Journal of Multivariate Analysis 115 (2013) 405–420

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A frequency domain bootstrap for Whittle estimation under long-range dependence

Young Min Kim [a,b,*], Daniel J. Nordman [a]

[a] Department of Statistics, Iowa State University, Ames, IA 50010, USA

[b] Department of Biostatistics, University at Buffalo, SUNY, Buffalo, NY 14214, USA

ARTICLE INFO

Article history:
Received 28 April 2012
Available online 21 November 2012

AMS subject classification:
62G09

Keywords:
FARIMA
Interval estimation
Long memory
Spectral density
Periodogram

ABSTRACT

Whittle estimation is a common technique for fitting parametric spectral density functions to time series, in an effort to model the underlying covariance structure. However, Whittle estimators from long-range dependent processes can exhibit slow convergence to their Gaussian limit law so that calibrating confidence intervals with normal approximations may perform poorly. As a remedy, we study a frequency domain bootstrap (FDB) for approximating the distribution of Whittle estimators. The method provides valid distribution estimation for a broad class of stationary, long-range (or short-range) dependent linear processes, without stringent assumptions on the distribution of the underlying process. A large simulation study shows that the FDB approximations often improve normal approximations for setting confidence intervals for Whittle parameters in spectral models with strong dependence.

1. Introduction

We consider a problem concerning parametric spectral density estimation for time series that could exhibit strong forms of dependence. Suppose a real-valued stationary time process $\{X_t\}$ has an integrable spectral density function $g(\lambda)$, $\lambda \in \Pi = (-\pi, \pi]$, behaving as

$$\lim_{\lambda \to 0} |\lambda|^{2d} g(\lambda) = C \tag{1}$$

for some $d \in [0, 1/2)$ and positive constant $C > 0$. We refer to the process $\{X_t\}$ as weakly or short-range dependent (SRD) when $d = 0$, and call the process strongly or long-range dependent (LRD) when $d > 0$; in the following, we also abbreviate short- or long-range dependence as "LRD/SRD" for simplicity. This dependence-type classification is a common, in which LRD entails a pole of g at the origin [20,4]. Time series exhibiting LRD often have applications in astronomy, hydrology and economics, where correlations may decrease particularly slowly between observations over time (cf. [36,18]). That is, LRD can be alternatively formulated in terms of a slow decay of process autocovariances $r(k) = \mathrm{Cov}(X_t, X_{t+k}) \approx ak^{-1+2d}$ as $|k| \to \infty$ for some $a > 0$, entailing that $\sum_{k=1}^{\infty} r(k)$ is not finitely summable unlike the usual SRD case; see [5, p. 240] or [43, p. 1634] for mild conditions under which this covariance behavior is equivalent to (1).

While some semiparametric approaches focus on estimating the long-memory exponent d of LRD processes (e.g., [13,43,44,37,3]), we consider an inference scenario involving a parametric collection of spectral densities

$$\mathcal{F} \equiv \left\{ g(\lambda; \sigma^2, \theta) = \frac{\sigma^2}{2\pi} f(\lambda; \theta) : \sigma^2 > 0, \theta \in \Theta \subset \mathbb{R}^p \right\}, \tag{2}$$

* Corresponding author at: Department of Statistics, Iowa State University, Ames, IA 50010, USA.
E-mail address: ykim25@buffalo.edu (Y.M. Kim).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2012.10.018

defined by a kernel density $f(\lambda;\theta)$ involving p parameters $\theta = (\theta_1, \ldots, \theta_p)'$ which can include d. This particular class form is often considered for modeling the covariance structure of broad classes of linear processes, which could exhibit SRD (e.g., autoregressive moving averages (ARMA) models) or LRD. Important spectral models in the latter case include the fractional Gaussian processes of [33] and the fractional autoregressive integrated moving average (FARIMA) models of [1,15,20]. For fitting such models to data, Whittle estimation is a common, computationally feasible approach [47]. For SRD linear processes, [46,17] established the consistency and asymptotic normality of Whittle estimators of the parameters θ. Fox and Taqqu [9] proved the same properties for Whittle estimators with LRD Gaussian processes, while [14] extended these results to more general linear LRD time series. These results enable confidence regions for θ to be calibrated from Whittle estimators with large-sample normal approximations. However, it has been noted that a normal approximation may not adequately reflect the finite-sample distribution of Whittle estimators (cf. [34]), which can be more asymmetric [25]. For example, [40] showed a stochastic tendency of estimators of the long-memory parameter to be smaller than the true values of d in FARIMA models.

