*Известия НАН Армении, Математика, том 59, н. 1, 2024, стр. 20 – 31.*

# SIMPLIFIED WHITTLE ESTIMATORS FOR SPECTRAL PARAMETERS OF STATIONARY LINEAR MODELS WITH TAPERED DATA

M. S. GINOVYAN

*Boston University, Boston, MA, USA*
E-mail: *ginovyan@math.bu.edu*

Abstract. The paper is concerned with the statistical estimation of the spectral parameters of stationary models with tapered data. As estimators of the unknown parameters we consider the tapered Whittle estimator and the simplified tapered Whittle estimators. We show that under broad regularity conditions on the spectral density of the model these estimators are asymptotically statistically equivalent, in the sense that these estimators possess the same asymptotic properties. The processes considered will be discrete-time and continuous-time Gaussian, linear or Lévy-driven linear processes with memory.

**MSC2020 numbers:** 62F10; 62F12; 60G10; 60G12.

**Keywords:** tapered data; stationary processes; spectral density; parametric estimation; simplified Whittle estimator.

## 1. Introduction

The present paper is concerned with the following general parametric estimation problem. Let $\theta := (\theta_1, \ldots, \theta_p) \in \Theta \subset \mathbb{R}^p$ be an unknown vector parameter appearing (a) in the probability density of some random variable $X$, or (b) in the finite-dimensional probability densities of a random process $\{X(t),\ t \in \mathbb{U}\}$, where $\mathbb{U} = \mathbb{R}$ in the continuous-time (c.t.) case and $\mathbb{U} = \mathbb{Z}$ in the discrete-time (d.t.) case. The problem of interest is to estimate the value of the parameter $\theta$ based on the sample $\mathbf{X}_T$, where in case (a) $\mathbf{X}_T := \{X_1, \ldots, X_T\}$, $X_1, \ldots, X_T$ being $T$ independent observations of the random variable $X$, and in case (b) $\mathbf{X}_T$ is an observed finite realization 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. The usual methods of constructing estimators of the unknown parameter $\theta$ used in mathematical statistics (for example, the method of moments, the maximum likelihood method, the least-squares method, the Whittle method, etc.), as a rule, require finding the roots of some system of (possibly non-linear) estimating equations with respect to the unknown $\theta = (\theta_1, \ldots, \theta_p)$ of the form:

$$F_i(\mathbf{X}_T, \theta) = 0, \quad i = 1, \ldots p, \tag{1.1}$$

where $F_i(\theta) := F_i(\mathbf{X}_T, \theta)$ are certain functionals of $\mathbf{X}_T$ depending on $\theta$.

The classical estimation methods often lead to estimators with good asymptotic properties. For example, in many cases one can prove that for sufficiently large values $T$ there exist, with probability near 1, a root $\hat{\theta}_T$ of the system of estimating equations (1.1) which is a consistent estimator of $\theta$, that is, $p - lim_{T\to\infty}\hat{\theta}_T = \theta_0$, where $p - lim$ denotes the limit in probability, and $\theta_0 \in \Theta$ is the unknown true value of the parameter $\theta$. Moreover, under broad regularity conditions, the classical estimation methods often lead to $\tau_T$-consistent and asymptotically normal estimators, where $\tau_T$ is a comparatively rapidly increasing function. (Recall that for a non-random function $\tau_T = \tau(T)$ increasing without bound as $T \to \infty$, we say that the statistic $\hat{\theta}_T$ is a $\tau_T$-consistent estimator for $\theta$ if the distribution of the random vector $\tau_T(\hat{\theta}_T - \theta_0)$ converges (as $T \to \infty$) to a non-degenerate distribution.

These classical estimation methods, however, have two disadvantages. First, it is only for relatively simple situations that the system of estimating equations (1.1) has an explicit solution, and finding the roots of the system (1.1) often turns out to be very hard problem. Second, for the roots to be consistent, the estimating equations need to behave well throughout the parameter set. Another issue that arise in the statistical analysis of stationary models is that 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*.

