary model $X(t)$ by the non-stationary $Y(t)$. To this end, similar to the non-tapered case, we first establish an asymptotic relation between stationary and contaminated tapered periodograms. For simplicity, the results that follow we prove in the c.t. case, the proofs in the d.t. case are similar.

5.1. **A relation between stationary and contaminated tapered periodograms.** The next result shows that a small trend of the form $|M(t)| \le C|t|^{-\beta}$, $\beta > 1/4$, does not effect the asymptotic properties of the empirical spectral linear functionals of a tapered periodogram. Note that this result is of general nature, and do not require from the model $X(t)$ to be linear.

Theorem 5.1. *Let $\{X(t),\, t \in \mathbb{U}\}$ be a stationary mean zero process, $\{M(t),\, t \in \mathbb{U}\}$ be a deterministic trend, $Y(t) = X(t) + M(t)$, and let $I^h_{TX}(\lambda)$ and $I^h_{TY}(\lambda)$ be the tapered periodograms of $X(t)$ and $Y(t)$, respectively. Let $g(\lambda)$, $\lambda \in \Lambda$ be an even integrable function. If the trend $M(t)$ and the Fourier transform $a(t) := \widehat{g}(t)$ of $g(\lambda)$ are such that $M(t)$ is locally integrable on $\mathbb{R}$ and*

$$|M(t)| \le C|t|^{-\beta}, \qquad |a(t)| \le C|t|^{-\gamma}, \quad t \in \Lambda, \quad 2\beta + \gamma > 3/2, \tag{5.2}$$

with some constants $C > 0$, $\gamma > 0$ and $\beta > 1/4$, then

$$T^{1/2} \int_\Lambda g(\lambda) \left[I^h_{TY}(\lambda) - I^h_{TX}(\lambda) \right] d\lambda \xrightarrow{P} 0 \quad \text{as} \quad T \to \infty, \tag{5.3}$$

where $\xrightarrow{P}$ stands for convergence in probability, provided that one of the following conditions holds:

(i) *the process $X(t)$ has short or intermediate memory, that is, the covariance function $r(t) := r_X(t)$ of $X(t)$ satisfies $r \in L^1(\Lambda)$, and $\beta + \gamma > 1$,*

(ii) *the process $X(t)$ has long memory with covariance function $r(t)$ satisfying*

$$|r(t)| \le C|t|^{-\alpha}, \quad t \in \Lambda, \quad \alpha + \gamma \ge 3/2 \tag{5.4}$$

with some constants $C > 0$, $0 < \alpha \le 1$, and $\alpha + 2\beta > 1$ if $\beta < 1 < \gamma$.

Proof. In view of (2.18) and (5.1) we can write

$$
\begin{aligned}
& I^h_{T,X}(\lambda) - I^h_{T,Y}(\lambda) = \\
= \;& \frac{1}{C_T}\left(\left|\int_0^T e^{-i\lambda t} h_T(t) X(t)\, dt\right|^2 - \left|\int_0^T e^{-i\lambda t} h_T(t) Y(t)\, dt\right|^2\right) \\
= \;& \frac{1}{C_T}\left(\left|\int_0^T e^{-i\lambda t} h_T(t) [Y(t)+M(t)]\, dt\right|^2 - \left|\int_0^T e^{-i\lambda t} h_T(t) Y(t)\, dt\right|^2\right) \\
= \;& \frac{1}{C_T}\int_0^T\int_0^T e^{i\lambda(t-s)} h_T(t) h_T(s)\left[Y(t)M(s) + Y(s)M(t) + M(t)M(s)\right] dtds
\end{aligned}
$$

and

$$
\begin{aligned}
& \int_{-\infty}^{+\infty} g(\lambda,\theta)\left[I^h_{T,X}(\lambda) - I^h_{T,Y}(\lambda)\right] d\lambda \\
& = \frac{1}{H_2 T}\int_0^T\int_0^T [Y(t)M(s) + Y(s)M(t) + M(t)M(s)]\, h_T(t) h_T(s) a(t-s)\, dtds \\
(5.5) \le\; & \frac{C}{T}\int_0^T\int_0^T |Y(t)M(s) + Y(s)M(t) + M(t)M(s)|\, |a(t-s)|\, dtds,
\end{aligned}
$$

since the function h is bounded on $\mathbb{R}$ by Assumption 2.1.

