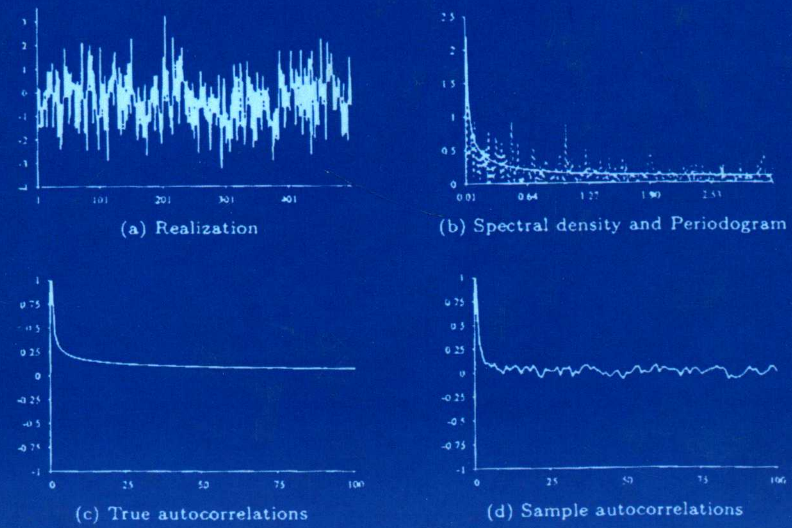
Liudas Giraitis
Hira L Koul
Donatas Surgailis



Large Sample Inference *for* Long Memory Processes

Imperial College Press

Large Sample Inference *for* Long Memory Processes

Liudas Giraitis
Queen Mary, University of London, UK

Hira L. Koul
Michigan State University, USA

Donatas Surgailis
Vilnius University, Lithuania

ICP
Imperial College Press

Published by

Imperial College Press
57 Shelton Street
Covent Garden
London WC2H 9HE

Distributed by

World Scientific Publishing Co. Pte. Ltd.
5 Toh Tuck Link, Singapore 596224
USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

LARGE SAMPLE INFERENCE FOR LONG MEMORY PROCESSES

ISBN-13 978-1-84816-278-5
ISBN-10 1-84816-278-2

Printed in Singapore by World Scientific Printers.

Large Sample Inference *for* Long Memory Processes

Preface

A discrete-time stationary stochastic process is said to have long memory if its autocovariances tend to zero hyperbolically slowly as the lag tends to infinity, but their sum diverges. Such processes have unbounded spectral densities at the origin. This is unlike the so-called short memory processes where autocovariances are summable, often tending to zero exponentially fast and whose spectral densities are bounded at the origin.

Ever since the work of Hurst (1951, 1956), the proponent of the Aswan High Dam, more and more scientists have found the presence of long memory in their data. Hurst noticed that measurements on the Nile River were not consistent with the assumption of independence. Mandelbrot and van Ness (1968) and Mandelbrot and Wallis (1969a,b) were the first to provide a theoretical justification for this by advancing the idea that the data observed by Hurst follow a long memory process. They proposed the use of fractional Brownian motion, as opposed to classical Brownian motion, to model and analyze various phenomena in hydrology, and more generally, for modeling a long memory time series. The review paper of Lawrance and Kottegoda (1977) nicely summarizes this fact and some of the other stochastic models used in the modeling of river-flow time series.

Box and Jenkins (1970) popularized the idea of obtaining a stationary time series by differencing the given, possibly non-stationary, time series. Numerous time series in economics are found to have this property, i.e., even though the initial time series is not stationary, its dth-order difference, for some positive integer d, is stationary. Subsequently, Granger and Joyeux (1980) and Hosking (1981) found examples of time series whose fractional difference is a short memory process, in particular, white-noise, while the initial series has unbounded spectral density at the origin. Beran (1992a) gives examples from numerous other sciences where data follow long memory. Baillie (1996) cites several references that demonstrate the dramatic empirical success of long memory processes in modeling the volatility of the asset prices and power transforms of stock market returns while Ding and Granger (1996) point out the long memory property of the absolute values and the squares of S&P 500 daily stock market index returns. Willinger, Paxson, Riedi and Taqqu (2003) discuss the importance of long memory processes in network traffic data.

For classical time-series analysis the text of Brockwell and Davis (1991) is an excellent source. The monograph of Beran (1994) provides a nice introduction to some basic notions and applications of long memory processes. However, there have been significant advances in theoretical aspects of long memory processes since the mid-1990s that need to be made available in a unified fashion in one place. The monographs by Doukhan, Oppenheim and Taqqu (2003), Dehling, Mikosch and Sørensen (2002), and Robinson (2003) consist of collections of papers that discuss and review various theoretical results of long memory processes and their applications, while that of Teyssière and Kirman (2007) contains a collection of papers that emphasize the presence of long memory in economics and finance. The text of Palma (2007) summarizes some statistical theory and applications of these processes. Some theoretical aspects of asymptotic theory for long memory processes are discussed in Chapter 5 of the monograph by Taniguchi and Kakizawa (2000). Various connections of long memory with non-stationary and regime switching processes, self-similar processes, and the Hurst phenomenon are outlined in Samorodnitsky (2007).