Therefore it is of considerable interest to find more easily constructed (simplified) estimators $\check{\theta}_T$ that are asymptotically statistically equivalent to $\hat{\theta}_T$, that is, having the same asymptotic (as $T \to \infty$) properties as the estimator $\hat{\theta}_T$. The problem of constructing simplified estimators with good asymptotic properties based on the standard (non-tapered) data $\mathbf{X}_T$ goes back to the classical work of Le Cam [16], and then it was developed by Dzhaparidze [8, 9] (see also Beinicke and Dzhaparidze [1] and Dzhaparidze [10]).

In this paper we focus on the Whittle estimation method of the spectral parameters of stationary models with tapered data. We provide sufficient conditions for the tapered Whittle estimator to be $\sqrt{T}$-consistent and asymptotically normal. Then we construct simplified Whittle estimators based on the tapered data, and show that under broad regularity conditions on the spectral density of the model the Whittle estimator and the simplified Whittle estimator are asymptotically statistically equivalent, in the sense that these estimators possess the same asymptotic properties.

The processes considered will be discrete-time and continuous-time Gaussian, linear or Lévy-driven linear processes with memory.

## 2. The model

We will consider here stationary processes possessing spectral density functions, and will distinguish the following three models.

(a) Discrete-time linear model. The process $\{X(t),\, t \in \mathbb{Z}\}$ is a discrete-time linear process of the form:

$$X(t) = \sum_{k=-\infty}^{\infty} a(t-k)\xi(k), \qquad \sum_{k=-\infty}^{\infty} |a(k)|^2 < \infty, \tag{2.1}$$

where $\{\xi(k), k \in \mathbb{Z}\} \sim$ WN(0,1) is a standard white-noise, that is, a sequence of orthonormal random variables. The spectral density $f(\lambda)$ of $X(t)$ is given by formula:

$$f(\lambda) = \frac{1}{2\pi} \left| \sum_{k=-\infty}^{\infty} a(k) e^{-ik\lambda} \right|^2 = \frac{1}{2\pi} |\widehat{a}(\lambda)|^2, \quad \lambda \in [-\pi, \pi]. \tag{2.2}$$

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

(b) Continuous-time linear model. The process $\{X(t),\, t \in \mathbb{R}\}$ is a continuous-time linear process of the form:

$$X(t) = \int_{\mathbb{R}} a(t-s) d\xi(s), \qquad \int_{\mathbb{R}} |a(s)|^2 ds < \infty, \tag{2.3}$$

where $\{\xi(s), s \in \mathbb{R}\}$ is a process with orthogonal increments and $\mathbb{E}|d\,\xi(s)|^2 = ds$. The spectral density $f(\lambda)$ of $X(t)$ is given by formula:

$$f(\lambda) = \frac{1}{2\pi} \left| \int_{\mathbb{R}} e^{-i\lambda t} a(t) dt \right|^2 = \frac{1}{2\pi} |\widehat{a}(\lambda)|^2, \quad \lambda \in \mathbb{R}. \tag{2.4}$$

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

(c) Lévy-driven linear model. We first recall that a Lévy process, $\{\xi(s),\ s \in \mathbb{R}\}$ is a process with independent and stationary increments, continuous in probability, with sample-paths which are right-continuous with left limits and $\xi(0) = \xi(0-) = 0$. The Wiener process $\{B(s),\ s \geq 0\}$ is a typical example 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 (2.3), where $\xi(s)$ is a Lévy process satisfying the conditions: $\mathbb{E}\xi(s) = 0$, $\mathbb{E}\xi^2(1) = 1$ and $\mathbb{E}\xi^4(1) < \infty$. In the case where $\xi(s) = B(s)$, $X(t)$ is a Gaussian process.

The function $a(\cdot)$ in representations (2.1) and (2.3) plays the role of a *time-invariant filter*, and the linear processes defined by (2.1) and (2.3) can be viewed as the output of a linear filter $a(\cdot)$ applied to the process $\{\xi(u),\ u \in \mathbb{U}\}$, called the innovation or driving process of $X(t)$.