Thus, to complete the proof it is enough to observe that under the conditions of the theorem we have (see Ginovyan and Sahakyan [33], relations (6.11) and (6.12)):

$$T^{-1/2}\int_0^T\int_0^T M(t)|M(s)a(t-s)|\, dtds \to 0 \quad \text{as} \quad T\to\infty$$

and

$$T^{-1/2}\int_0^T\int_0^T |Y(t)M(s)a(t-s)|\, dtds \xrightarrow{P} 0 \quad \text{as} \quad T\to\infty. \qquad \square$$

Remark 5.1. It is easy to check that the statement of Theorem 5.1 holds, in particular, if the parameters α, β and γ satisfy the following conditions:

in the case *(i)*: $\beta > 1/2,\ \gamma \ge 1/2$,

in the case *(ii)*: $\alpha \ge 3/4,\ \beta > 3/8,\ \gamma \ge 3/4$.

Remark 5.2. In the non-tapered d.t. case, Theorem 5.1 (with additional conditions $\gamma = 1$ in the case *(i)*, and $\gamma > 1,\ \alpha < 1/2$ in the case *(ii)*), was proved by Heyde and Dai [42] (see also Taniguchi and Kakizawa [50], Theorems 6.4.1 and 6.4.2).

5.2. **Robustness to small trends of nonparametric estimation.** The next result shows that a small trend of the form $|M(t)| \le C|t|^{-\beta}$ does not effect the asymptotic properties of the estimator of a linear spectral functional $J(f)$, that is, the nonparametric estimation procedure is robust to the presence of a small trend in the model.

Theorem 5.2. *Suppose that the assumptions of Theorems 3.6 and 5.1 are fulfilled. Then the statistic $J(I^h_{TY})$ is consistent and asymptotically normal estimator for functional $J(f)$ with asymptotic variance $\sigma^2_h(J)$ given by* (3.10) *and* (3.11), *that is, the asymptotic relation (3.14) is satisfied with $I^h_{TX}(\lambda)$ replaced by the contaminated periodogram $I^h_{TY}(\lambda)$:*

$$T^{1/2}\left[J(I^h_{TY}) - J(f)\right] \xrightarrow{d} \eta \quad \text{as} \quad T \to \infty, \tag{5.6}$$

where η is $N(0, \sigma^2_h(J))$ with $\sigma^2_h(J)$ given by (3.10) *and* (3.11).

Proof of Theorem 5.2. In view of (3.1) and (3.2) we can write

$$\begin{aligned} T^{1/2}\left[J(I^h_{TY}) - J(f)\right] &= T^{1/2}\left[\int_{\mathbb{R}} I^h_{TY}(\lambda)g(\lambda)d\lambda - \int_{\mathbb{R}} f(\lambda)g(\lambda)d\lambda\right] \\ &= T^{1/2}\left[\int_{\mathbb{R}} I^h_{TY}(\lambda)g(\lambda)d\lambda - \int_{\mathbb{R}} I^h_{TX}(\lambda)g(\lambda)d\lambda\right] \\ &+T^{1/2}\left[\int_{\mathbb{R}} I^h_{TX}(\lambda)g(\lambda)d\lambda - \int_{\mathbb{R}} f(\lambda)g(\lambda)d\lambda\right] \\ &= T^{1/2}\int_{\mathbb{R}} g(\lambda)\left[I^h_{TY}(\lambda) - I^h_{TX}(\lambda)\right]d\lambda + T^{1/2}\left[J(I^h_{TX}) - J(f)\right]. \end{aligned} \tag{5.7}$$

By Theorem 5.1, the first term on the right-hand side of (5.7) goes to zero in probability as $T \to \infty$, while by Theorem 3.6, the second term on the right-hand side of (5.7) goes in distribution to η, and the result follows. □

5.3. **Robustness to small trends of parametric estimation.** The next result shows that a small trend of the form $|M(t)| \le C|t|^{-\beta}$, $\beta > 1/4$, does not effect the asymptotic properties of the Whittle estimator of an unknown spectral parameter θ, that is, the Whittle parametric estimation procedure based on the tapered sample (2.13) is robust to the presence of a small trend in the model.

