2012

On nonparametric likelihood methods for weakly and strongly dependent time series

Young Min Kim

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Economics Commons, Mathematics Commons, and the Statistics and Probability Commons

Recommended Citation

Kim, Young Min, "On nonparametric likelihood methods for weakly and strongly dependent time series" (2012). Graduate Theses and Dissertations. 12365.
https://lib.dr.iastate.edu/etd/12365

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
On nonparametric likelihood methods for weakly and strongly dependent time series

by

Young Min Kim

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:
Daniel J. Nordman, Major Professor
Petruța C. Caragea
Heike Hofmann
Mark S. Kaiser
Stephen B. Vardeman

Iowa State University
Ames, Iowa
2012
Copyright © Young Min Kim, 2012. All rights reserved.
DEDICATION

I would like to dedicate this thesis to my parents Kim, Kwangwoong and Kang, Sooja and to my siblings Kim, Taeyoung and Kim, Jina whose support I would not have been able to complete this work. I would also like to thank my friends - Mr. and Mrs. Riddles, Eunice Kim, Jongho Im and other classmates at the Department of Statistics and Jihyoek Choi and Hyunsoo Kim for their loving guidance during the writing of this work. Also, I would like to thank our department staffs, Jeanette La Grange, Denise Riker, Sharon Shepard and Marlene Tjernagel.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .............................................................. vi

CHAPTER 1. INTRODUCTION .................................................. 1

CHAPTER 2. A PROGRESSIVE BLOCK EMPIRICAL LIKELIHOOD METHOD FOR TIME SERIES .................................................. 5

  2.1 Introduction ................................................................. 6

  2.2 Progression block empirical likelihood ................................. 8

      2.2.1 Description .......................................................... 8

      2.2.2 Main distributional results ....................................... 11

  2.3 Numerical studies ........................................................ 13

  2.4 Data examples ............................................................ 15

       2.4.1 U.S. unemployment rates ........................................ 15

       2.4.2 Records of hemispheric temperatures ........................ 16

  2.5 Conclusion remarks ..................................................... 18

  2.6 Proofs of main results ................................................. 19

       2.6.1 Proof of Theorem 1 ............................................... 21

       2.6.2 Proof of Theorem 2 ............................................... 22

CHAPTER 3. PROPERTIES OF A BLOCK BOOTSTRAP UNDER LONG-RANGE DEPENDENCE ................................................. 33

  3.1 Introduction ............................................................... 33

  3.2 Block bootstrap distribution estimation under LRD ................ 35

       3.2.1 Target processes .................................................. 35

       3.2.2 Block bootstrap distributional results ......................... 36
3.3 Simulation study of bootstrap confidence intervals .......................... 40
  3.3.1 Simulation design ............................................. 40
  3.3.2 Summary of results ........................................... 42
3.4 Block bootstrap variance estimation under LRD ............................. 44
  3.4.1 Large-sample bias and variance properties ............................ 45
  3.4.2 Optimized block sizes and mean-squared error ..................... 47
3.5 Concluding remarks .................................................. 50
3.6 Proofs of main results ................................................ 50
  3.6.1 Proof of Theorem 1 .............................................. 51
  3.6.2 Proof of Theorem 2 .............................................. 53

CHAPTER 4. A FREQUENCY DOMAIN BOOTSTRAP FOR WHITTLE
ESTIMATION UNDER LONG-RANGE DEPENDENCE .......................... 65
4.1 Introduction ..................................................................... 66
4.2 Estimation Problem and Bootstrap Method ................................ 69
  4.2.1 Whittle estimation with linear time processes .................... 69
  4.2.2 A frequency domain bootstrap method for Whittle estimation . 72
4.3 Main Distributional Results ............................................. 74
  4.3.1 Assumptions .......................................................... 74
  4.3.2 Distributional results ................................................ 75
4.4 Simulation Studies .......................................................... 77
  4.4.1 Properties of Whittle estimators ................................ 78
  4.4.2 Numerical studies of coverage accuracy in estimation ............ 83
4.5 Concluding Remarks ....................................................... 96
4.6 Preliminary results and lemmas ......................................... 97
  4.6.1 Preliminary results for proving Theorem 1: Lemmas 1-7 .......... 98
  4.6.2 Preliminary results for proving Theorem 2: Lemmas 8-10 ....... 105
4.7 Proofs of main results .................................................... 115
  4.7.1 Proof of Theorem 1 ............................................... 115
  4.7.2 Proof of Theorem 2 ............................................... 118
CHAPTER 5. GENERAL CONCLUSIONS ................................. 140