## 3. Data tapers and the tapered periodogram

In this section we introduce the data tapers and tapered periodogram. Our inference procedures will be based on the tapered data $\mathbf{X}_T^h$:

$$\mathbf{X}_T^h := \{h_T(t)X(t),\ t \in D_T\}, \tag{3.1}$$

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

$$h_T(t) := h(t/T) \tag{3.2}$$

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

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

$$H_{k,T}(\lambda) := \begin{cases} \sum_{t=1}^{T} h_T^k(t)e^{-i\lambda t} & \text{in the d.t. case,} \\ \int_0^T h_T^k(t)e^{-i\lambda t}dt & \text{in the c.t. case,} \end{cases} \tag{3.3}$$

and put

$$H_{k,T} := H_{k,T}(0). \tag{3.4}$$

Define the finite Fourier transform of the tapered data (3.1):

$$d_T^h(\lambda) := \begin{cases} \sum_{t=1}^{T} h_T(t)X(t)e^{-i\lambda t} & \text{in the d.t. case,} \\ \int_0^T h_T(t)X(t)e^{-i\lambda t}dt & \text{in the c.t. case.} \end{cases} \tag{3.5}$$

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

$$I_T^h(\lambda) := \frac{1}{C_T}\, d_T^h(\lambda)d_T^h(-\lambda), \tag{3.6}$$

where

$$C_T := 2\pi H_{2,T}(0) = 2\pi H_{2,T} \neq 0. \tag{3.7}$$

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

$$H_k := \lim_{T\to\infty} (1/T)H_{k,T}, \quad \textit{and } H_{k,T} \textit{ is as in } (3.4). \tag{3.8}$$

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

**Remark 3.1.** 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 $t \in [0, 1]$.

The benefits of tapering the data have been widely reported in the literature (see, e.g., Brillinger [2], Dahlhaus [3]–[6], Dahlhaus and Künsch [7], Ginovyan and Sahakyan [13, 14], Guyon [15], 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. 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. 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 Dahlhaus [5, 6], and Ginovyan and Sahakyan [14]).

## 4. Estimation of linear spectral functionals

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., Ginovyan and Sahakyan [14], and references therein). The result that follow is used to prove consistency and asymptotic normality of the minimum contrast estimators based on the Whittle functionals for linear models with tapered data. Specifically, we are interested in the nonparametric estimation problem, based on the tapered data (3.1), of the following linear spectral functional:

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

where $g(\lambda) \in L^q(\Lambda)$, $1/p + 1/q = 1$. Here, and in what follows, $\Lambda = \mathbb{R}$ in the c.t. case, and $\Lambda = [-\pi.\pi]$ in the d.t. case.

As an estimator $J_T^h$ for functional $J(f)$, given by (4.1), based on the tapered data (3.1), 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 (3.6). 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. To state the corresponding results we first introduce the following assumptions.

**Assumption 4.1.** 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.

**Assumption 4.2.** (A) (d.t. case). The spectral density $f$ and the generating function $g$ are such that $f \in L^p(\Lambda)$ $(p \geq 1)$ and $g \in L^q(\Lambda)$ $(q \geq 1)$ with $1/p + 1/q \leq 1/2$.
(B) (c.t. case). The spectral density $f$ and the generating function $g$ are such that $f \in L^1(\Lambda) \cap L^p(\Lambda)$ $(p \geq 1)$ and $g \in L^1(\Lambda) \cap L^q(\Lambda)$ $(q \geq 1)$ with $1/p + 1/q \leq 1/2$.
(C) (c.t. Lévy-driven case). The filter $a$ and the generating kernel $\widehat{g}$ are such that $a \in L^2(\Lambda) \cap L^p(\Lambda)$ and $\widehat{g} \in L^q(\Lambda)$ with $1 \leq p, q \leq 2$ and $2/p + 1/q \geq 5/2$.

Denote

$$e(h) := \lim_{T \to \infty} \frac{T H_{4,T}}{H_{2,T}^2}, \tag{4.3}$$

where $H_{k,T}$ is as in (3.4), and

$$\sigma_h^2(J) := 4\pi e(h) \int_\Lambda f^2(\lambda) g^2(\lambda) d\lambda + \kappa_4 e(h) \left[\int_\Lambda f(\lambda) g(\lambda) d\lambda\right]^2, \tag{4.4}$$

where $\kappa_4$ is the fourth cumulant of $\xi(1)$.