Theorem 5.3. *Suppose that the assumptions of Theorem 5.1 with $g = f^{-1}(\lambda, \theta) \cdot w(\lambda)$ are satisfied. Then under the conditions of Theorems 4.2 the Whittle estimator $\widehat{\theta}_{TY,h}$, constructed on the basis of the contaminated tapered periodogram $I^h_{T,Y}(\lambda)$, is consistent and asymptotically normal estimator for an unknown spectral parameter θ, that is, the asymptotic relation (4.8) is satisfied with $I^h_{TX}(\lambda)$ replaced by the contaminated periodogram $I^h_{TY}(\lambda)$:*

$$T^{1/2}\left(\widehat{\theta}_{TY,h} - \theta_0\right) \xrightarrow{d} N_p\left(0, e(h)\Gamma(\theta_0)\right) \quad \text{as} \quad T \to \infty, \tag{5.8}$$

where the matrix $\Gamma(\theta_0)$ is defined in (4.9).

Proof of Theorem 5.3. By Taylor's formula for $\frac{\partial}{\partial\theta}U^h_{TX}\left(\widehat{\theta}_{TX,h}\right)$, where $U^h_{TX}(\cdot)$ is the tapered Whittle functional defined by (4.2) and $\widehat{\theta}_{TX,h}$ is the Whittle estimator constructed on the basis of observation $\mathbf{X}_T = \{X(t),\, 0 \le t \le T\}$, for $|\widehat{\theta}^*_T - \theta_0| < |\widehat{\theta}_{TX,h} - \theta_0|$ and for sufficiently large T, we can write

$$(5.9) \quad T^{1/2}\left[\widehat{\theta}_{TX,h} - \theta_0\right] = -T^{1/2}\left[\frac{\partial^2}{\partial\theta\partial\theta'}U^h_{TX}(\theta^*_T)\right]^{-1}\left[\frac{\partial}{\partial\theta}U^h_{TX}(\theta_0)\right] + o_P(1).$$

Next, by Theorem 5.1, we have

$$(5.10) \qquad U^h_{TY}(\theta_T) = U^h_{TX}(\theta_T) + o_P(1).$$

Again using Taylor's formula for $\frac{\partial}{\partial\theta}U^h_{TY}\left(\widehat{\theta}_{TY,h}\right)$, where now $U^h_{TY}(\cdot)$ and $\widehat{\theta}_{TY,h}$ are respectively the Whittle functional and the Whittle estimator, constructed on the basis of the contaminated observation $\mathbf{Y}_T = \{Y(t),\, 0 \le t \le T\}$, and taking into account the relations (5.9) and (5.10), we can infer that

$$T^{1/2}\left[\widehat{\theta}_{TY,h} - \theta_0\right] = T^{1/2}\left[\widehat{\theta}_{TX,h} - \theta_0\right] + o_P(1),$$

showing that the estimator $\widehat{\theta}_{TY,h}$ possesses the same asymptotic properties as $\widehat{\theta}_{TX,h}$.

Hence the result follows from Theorems 4.2. $\square$

Remark 5.3. In the non-tapered case, Theorems 5.1 – 5.3 were proved in Ginovyan and Sahakyan [33].

6. Methods and tools

In this section we briefly discuss the methods and tools, used to prove the results stated in Sections 3–5.

6.1. **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. [36] 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 3–5.

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 $T = 1, 2, \ldots$, the $(T \times T)$-*truncated tapered Toeplitz matrix* generated by ψ and h, denoted by $B^h_T(\psi)$, is defined by the following equation:

$$(6.1) \qquad B^h_T(\psi) := \|\widehat{\psi}(t-s)h_T(t)h_T(s)\|_{t,s=1,2\ldots,T},$$

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 $W^h_T(\psi)$ is defined as follows:

$$(6.2) \qquad [W^h_T(\psi)u](t) = \int_0^T \hat{\psi}(t-s)u(s)h_T(s)ds, \quad u(s) \in L^2([0,T]; h_T),$$

where $\hat{\psi}(\cdot)$ 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 2.1, 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 (6.1) and (6.2)).