At present there is a need for a text where an interested reader can methodically learn some basic asymptotic theory techniques found useful in the analysis of statistical inference procedures for long memory processes. This text makes an attempt in this direction. Our goal here is to provide in a concise style a text at the graduate level summarizing theoretical developments both for short and long memory processes and their applications to statistics. It also contains some real-data applications and mentions some unsolved inference problems for interested researchers in the field at the time of writing this monograph.

This book can be used by doctoral students needing to familiarize themselves with the detailed proofs and derivations. It can also be used as a source of theoretical tools for further investigations in econometrics and statistics and as a theoretical background for practical applications and modeling. Parts of the text can also be used by students in Masters' programs in statistics or econometrics.

The literature on long memory processes is vast. We were influenced by numerous works and the ideas of researchers working in the field. We would like to thank our teachers, colleagues, co-authors, and students. In particular, we are grateful to Rainer Dahlhaus, Ronald L. Dobrushin, Peter C.B. Phillips, Peter M. Robinson, and Murad S. Taqqu. We thank Violetta Dalla for providing some of the graphs and simulations included in this text, and Natalia Bailey and Remigijus Leipus for careful reading of the manuscript. We are also grateful to Shama Koul and Rūta Surgailienė for their patience and support during this long quest. We thank Shama and Ajeet Koul for hosting numerous dinners and facilitating a conducive atmosphere for the completion of this project.

The Department of Statistics and Probability at Michigan State University, the School of Economics and Finance of Queen Mary, University of London, and the Institute of Mathematics and Informatics, Vilnius University, have been particularly generous in providing excellent support and working environments.

During the preparation of this monograph the authors were supported in part by the NSF DMS Grants 00-71619, 07-04130, ESRC Grant RES062230790, and a grant from the Lithuanian State Science and Studies Foundation and the Research Council of Lithuania.

Liudas Giraitis
School of Economics and Finance
Queen Mary, University of London

Hira L. Koul
Department of Statistics and Probability
Michigan State University

Donatas Surgailis
Institute of Mathematics and Informatics
Vilnius University

February 10, 2012

Notation and Conventions

The following notation and conventions are used throughout the book. The symbol " $:=$ " stands for "by definition". All limits are taken as $n \to \infty$, unless specified otherwise.

$$\mathbb{R} := (-\infty, \infty), \quad \bar{\mathbb{R}} := [-\infty, \infty], \quad \mathbb{R}^+ := [0, \infty).$$

$$\mathbb{Z} := \{0, \pm 1, \pm 2, \cdots\}, \quad \Pi := [-\pi, \pi], \quad \mathbf{i} := (-1)^{1/2}.$$

$a.e.$:= almost everywhere.

$a.s.$:= almost surely.

$a \wedge b := \min(a,\, b), \quad a \vee b := \max(a,\, b).$

$$B(a, b) := \int_0^1 u^{a-1}(1-u)^{b-1} du, \; a \wedge b > 0.$$

CLT := the central limit theorem.

$\mathcal{C}(A)$:= class of continuous functions defined on a set A.

C–S := the Cauchy–Schwarz inequality.

DCT := the dominated convergence theorem.

ET := the ergodic theorem.

d.f. := distribution function.

$\|g\|_\infty$:= the supremum norm over the domain of a function g.

$I(A)$:= indicator of the set A.

i.i.d. := independent identically distributed.

l.h.s. := left-hand side.

$L_p(A)$:= class of p-integrable functions defined on a set A, $p \in \mathbb{R}$.

MSE := mean-squared error.

$\mathcal{N}_q(\mathbf{0}, \mathbf{C})$:= a q-variate normal distribution with the mean vector $\mathbf{0}$ and the covariance matrix $\mathbf{C}$.

$\mathcal{N}(0,1) := \mathcal{N}_1(0,1)$.

Φ := the d.f. of $\mathcal{N}(0,1)$ r.v.

r.h.s. := right-hand side.

r.v. := random variable or vector.

$O_p(1)$:= a sequence of r.v.'s that is bounded, in probability.

$o_p(1)$:= a sequence of r.v.'s converging to zero, in probability.

$\text{sgn}(\cdot)$:= sign function.

$\lfloor x \rfloor$:= the largest integer not greater than x.

$[x] := \lfloor x \rfloor,\ x \geq 0;\ \lfloor x \rfloor + 1,\ x < 0$.

z_α := $(1-\alpha)$th percentile of $\mathcal{N}(0,1)$ distribution, $0 \leq \alpha \leq 1$.

Z := a $\mathcal{N}(0,1)$ r.v., unless mentioned otherwise.

w.r.t. := with respect to.

$u_p(1)$:= a sequence of stochastic processes converging to zero uniformly over the time domain, in probability.