As an alternative to the normal approximation, we develop a frequency domain bootstrap (FDB) for approximating the distribution of Whittle estimators under LRD. This method has the advantage of allowing inference without knowledge or stringent assumptions on the full probability structure of the time process. Under SRD, [7] originally established a FDB for so-called "ratio" statistics. The main idea is that a data transformation (i.e., Fourier transform) can weaken the dependence structure so that the periodogram ordinates, when properly scaled to normalize variances, can be independently resampled to create bootstrap versions of spectral estimators. Our results provide a type of extension of the FDB to LRD processes and to Whittle estimation in particular. However, there is a difference in formulating the FDB here with regard to the scaling or normalizing step for the periodogram. Namely, the resampling scheme in [23] involves first scaling the periodogram by a nonparametric kernel estimator of the spectral density. Similar approaches have been applied to formulating bootstraps for other problems in the frequency domain under SRD, such as nonparametric spectral density estimation (cf. [11,38]). Presently, defining a FDB under LRD as a straightforward copy of the mechanics used in the SRD case is difficult because, to our knowledge, appropriate nonparametric estimators of the spectral density are currently unavailable under LRD for analogous purposes of periodogram scaling in a FDB. That is, under LRD, it is still an open problem to develop a nonparametric estimator of the spectral density which is uniformly consistent on the entire spectrum $(0, \pi]$. But such estimators do exist under SRD (e.g., [48]) which is a critical component underlying the FDB of [23] and other frequency domain resampling methods under SRD (cf. [29]). However, for Whittle estimation, it becomes possible to define a valid FDB under LRD by re-scaling periodogram ordinates with an estimated spectral density from the model class (2). While this extension of the FDB under LRD is then particular to Whittle estimation, interval estimation of Whittle parameters may be the most relevant application of the FDB under LRD, and the resulting FDB under LRD requires no assumptions about the full probability structure of the time process. A simulation study to follow also suggests that the FDB method generally outperforms a normal approximation for estimating Whittle parameters. Additionally, resampling methods developed for SRD can fail under LRD (e.g., block bootstrap, [31]), so that the validity of the FDB under LRD does not follow from the SRD case.

The paper is organized as follows. We end this section by briefly summarizing other resampling literature under LRD. Section 2 describes the Whittle estimation problem and the associated FDB method. Section 3 gives the main distributional results on the consistency of FDB method for distribution estimation of Whittle estimators under LRD. Section 4 summarizes a simulation study to compare the performance of the FDB method against normal approximations for interval estimation of Whittle parameters under several LRD models. Perhaps surprisingly, non-studentized versions of FDB distribution estimators are often better for interval estimation than studentized FDB versions or normal approximations. Section 5 provides some concluding remarks. The proofs of all results are deferred to supplementary material and, to ease the exposition, the supplementary material includes complete numerical tables of simulation results, which are graphically summarized in Appendix A here.

Lahiri [31] showed the moving block bootstrap is invalid for estimating the distribution of a sample mean from a class of LRD transformed Gaussian processes, though [16,39] established that a subsampling method is consistent for the same inference problem with several types of LRD processes. McElroy and Politis [35] proposed a subsampling method for distribution estimation of self-normalized sample means from processes having infinite invariance and LRD properties. Kim and Nordman [27] studied optimal block sizes for a block bootstrap variance estimator with stationary linear processes exhibiting LRD. Zhou and Taqqu [50,51] studied how three types of permutation methods (e.g., external, internal and two-level) for stationary time series exhibiting LRD perform in capturing the autocovariance structure; these are variations on block bootstrap methods with the external permutation representing a non-overlapping block bootstrap. Kapetanios and Psaradakis [26] and Poskitt [42] extended an autoregressive, or sieve bootstrap (cf. [12,6]) to causal linear LRD processes. Andrews et al. [2] studied the coverage accuracy of a parametric bootstrap for LRD Gaussian processes. In the frequency domain, [19] investigated a bootstrap for regression models with causal linear LRD processes. Kreiss and Paparoditis [29] developed a modified (i.e., autoregressive-aided) FDB for causal linear SRD processes which is valid for more spectral estimation than the version of [23]. Under SRD, their modified resampling scheme has been extended to obtaining time domain resamples from a technique of inverting frequency domain resamples (cf. [21,24,28,30]). See also [32] for a further summary of resampling methods for dependent data, including LRD.

While we adopt frequency domain resampling (i.e., bootstrapping the periodogram) to estimate the distribution of Whittle estimators, similar distributional estimators may be possible by re-creating data samples in the time domain through permutation methods (cf. [41, Chapter 8]) or time-domain bootstrap approaches (cf. [32, Chapter 4]). One possibility