The proof of the next theorem can be found in Ginovyan and Sahakyan [13] (see also Ginovyan [11]).

**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, 4.1 and 4.2 the following asymptotic relation holds:*

$$\begin{aligned}
&(a) \quad \mathbb{E}(J_T^h) - J \to 0 \quad \text{as} \quad T \to \infty. \\
&(b) \quad T^{1/2}\left[\mathbb{E}(J_T^h) - J\right] \to 0 \quad \text{as} \quad T \to \infty. \\
&(c) \quad \lim_{T \to \infty} T \operatorname{Var}(J_T^h) = \sigma_h^2(J), \\
&(d) \quad T^{1/2}\left[J_T^h - J\right] \xrightarrow{d} \eta \quad \text{as} \quad T \to \infty,
\end{aligned}$$

*where $\mathbb{E}[\cdot]$ is the expectation operator, the symbol $\xrightarrow{d}$ stands for convergence in distribution, and $\eta$ is a normally distributed random variable with mean zero and variance $\sigma_h^2(J)$ given by (4.4).*

## 5. The Whittle estimation procedure

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 $\Theta$ is a subset of the Euclidean space $\mathbb{R}^p$. The problem of interest is to estimate $\theta$ on the basis of the tapered data (3.1), and investigate the asymptotic (as $T \to \infty$) properties of the suggested estimators. We use here the Whittle estimation method to estimate $\theta$. This method, originally devised by P. Whittle for d.t. stationary processes (see Whittle [17]), is based on the smoothed periodogram analysis on a frequency domain, involving approximation of the likelihood function and asymptotic properties of empirical spectral functionals. The Whittle procedure of estimation of a spectral parameter $\theta$ based on the tapered sample (3.1) is to choose the estimator $\widehat{\theta}_{T,h}$ to minimize the weighted tapered Whittle functional:

$$U_{T,h}(\theta) := \frac{1}{4\pi}\int_\Lambda \left[\log f(\lambda,\theta) + \frac{I_T^h(\lambda)}{f(\lambda,\theta)}\right]\cdot w(\lambda)\,d\lambda, \tag{5.1}$$

where $I_T^h(\lambda)$ is the tapered periodogram of $X(t)$, given by (3.6), and $w(\lambda)$ is a weight function (that is, $w(-\lambda) = w(\lambda)$, $w(\lambda) \ge 0$, $w(\lambda) \in L^1(\mathbb{R})$) for which the integral in (5.1) is well defined. In the d.t. case as a weight function we take $w(\lambda) \equiv 1$. In the c.t. case, an example of common used weight function is $w(\lambda) = 1/(1+\lambda^2)$. So, the Whittle estimator $\widehat{\theta}_{T,h}$ of $\theta$ based on the tapered sample (3.1) is defined by

$$\widehat{\theta}_{T,h} := \operatorname*{Arg\,min}_{\theta\in\Theta} U_{T,h}(\theta), \tag{5.2}$$

where $U_{T,h}(\theta)$ is given by (5.1). Thus, the tapered Whittle estimator $\widehat{\theta}_{T,h}$ of $\theta$ is the root of the following system of estimating equations:

$$\begin{aligned} F_{h,i}(\theta) &= F_{T,h,i}(\theta) := (\partial/\partial\theta_i)U_{T,h}(\theta) \\ &= \frac{1}{4\pi}\int_\Lambda \left[(\partial/\partial\theta_i)\log f(\lambda,\theta) + I_T^h(\lambda)(\partial/\partial\theta_i)f^{-1}(\lambda,\theta)\right]\cdot w(\lambda)\,d\lambda = 0, \quad i = 1,\dots,p. \end{aligned} \tag{5.3}$$

The tapered Whittle estimator $\widehat{\theta}_{T,h}$ of $\theta$ possesses good asymptotic properties. To state these properties of $\widehat{\theta}_{T,h}$, we first introduce the following set of assumptions.