Equality of distribution is denoted by $=_D$, while $\rightarrow_p$ and $\rightarrow_D$, respectively, denote the convergence in probability and in distribution.

For a sequence of stochastic processes $\{Y, Y_n,\ n \geq 1\}$, $Y_n \rightarrow_{fdd} Y$ denotes the weak convergence of finite-dimensional distributions of Y_n to the corresponding finite-dimensional distributions of Y, and $Y_n \Rightarrow Y$ means that Y_n converges weakly to Y in the given topology.

For any two real sequences a_n, $b_n, n \geq 1$, $a_n \sim b_n$ denotes convergence $a_n/b_n \rightarrow 1$, and $a_n \propto b_n$ means that $C_1 \leq a_n/b_n \leq C_2$, for some $C_1, C_2 > 0$, as $n \rightarrow \infty$.

The kth derivative of a smooth function g is denoted by $g^{(k)}$, $k = 1, 2, \cdots$. Often we write $\dot{g} = g^{(1)}$, $\ddot{g} = g^{(2)}$.

Contents

Chapter 1

Introduction

In this text the dependence structure of a stationary process is described by its autocovariance and spectral density functions. In addition, it is assumed that the given stationary process has a linear structure with either i.i.d. or white-noise innovations. The rate of decay of the coefficients of a linear process determines the type of dependence, which may be weak or strong.

A discrete-time stationary stochastic process X_j, $j \in \mathbb{Z}$, with finite variance is said to have long memory if its autocovariances $\gamma(k) := \mathrm{Cov}(X_0, X_k) \sim c_\gamma |k|^{-2d}$ tend to zero hyperbolically slowly in the lag k, for some $0 < d < 1/2$ and finite $c_\gamma \neq 0$. The sum of autocovariances of a long memory process diverges, and such processes have an unbounded spectral density at the origin. In contrast, a weakly dependent or short memory process has absolutely summable autocovariances that often tend to zero exponentially fast, and a continuous and bounded spectral density. Numerous classical methods that are useful in analyzing weakly dependent stationary time series are inapplicable to long memory processes. The ultimate statistical inference theory must be broad enough to accommodate both short and long memory processes.

The need for new statistical inference methods arises, for instance, from the fact that under long memory the sample mean estimate $\bar{X}_n := n^{-1} \sum_{j=1}^{n} X_j$ of the mean $\mu = EX_1$ of a linear process is consistent at the rate $n^{d-1/2}$, which is slower than the classical rate $n^{-1/2}$, while $n^{1/2-d}(\bar{X}_n - \mu)$ is still asymptotically normally distributed. Thus the use of the classical confidence intervals for μ based on $n^{1/2}$ scaling is unjustified.

In long memory processes, the dependence between the current observation and the one at a distant future is persistent, i.e., observations that are distant from each other continue to have a linear relationship. This fact alone leads to some surprising results. For example, under long memory, any other location invariant estimator $\hat{\mu}_n$ of μ is a first-order equivalent to $\bar{X}_n$, i.e., $n^{1/2-d}(\hat{\mu}_n - \bar{X}_n)$ tends to zero, in probability. In comparison, for weakly dependent processes, observations distant from each other are approximately uncorrelated, which in turn yields that the weak limit of $n^{1/2}(\hat{\mu}_n - \bar{X}_n)$ is non-degenerate Gaussian, for a large class of estimators $\hat{\mu}_n$. Similar first-order degeneracy is observed between the least-squares estimator and the robust estimators in linear and non-linear regression models when errors follow a long memory moving-average process. This is discussed in Chapters 10 and 11. A resounding conclusion from these results is that in the presence of long memory moving-average errors, there is no gain in using robust estimators of regression parameters over the least-squares estimator, for the purpose of the first-order large sample inference.

In order to develop a rigorous asymptotic theory we need various preliminaries. Chapter 2 reviews some facts from trigonometric series, analysis, and probability. It includes some examples of time-series models and definitions of slowly varying functions and Hermite and Appell polynomials.

Chapter 3 defines short and long memory processes in the time and spectral domains, and provides some characteristics of these processes. This chapter also discusses self-similar processes and fractional Brownian motion and their connection with long memory processes.

A large class of tests and estimators in many time-series models are based on sums and weighted sums of underlying observations. Chapter 4 introduces the basic theoretical tools and provides minimal sufficient conditions for asymptotic normality of the weighted sums of a linear process. The methods of proof include the cumulant method and the method of approximation by m-dependent variables. The methodology presented also provides techniques for analyzing asymptotic distributions of sums of functions of long and short memory Gaussian process. The results of this chapter are useful in deriving asymptotic distributions of various inference procedures in regression models with dependent errors in Chapters 9 and 12.

Given that the sample $X_1, \cdots, X_n$ is fully represented by discrete Fourier transforms (DFTs) $w(u_1), \cdots, w(u_n)$, computed at Fourier frequencies $u_1, \cdots, u_n$, and that the periodogram at frequency u_j equals $|w(u_j)|^2$,