Observe that, in view of (2.15), (2.19), (6.1) and (6.2), we have

$$\frac{1}{T}\mathrm{tr}\left[A_T^h(\psi)\right] = \frac{1}{T}\cdot\widehat{\psi}(0)\cdot\int_0^T h_T^2(t)dt = 2\pi H_2\int_\Lambda \psi(\lambda)d\lambda. \tag{6.3}$$

What happens to the relation (6.3) 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

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

where H_m is as in (2.15), and let

$$\Delta(T) := \Delta_{A,\Lambda,\mathcal{H},h}(T) = |S_{A,\mathcal{H},h}(T) - M_{\Lambda,\mathcal{H},h}|. \tag{6.4}$$

Proposition 6.1. *Let $\Delta(T)$ be as in (6.4). Each of the following conditions is sufficient for*

$$\Delta(T) = o(1) \quad \text{as} \quad T\to\infty. \tag{6.5}$$

- **(C1)** $\psi_i\in L^1(\Lambda)\cap L^{p_i}(\Lambda)$, $p_i>1$, $i=1,2,\dots,m$, *with* $1/p_1+\cdots+1/p_m\le 1$.
- **(C2)** *The function* $\varphi(\mathbf{u})$ *defined by*

$$\varphi(\mathbf{u}) := \int_\Lambda \psi_1(\lambda)\psi_2(\lambda-u_1)\psi_3(\lambda-u_2)\cdots\psi_m(\lambda-u_{m-1})\,d\lambda, \tag{6.6}$$

  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 6.1. In the non-tapered case, Proposition 6.1 was proved in Ginovyan et al. [36], while in the tapered case, it was proved in Ginovyan [28]. Proposition 6.1 was used to prove Theorems 3.5, 3.6, and 4.2. More results concerning the trace approximation problem for truncated Toeplitz operators and matrices can be found in Ginovyan and Sahakyan [31, 32], and in Ginovyan et al. [36].

6.2. **Central limit theorems for tapered quadratic functionals.** In this subsection we state central limit theorems for tapered quadratic functional Q_T^h given by (3.3), which were used to prove the results stated in Sections 3–5.

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 (6.1) and (6.2)). Similar to the non-tapered case, we have the following results (cf. Ginovyan et al. [36], Ibragimov [43]).

1. The quadratic functional Q_T^h in (3.3) has the same distribution as the sum $\sum_{j=1}^{\infty}\lambda_{j,T}\xi_j^2$, where $\{\xi_j, j\ge 1\}$ are independent $N(0,1)$ Gaussian random variables and $\{\lambda_{j,T}, j\ge 1\}$ 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: $\varphi(t)=\prod_{j=1}^{\infty}|1-2it\lambda_{j,T}|^{-1/2}$.
3. The k–th order cumulant $\chi_k(Q_T^h)$ of Q_T^h is given by formula:

$$\chi_k(Q_T^h)=2^{k-1}(k-1)!\sum_{j=1}^{\infty}\lambda_{j,T}^k=2^{k-1}(k-1)!\,\mathrm{tr}\,[A_T^h(f)\,A_T^h(g)]^k. \tag{6.7}$$

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.

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

$$\widetilde{Q}_T^h=T^{-1/2}\left(Q_T^h-\mathbb{E}[Q_T^h]\right). \tag{6.8}$$

Also, we set

$$\sigma_h^2:=16\pi^3H_4\int_\Lambda f^2(\lambda)g^2(\lambda)\,d\lambda, \tag{6.9}$$

where H_4 is as in (2.15). The notation

$$\widetilde{Q}_T^h\xrightarrow{d}\eta\sim N(0,\sigma_h^2)\quad\text{as}\quad T\to\infty \tag{6.10}$$

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

The following theorems were proved in Ginovyan and Sahakyan [35].

Theorem 6.1. *Each of the following conditions is sufficient for the quadratic form Q_T^h to obey the CLT, that is, for $\widetilde{Q}_T^h\xrightarrow{d}\eta\sim N(0,\sigma_h^2)$ as $T\to\infty$ with σ_h^2 is as in (6.9).*

(T1) *$f \cdot g \in L^1(\Lambda) \cap L^2(\Lambda)$, the taper function h satisfies Assumption 2.1, and for $T \to \infty$*

$$\chi_2(\widetilde{Q}_T^h) = \frac{2}{T} tr\big[A_T^h(f) A_T^h(g)\big]^2 \longrightarrow \sigma_h^2. \tag{6.11}$$

(T2) *The function*

$$\varphi(x_1, x_2, x_3) = \int_\Lambda f(u) g(u - x_1) f(u - x_2) g(u - x_3)\, du \tag{6.12}$$

belongs to $L^2(\Lambda^3)$ and is continuous at $(0,0,0)$, and the taper function h satisfies Assumption 2.1.