**Assumption 5.1.** The true value $\theta_0$ of the parameter $\theta$ belongs to a compact set $\Theta$ 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 $\Lambda$ with respect to the Lebesgue measure.

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

**Assumption 5.3.** The functions $f := f(\lambda, \theta)$ and $g := w(\lambda)(\partial/\partial\theta_k)f^{-1}(\lambda, \theta)$ satisfy Assumption 4.1 for all $k = 1, \dots, p$ and $\theta \in \Theta$.

**Assumption 5.4.** The functions $f$, $g$, $a := a(\lambda, \theta)$ and $b := \widehat{g}$, where $g$ is as in Assumption 5.3, satisfy Assumption 4.2.

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

**Assumption 5.6.** The matrices

$$W(\theta) := \|w_{ij}(\theta)\|, \ A(\theta) := \|a_{ij}(\theta)\|, \ B(\theta) := \|b_{ij}(\theta)\|, \quad i, j = 1, \dots, p \tag{5.4}$$

are positive definite, where

$$w_{ij}(\theta) = \frac{1}{4\pi}\int_\Lambda \frac{\partial}{\partial\theta_i}\ln f(\lambda, \theta)\frac{\partial}{\partial\theta_j}\ln f(\lambda, \theta)w(\lambda)d\lambda, \tag{5.5}$$

$$a_{ij}(\theta) = \frac{1}{4\pi}\int_\Lambda \frac{\partial}{\partial\theta_i}\ln f(\lambda, \theta)\frac{\partial}{\partial\theta_j}\ln f(\lambda, \theta)w^2(\lambda)d\lambda, \tag{5.6}$$

$$b_{ij}(\theta) = \frac{\kappa_4}{16\pi^2}\int_\Lambda \frac{\partial}{\partial\theta_i}\ln f(\lambda, \theta)w(\lambda)d\lambda\int_\Lambda \frac{\partial}{\partial\theta_j}\ln f(\lambda, \theta)w(\lambda)d\lambda, \tag{5.7}$$

and $\kappa_4$ is the fourth cumulant of $\xi(1)$.

The next theorem, which was proved in Ginovyan [12], contains sufficient conditions for the tapered Whittle estimator $\widehat{\theta}_{T,h}$ to be $\sqrt{T}$-consistent and asymptotically normal.

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

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

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

$$\Gamma(\theta_0) = W^{-1}(\theta_0)\left(A(\theta_0) + B(\theta_0)\right)W^{-1}(\theta_0). \tag{5.9}$$

*Here the matrices $W$, $A$ and $B$ are defined in (5.4)-(5.7), and the tapering factor $e(h)$ is given by formula* (4.3).

**Remark 5.1** (The variance effect)**.** Since tapering of the data, roughly speaking, reduces the effective length of the data, it is not surprising that 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 (4.3)) 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 (see [6], p.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 $e(h) = 1$. Moreover, in the classical asymptotic setting, for d.t. Gaussian processes it is possible to choose the taper function $h(t)$ so that the corresponding tapered estimator will be asymptotically Fisher-efficient (for details see Dahlhaus [4, 6], Ginovyan and Sahakyan [14]).

## 6. The Le Cam-Dzhaparidze simplified estimators

We describe here the Le Cam-Dzhaparidze approach of constructing simplified estimators in the general setting (see Le Cam [16] and Dzhaparidze [8, 9]).

We first introduce the following set of assumptions (see Dzhaparidze [8]). In what follows, $\tau_T = \tau(T)$ stands for a non-random function increasing without bound as $T \to \infty$.

**Assumption 6.1.** The system of estimating equations (1.1) has a root $\hat{\theta}_T$ which is a consistent estimator of $\theta$, that is, $p - lim_{T\to\infty}\hat{\theta}_T = \theta_0$.

**Assumption 6.2.** For $\theta \in \Theta$ the derivatives $F_i^{(k)}(\theta) = (\partial/\partial\theta_k)F_i(\theta)$, $i, k = 1, \dots p$, exist, and for any arbitrarily small $\varepsilon > 0$ and $\delta > 0$

$$\mathbb{P}\left(|F_i^{(k)}(\theta_0) - w_{ik}(\theta_0)| < \varepsilon\right) \geq 1 - \delta, \tag{6.1}$$

where $F_i(\theta)$ is as in (1.1) and $W(\theta) := \|w_{ik}(\theta),\, i, k = 1, \dots, p\|$ is a non-random matrix, which is non-degenerate for $\theta = \theta_0$.

**Assumption 6.3.** The second derivatives $F_i^{(k,j)}(\theta) = (\partial^2/\partial\theta_k\partial\theta_j)F_i(\theta)$ exist, which are continuous for $\theta \in \Theta$ and $i, k, j = 1, \dots, p$, and such that for any arbitrarily small $\delta > 0$ and some $M < \infty$,

$$\mathbb{P}\left(|F_i^{(k,j)}(\theta)| < M\right) \geq 1 - \delta. \tag{6.2}$$

**Assumption 6.4.** Along with (6.1), for sufficiently large $T$, the following stronger inequality holds:

$$\mathbb{P}\left(\sqrt{\tau_T}|F_i^{(k)}(\theta_0) - w_{ik}(\theta_0)| < \varepsilon\right) \geq 1 - \delta. \tag{6.3}$$

**Assumption 6.5.** There exists a random matrix $D_* := \|d^*_{ik}\|$, $i, k = 1, \dots, p$, such that for any arbitrarily small $\varepsilon > 0$ and $\delta > 0$, the inequality

$$\mathbb{P}\left(\sqrt{\tau_T}|d^*_{ik} - d_{ik}(\theta_0)| < \varepsilon\right) \geq 1 - \delta \tag{6.4}$$

holds for sufficiently large $T$ and all $i, k = 1, \dots, p$, where $d_{ik}(\theta_0)$ are the elements of the matrix $D(\theta_0) := W^{-1}(\theta_0)$, and $W(\theta)$ is as in Assumption 6.2.

**Theorem 6.1** (Dzhaparidze [8]). *Let $\mathbf{F}(\theta)$ be a $p$-dimensional vector with elements $F_i(\theta)$, $i = 1, \dots p$, $\mathcal{F}(\theta)$ be a matrix with elements $F_i^{(k)}(\theta)$ $i, k = 1, \dots p$, and $\theta_T^*$ be an arbitrary $\tau_T^*$-consistent estimator of $\theta$, where $\sqrt{\tau_T}/\tau_T^* \to 0$ as $T \to \infty$. The following assertions hold:*

(a) *Under Assumptions 6.1-6.3 the estimator $\check{\theta}_{1,T} := \theta_T^* - \mathcal{F}^{-1}(\theta_T^*)F(\theta_T^*)$ is asymptotically equivalent to $\hat{\theta}_T$ in the sense that*

$$p - \lim_{T\to\infty} \tau_T \left(\hat{\theta}_T - \check{\theta}_{1,T}\right) = 0.$$

(b) *Under Assumptions 6.1-6.5 the estimators of the form $\check{\theta}_T := \theta_T^* - D_* F(\theta_T^*)$ are asymptotically equivalent to $\hat{\theta}_T$ in the sense that*

$$p - \lim_{T\to\infty} \tau_T \left(\hat{\theta}_T - \check{\theta}_T\right) = 0.$$

**Remark 6.1.** Comparing assertions (a) and (b) of Theorem 6.1 one easily sees that if $D_* = \mathcal{F}^{-1}(\theta_T^*)$, then the estimator $\check{\theta}_{1,T}$ coincides with $\check{\theta}_T$.

## 7. Simplified Whittle estimators for spectral parameters with tapered data

As it was stated above (see Theorem 5.1), the tapered Whittle estimator $\widehat{\theta}_{T,h}$ of $\theta$ possesses good asymptotic properties, that is, the estimator $\widehat{\theta}_{T,h}$ is $\sqrt{T}$-consistent and asymptotically normal. Moreover, for d.t. Gaussian models it is also asymptotically Fisher-efficient (see Remark 5.1).