(T3) *$f(\lambda) \in L^p(\Lambda)$ $(p \geq 1)$ 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 2.1.*

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 $[-a, a]$, $\lim_{\lambda \to 0} u(\lambda) = 0$, $u(\lambda) = u(-\lambda)$ and $0 < u(\lambda) < u(\mu)$ for $0 < \lambda < \mu < a$.

Theorem 6.2. *Assume that the functions f and g are integrable on $\mathbb{R}$ and bounded outside any neighborhood of the origin, and satisfy for some $a > 0$*

$$f(\lambda) \leq |\lambda|^{-\alpha} L_1(\lambda), \quad |g(\lambda)| \leq |\lambda|^{-\beta} L_2(\lambda), \quad \lambda \in [-a, a], \tag{6.13}$$

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

$$L_i \in SV_0(\mathbb{R}), \quad \lambda^{-(\alpha+\beta)} L_i(\lambda) \in L^2[-a, a], \; i = 1, 2. \tag{6.14}$$

Also, let the taper function h satisfy Assumption 2.1. Then $\widetilde{Q}_T^h \xrightarrow{d} \eta \sim N(0, \sigma_h^2)$ as $T \to \infty$.

The condition $\alpha < 1$, $\beta < 1$ in Theorem 6.2 ensure that the Fourier transforms of f and g are well defined. For $\alpha > 0$ 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 (6.14) imply that $f \cdot g \in L^2(\Lambda)$, so that the variance σ_h^2 in (6.9) is finite.

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 (2.5), 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 [34].

Theorem 6.3. *Assume that the filter $a(\cdot)$ and the generating kernel $\widehat{g}(\cdot)$ are such that*

$$a(\cdot) \in L^p(\mathbb{R}) \cap L^2(\mathbb{R}), \quad \widehat{g}(\cdot) \in L^q(\mathbb{R}), \quad 1 \le p, q \le 2, \quad 2/p + 1/q \ge 5/2, \tag{6.15}$$

and the taper h satisfies Assumption 2.1. Then $\widetilde{Q}_T^h \xrightarrow{d} \eta \sim N(0, \sigma^2_{L,h})$ as $T \to \infty$, where

$$\sigma^2_{L,h} = 16\pi^3 H_4 \int_{\mathbb{R}} f^2(\lambda) g^2(\lambda) d\lambda + \kappa_4 4\pi^2 H_4 \left[\int_{\mathbb{R}} f(\lambda) g(\lambda) d\lambda \right]^2, \tag{6.16}$$

where H_4 is as in (2.15)*.*

Remark 6.2. Notice that if the underlying process $X(t)$ is Gaussian, then in formula (6.16) we have only the first term and so $\sigma^2_{L,h} = \sigma^2_h$ (see (6.9)), because in this case $\kappa_4 = 0$. On the other hand, the condition (6.15) is more restrictive than the conditions in Theorems 6.1 and 6.2. Thus, for Gaussian processes Theorems 6.1 and 6.2 improve Theorem 6.3. For non-tapered case Theorem 6.3 was proved in Bai et al. [4].

6.3. **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 ($k = 2, 3, \ldots$) and a taper function h satisfying Assumption 2.1 consider the following Fejér-type tapered kernel function:

$$F^h_{k,T}(\mathbf{u}) := \frac{H_T(\mathbf{u})}{(2\pi)^{k-1} H_{k,T}(0)}, \quad \mathbf{u} = (u_1, \ldots, u_{k-1}) \in \mathbb{R}^{k-1}, \tag{6.17}$$

where

$$H_T(\mathbf{u}) := H_{1,T}(u_1) \cdots H_{1,T}(u_{k-1}) H_{1,T}\left(-\sum_{j=1}^{k-1} u_j \right), \tag{6.18}$$

and the function $H_{k,T}(\cdot)$ is defined by (2.16) with $H_{k,T}(0) = T \cdot H_k \neq 0$ (see (2.15)).

The next result shows that, similar to the classical Fejér kernel, the tapered kernel $F^h_{k,T}(\mathbf{u})$ is an approximation identity (see Ginovyan and Sahakyan [34], Lemma 3.4).