However, generally, the estimating equations (5.3) are non-linear, and it is a challenging problem to find the estimator $\widehat{\theta}_{T,h}$. So, it is important finding simpler estimators of the parameter $\theta$ having the same asymptotic properties as $\widehat{\theta}_{T,h}$. The estimators proposed here are asymptotically equivalent to the estimator $\widehat{\theta}_{T,h}$ under rather broad regularity conditions on the spectral density function $f(\lambda, \theta)$.

**Theorem 7.1.** *Let $\mathbf{F}_h(\theta)$ be a $p$-dimensional vector with elements $F_{i,h}(\theta)$ ($i = 1, \dots p$) given by* (5.3)*, $\theta_{T,h}^*$ be an arbitrary $\tau_T^*$-consistent estimator of $\theta$, where $\sqrt[4]{T}/\tau_T^* \to 0$ as $T \to \infty$, and let $D_* := \|d_{ik}^*\|$, $i, k = 1, \dots, p$, be a random matrix whose elements $d_{ik}^*$ satisfy the condition* (6.4)*. Then under assumptions of Theorem 5.1 the estimators of the form*

$$\check{\theta}_{T,h} := \theta_{T,h}^* - D_* \mathbf{F}_h(\theta_{T,h}^*) \tag{7.1}$$

*are asymptotically equivalent to the tapered Whittle estimator $\hat{\theta}_{T,h}$ in the sense that*

$$p - \lim_{T\to\infty} \sqrt{T} \left(\hat{\theta}_{T,h} - \check{\theta}_{T,h}\right) = 0.$$

**Proof.** The result we deduce from Theorem 6.1 by using Theorem 4.1. We show that the Assumptions 6.2–6.4 are satisfied for functions $F_{h,i}(\theta)$ $(i = 1, \dots p)$.

First, applying Theorem 4.1(a) we easily conclude that for $k, j = 1, \dots, p$,

$$\lim_{T\to\infty} \mathbb{E}_0 \left[U^{(kj)}_{T,h}(\theta_0)\right] = w_{kj}(\theta_0) =$$

$$= \frac{1}{4\pi} \int_\Lambda \frac{\partial}{\partial\theta_k} \ln f(\lambda, \theta_0) \frac{\partial}{\partial\theta_j} \ln f(\lambda, \theta_0) w(\lambda) d\lambda, \tag{7.2}$$

where $\mathbb{E}_0[\xi]$ stands for expectation with respect to probability $P_0$, corresponding to spectral density $f(\lambda, \theta_0)$, and $U^{(kj)}_{T,h}(\theta) = (\partial^2/\partial\theta_k\partial\theta_j)U_{T,h}(\theta)$ with $U_{T,h}(\theta)$ as in (5.1) (for details see Ginovyan [12]).

Next, by applying Theorem 4.1(c), for the variance of $U^{(kj)}_{T,h}(\theta_0)$ $(k, j = 1, \dots, p)$, we have

$$\lim_{T\to\infty} \sqrt{T}\mathrm{Var}\left(U^{(kj)}_{T,h}(\theta_0)\right) = 0. \tag{7.3}$$

Therefore, by Chebyshev's inequality it follows that, for sufficiently large $T$,

$$\mathbb{P}\left(\sqrt[4]{T}|F^{(k)}_{h,i}(\theta_0) - w_{ik}(\theta_0)| < \varepsilon\right) \geq 1 - \delta, \tag{7.4}$$

where $\varepsilon > 0$ and $\delta > 0$ are arbitrary small numbers. Hence, Assumption 6.4 is satisfied with $\tau_T = \sqrt{T}$. Since the matrix $W(\theta) := \|w_{ik}(\theta), i, k = 1, \dots, p\|$ is assumed to be non-degenerate for $\theta = \theta_0$, Assumption 6.2 also holds. Finally, using Theorem 4.1(a) and (c), we easily infer that the function

$$F^{(kj)}_{T,h,i}(\theta) = \frac{1}{4\pi} \int_\Lambda I^h_T(\lambda) \frac{\partial^3}{\partial\theta_i\partial\theta_k\partial\theta_j} f^{-1}(\lambda, \theta) w(\lambda) d\lambda$$

satisfies Assumption 6.3. Thus, the result follows from Theorem 6.1(b). □

**Corollary 7.1.** *Let $\mathcal{F}(\theta)$ be a matrix with elements $F^{(k)}_{h,i}(\theta)$ $(i, k = 1, \dots p)$. Then under the conditions of Theorem 7.1 the estimator*

$$\check{\theta}_{1,T,h} := \theta^*_{T,h} - \mathcal{F}^{-1}(\theta^*_{T,h})\mathbf{F}_h(\theta^*_{T,h}), \tag{7.5}$$

*is also asymptotically equivalent to the estimator $\hat{\theta}_{T,h}$.*

**Corollary 7.2.** *Assume that for $\theta \in \Theta$ there exist continuous derivatives $(\partial/\partial\theta_i)w_{k,j}(\theta)$ $(i, k, j = 1, \dots, p)$ satisfying $|(\partial/\partial\theta_i)w_{i,k}(\theta)| < C$, where the constant $C$ does not depend on $\theta$. Then under the conditions of Theorem 7.1, the estimator*

$$\check{\theta}_{2,T,h} := \theta^*_{T,h} - W^{-1}(\theta^*_{T,h})\mathbf{F}_h(\theta^*_{T,h}), \tag{7.6}$$

*is also asymptotically equivalent to the estimator $\hat{\theta}_{T,h}$.*

Indeed, applying the theorem on the mean we easily conclude that the elements of the matrix $D(\theta^*_{T,h}) = W^{-1}(\theta^*_{T,h})$ satisfy the condition (6.4), and hence may be chosen as the $d^*_{i,k}$.

**Remark 7.1.** It is easy to see that, similar to the non-tapered case (see Dzhaparidze [8]), the estimators $\check{\theta}_{1,T,h}$ and $\check{\theta}_{2,T,h}$ can be constructed comparatively easily. In fact, to find them it is necessary to have available some $\tau_T^*$-consistent estimator with $\sqrt[4]{T}/\tau_T^* \to 0$ as $T \to \infty$ and to determine the matrices $\mathcal{F}^{-1}(\theta)$ and $W^{-1}(\theta)$, respectively. Observe also that the estimators $\check{\theta}_{T,h}$, $\check{\theta}_{1,T,h}$ and $\check{\theta}_{2,T,h}$ are of interest only if it is too difficult to solve the system of estimating equation (5.3) directly for practical use. In the cases where the equations in (5.3) are linear (and so easily solved), then clearly the estimator $\check{\theta}_{1,T,h}$ coincides with $\hat{\theta}_{T,h}$.

## Список литературы


[1] G. Beinicke, K. O. Dzhaparidze, "On parameter estimation by the Davidon-Fletcher-Powell method", Theory Probab. Appl., **27**, 396 – 402 (1982).

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

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

[4] 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).

[5] R. Dahlhaus, "Small sample effects in time series analysis: a new asymptotic theory and a new estimate", Ann. Stat., **16**, 808 – 841 (1988).

[6] R. Dahlhaus, "Nonparametric high resolution spectral estimation", Probab. Th. Rel. Fields, **85**, 147 – 180 (1990).

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

[8] K. O. Dzhaparidze, "On simplified estimators of unknown parameters with good asymptotic properties", Theory Probab. Appl., **19**, 347 – 358 (1974).

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

[10] K. O. Dzhaparidze, "On iterative procedures of asymptotic inference", Statistica Neeriandica, **37**, 181 – 189 (1983).

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

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

[13] M. S. Ginovyan, 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).

[14] M. S. Ginovyan, A. A. Sahakyan, "Statistical inference for stationary models with tapered data", Statistics Surveys, **15**, 154 – 194 (2021).

[15] X. Guyon, Random Fields on a Network: Modelling, Statistics and Applications, Springer, New York (1995).

[16] L. Le Cam, "On the asymptotic theory ofestimation and testing hypotheses", Proc. 3rd Berkeley Sympos. Math. Stat. Probab., **1**, 129 – 156 (1956).

[17] P. Whittle, Hypothesis Testing in Time Series, Hafner, New York (1951).



Поступила 03 апреля 2023

После доработки 03 апреля 2023

Принята к публикации 15 августа 2023