Proposition 6.2. *For any $k = 2, 3, \ldots$ and a taper function h satisfying Assumption 2.1 the kernel $F^h_{k,T}(\mathbf{u})$, $\mathbf{u} = (u_1, \ldots, u_{k-1}) \in \mathbb{R}^{k-1}$, possesses the following properties:*

a) $\sup_{T>0} \int_{\mathbb{R}^{k-1}} \left| F^h_{k,T}(\mathbf{u}) \right| d\mathbf{u} = C_1 < \infty$;

b) $\int_{\mathbb{R}^{k-1}} F^h_{k,T}(\mathbf{u})\, d\mathbf{u} = 1$;

c) $\lim_{T\to\infty} \int_{\mathbb{E}^c_\delta} \left| F^h_{k,T}(\mathbf{u}) \right| d\mathbf{u} = 0$ *for any* $\delta > 0$;

d) *If $k > 2$ for any $\delta > 0$ there exists a constant $M_\delta > 0$ such that $\left\| F^h_{k,T} \right\|_{L^{p_k}(\mathbb{E}^c_\delta)} \le M_\delta$ for $T > 0$, where $p_k = \frac{k-2}{k-3}$ for $k > 3$, $p_3 = \infty$, $\mathbb{E}^c_\delta = \mathbb{R}^{k-1} \setminus \mathbb{E}_\delta$, and*

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

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

$$\lim_{T\to\infty} \int_{\mathbb{R}^{k-1}} Q(\mathbf{u}+\mathbf{v}) F^h_{k,T}(\mathbf{u}) d\mathbf{u} = Q(\mathbf{v}). \tag{6.19}$$

Denote

$$\Delta^h_{2,T} := \int_{\mathbb{R}^2} f(\lambda) g(\lambda+\mu) F^h_{2,T}(\mu) d\lambda d\mu - \int_{\mathbb{R}} f(\lambda) g(\lambda) d\lambda, \tag{6.20}$$

where $F^h_{2,T}(\mu)$ is given by (6.17) and (6.18).

The next two propositions give information on the rate of convergence to zero of $\Delta^h_{2,T}$ as $T \to \infty$ (see Ginovyan and Sahakyan [34], Lemmas 4.1 and 4.2).

Proposition 6.3. *Assume that Assumptions 2.1 and 3.3 are satisfied. Then the following asymptotic relation holds:*

$$\Delta^h_{2,T} = o\left(T^{-1/2}\right) \quad \text{as} \quad T \to \infty. \tag{6.21}$$

Proposition 6.4. *Assume that Assumptions 2.1 and 3.2 are satisfied. Then the following inequality holds:*

$$|\Delta^h_{2,T}| \le C_h \begin{cases} T^{-(\beta_1+\beta_2)}, & \textit{if} \ \ \beta_1+\beta_2<1 \\ T^{-1}\ln T, & \textit{if} \ \ \beta_1+\beta_2=1 \\ T^{-1}, & \textit{if} \ \ \beta_1+\beta_2>1, \end{cases} \qquad T>0, \tag{6.22}$$

where C_h *is a constant depending on* h.

Notice that for non-tapered case ($h(t) = \mathbb{I}_{[0,1]}(t)$), Propositions 6.3 and 6.4 were proved in Ginovyan and Sahakyan [30] (see also Ginovyan and Sahakyan [31, 32]). In the d.t. tapered case, Proposition 6.3 under different conditions was proved in Dahlhaus [10].

Список литературы

[1] V. V. Anh, J. M. Angulo and M. D. Ruiz-Medina, "Possible long-range dependence in fractional random fields", J. Statist. Plann. Inference, **80**, 95 – 110 (1999).

[2] V. V. Anh, N. N. Leonenko and R. McVinish, "Models for fractional Riesz-Bessel motion and related processes", Fractals, **9**, 329 – 346 (2001)..

[3] F. Avram, N. N. Leonenko and L. Sakhno, "On a Szegö 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 (2010).

[4] S. Bai, M. S. Ginovyan and M. S. Taqqu, "Limit theorems for quadratic forms of Levy-driven continuous-time linear processes", Stochast. Process. Appl., **126**, 1036 – 1065 (2016).

[5] J. Beran, Y. Feng, S. Ghosh and R. Kulik, Long-Memory Processes: Probabilistic Properties and Statistical Methods, Springer-Verlag, Berlin (2013).

[6] D. R. Brillinger, Time Series: Data Analysis and Theory, Holden Day, San Francisco (1981).

[7] P. J. Brockwell and R. A. Davis, Time Series: Theory and Methods, 2nd ed., Springer-Verlag, New York (1991).

[8] I. Casas and J. Gao, "Econometric estimation in long-range dependent volatility models: Theory and practice", Journal of Econometrics, **147**, 72 – 83 (2008).

[9] H. Cramér, M. R. Leadbetter, Stationary and Related Stochastic Processes. Sample Function Properties and Their Applications. John Wiley & Sons, New York (1967).

[10] R. Dahlhaus, "Spectral analysis with tapered data", J. Time Ser. Anal., **4**, 163 – 174 (1983).

[11] R. Dahlhaus, "Parameter estimation of stationary processes with spectra containing strong peaks", Robust and Nonlinear Time Series Analysis (Franke, Hardle and Martin, eds.) Lecture Notes in Statistics, no. 26, 50 – 67 (1984).

[12] R. Dahlhaus, "Efficient parameter estimation for self-similar processes", Ann. Statist., **17**, 1749 – 1766 (1989).

[13] R. Dahlhaus and H. Künsch, "Edge effects and efficient parameter estimation for stationary random fields", Biometrika, **74**(4), 877 – 882 (1987).

[14] R. Dahlhaus and W. Wefelmeyer, "Asymptotically optimal estimation in misspecified time series models", Ann. Statist., **24**, 952 – 974 (1996).

[15] K. Dzhaparidze, Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series, Springer, New York (1986).

[16] K. O. Dzaparidze and A. M. Yaglom, Spectrum Parameter Estimation in Time Series Analysis, Developments in Statistics (P. R. Krishnaiah, ed.), Academic Press, New York, **4**, 1 – 181, (1983).

[17] R. Fox and M. S. Taqqu, "Large-sample properties of parameter estimation for strongly dependent stationary Gaussian time series", Ann. Statist., **14**, 517 – 532 (1986).

[18] J. Gao, "Modelling long-range dependent Gaussian processes with application in continuous-time financial models", J. Appl.Probab., **41**, 467 – 482 (2004).

[19] J. Gao, V. V. Anh, C. Heyde Q. Tieng, "Parameter estimation of stochastic processes with long-range dependence and intermittency", J. Time Ser. Anal., **22**, 517 – 535 (2001).

[20] M. S. Ginovyan, "Asymptotically efficient nonparametric estimation of functionals on spectral density with zeros", Theory Probab. Appl., **33**, 315 – 322 (1988).

[21] M. S. Ginovyan, "On estimate of the value of the linear functional in a spectral density of stationary Gaussian process", Theory Probab. Appl., **33**, 777 – 781 (1988).

[22] M. S. Ginovyan, "On Toeplitz type quadratic functionals in Gaussian stationary process", Probab. Theory Relat. Fields, **100**, 395 – 406 (1994).

[23] M. S. Ginovyan, "Asymptotic properties of spectrum estimate of stationary Gaussian process", J. Cont. Math. Anal., **30**, 1 – 17 (1995).

[24] M. S. Ginovyan, "Asymptotically efficient nonparametric estimation of nonlinear spectral functionals", Acta Appl. Math., **78**, 145 – 154 (2003).

[25] M. S. Ginovyan, "Efficient estimation of spectral functionals for continuous-time stationary models", Acta Appl. Math., **115**(2), 233 – 254 (2011).

[26] M. S. Ginovyan, "Efficient estimation of spectral functionals for Gaussian stationary models", Comm. Stochast. Anal., **5**(1), 211 – 232 (2011).

[27] M. S. Ginovyan, "Parameter estimation for Lévy-driven continuous-time linear models with tapered data", Acta Appl Math., **169**, 79 – 97 (2020).

[28] M. S. Ginovyan, "Goodness-of-fit tests for stationary Gaussian processes with tapered data", Acta Appl. Math., **171** (1), 1 – 12 (2021).

[29] M. S. Ginovyan and A. A. Sahakyan, "On the Central limit theorem for Toeplitz quadratic forms of stationary sequences, Theory Probab. and Appl., **49**, 612 – 628 (2005).

[30] M. S. Ginovyan and A. A. Sahakyan, "Limit theorems for Toeplitz quadratic functionals of c.t. stationary process", Probab. Theory Relat. Fields, **138**, 551 – 579 (2007).

[31] M. S. Ginovyan and A. A. Sahakyan, "Trace approximations of products of truncated Toeplitz operators", Theory Probab. Appl., **56**(1), 57 – 71 (2012).

[32] M. S. Ginovyan and A. A. Sahakyan, "On the trace approximations of products of Toeplitz matrices", Statist. Probab. Lett., **83**(3), 753 – 760 (2013).

[33] M. S. Ginovyan and A. A. Shakyan, "Robust estimation for continuous-time linear models with memory", Theory Probability and Mathematical Statistics", **95**, 81 – 98 (2018).

[34] M. S. Ginovyan and A. A. Sahakyan, "Estimation of spectral functionals for Levy-driven continuous-time linear models with tapered data", Electronic Journal of Statistics", **13**, 255 – 283 (2019).

[35] M. S. Ginovyan and A. A. Sahakyan, "Limit theorems for tapered Toeplitz quadratic functionals of continuous-time Gaussian stationary processes", *J. Cont. Math. Anal.*, **54**(4), 222 – 239 (2019).

[36] M. S. Ginovyan, Sahakyan, A. A. and M. S. Taqqu, "The trace problem for Toeplitz matrices and operators and its impact in probability", Probability Surveys, **11**, 393 – 440 (2014).

[37] L. Giraitis, H. Koul and D. Surgailis, Large Sample Inference for Long Memory Processes, Imperial College Press, London (2012).
[38] L. Giraitis and D. Surgailis, "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 (1990).
[39] D. Guégan, "How can we define the concept of long memory? An econometric survey", Econometric Reviews, **24**, 113 – 149 (2005).
[40] X. Guyon, Random Fields on a Network: Modelling, Statistics and Applications, Springer, New York (1995).
[41] R. Z. Hasminskii and I. A. Ibragimov, "Asymptotically efficient nonparametric estimation of functionals of a spectral density function", Probab. Theory Related Fields., **73**, 447 – 461 (1986).
[42] C. C. Heyde and W. Dai, "On the robustness to small trends of estimation based on the smoothed periodogram", J. Time Ser. Anal., **17**(2), 141 – 150 (1996).
[43] I. A. Ibragimov, "On estimation of the spectral function of a stationary Gaussian process", Theory Probab. Appl., **8**, 366 – 401 (1963).
[44] A. A. Ibragimov, "On maximum likelihood estimation of parameters of the spectrum of stationary time series", Theory Probab. Appl., **12**, 115 – 119 (1967).
[45] I. A. Ibragimov and R. Z. Khasminskii, "Asymptotically normal families of distributions and efficient estimation", Ann. Statist., **19**, 1681 – 1724 (1991).
[46] I. A. Ibragimov and Yu. V. Linnik, Independent and Stationary Sequences of Random Variables, Wolters-Noordhoff Publishing Groningen, The Netherlands (1971).
[47] N. N. Leonenko and L. M. Sakhno, "On the Whittle estimators for some classes of continuous-parameter random processes and fields", Stat. Probab. Lett., **76**, 781 – 795 (2006).
[48] C. Ludeña and M. Lavielle, The Whittle Estimator for Strongly Dependent Stationary Gaussian Fields, Scand. J. Stat., **26**, 433 – 450 (1999).
[49] M. Taniguchi, "Minimum contrast estimation for spectral densities of stationary processes", J. R. Stat. Soc. Ser. B-Stat. Methodol., **49**, 315 – 325 (1987).
[50] M. Taniguchi and Y. Kakizawa, Asymptotic Theory of Statistical Inference for Time Series, Academic Press, New York (2000).
[51] J. W. Tukey, "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 (1967).
[52] H. Tsai and K. S. Chan, "Quasi-maximum likelihood estimation for a class of continuous-time long memory processes", J. Time Ser. Anal., **26**(5), 691 – 713 (2005).
[53] A. M. Walker, "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 (1964).
[54] P. Whittle, Hypothesis Testing in Time Series, Hafner, New York (1951).
[55] A. M. Yaglom, The Correlation Theory of Stationary and Related Random Processes, Vol. 1., Springer, New York (1986).

Поступила 03 мая 2021

После доработки 03 мая 2020

Принята к публикации 30 июля 2021