5.1 General Discussion ............................................. 140
5.2 Recommendations for Future Research ...................... 141
ACKNOWLEDGEMENTS

I would like to take this opportunity to express my thanks to those who helped me with various aspects of conducting research and the writing of this thesis. First and foremost, Dr. Daniel J. Nordman for his guidance, patience and support throughout this research and the writing of this thesis. His insights and words of encouragement have often inspired me and renewed my hopes for completing my graduate education. I would also like to thank my committee members for their efforts and contributions to this work: Professors Petruța C. Caragea, Heike Hofmann, Mark S. Kaiser and Stephen B. Vardeman. I would additionally like to thank Dr. Kim and Dr. Lahiri for his guidance throughout the initial stages of my graduate career.
CHAPTER 1. INTRODUCTION

This dissertation investigates and develops three different nonparametric likelihood methods (e.g., resampling or empirical likelihood) for time series. For handling dependent data, current statistical methodology often relies on selecting parametric distributional models to accurately represent the data-generating process, which can be challenging. Additionally, inference drawn from a mistaken or misspecified probability model can potentially be misleading. Hence, the general advantage of nonparametric likelihood methods for time series is that these often allow valid statistical inference without stringent modeling assumptions about the underlying data distribution or exact dependence structure.

Specifically, the three nonparametric likelihood methods considered are

Method 1: a blockwise empirical likelihood (EL),

Method 2: a block bootstrap,

Method 3: a frequency domain bootstrap.

Each method differs in form for building a nonparametric likelihood, but all methods involve “setting empirical probabilities” on observed data. EL creates a multinomial likelihood through a process of “probability profiling” data values. That is, probabilities are placed on data values, typically under a constraint involving expectations and estimating functions, and the product of these probabilities produces an EL function for inference (cf. Owen, 2001). In contrast, bootstrap methods resample data values, through an empirical probability distribution placed on these, in an effort to re-create pseudo versions of data (bootstrap data sets) which can be applied for inference (cf. Lahiri, 2003).

An important consideration in designing such methods for time series is appropriately accommodating the underlying (unknown and potentially complicated) time dependence struc-
ture. This issue has been recognized since Singh’s (1981) example demonstrating that Efron’s (1979) original bootstrap for independent data fails under dependence. There are two general strategies for handling time dependence in resampling/EL methods (cf. Lahiri, 2003). One approach involves so-called “data blocking” whereby the original time series data are replaced by data blocks, consisting of consecutive groups of data points in time. Such data blocking helps to capture or preserve the dependence between neighboring temporal observations, and this approach is used in methods 1 and 2 above. A different technique for treating the dependence structure in a resampling method is a data transformation. The goal of a data transformation is to weaken the dependence structure, without completely distorting it, whereby transformed observations can be handled as if these were (approximately) independent. This principle is applied in developing method 3.

An additional theme of this dissertation is the type or strength of the dependence in a stationary time process \( \{X_t\} \). If \( r(k) = \text{Cov}(X_0, X_k), k \geq 0 \), denote the process autocovariance function, then we may generally classify the process as weakly or short-range dependent (SRD) if the covariances decay fast enough (i.e., \( r(k) \to 0 \) as \( k \to \infty \)) so that \( \sum_{k=1}^{\infty} |r(k)| < \infty \) holds. In contrast, strongly or long-range dependent (LRD) processes are characterized by a slow covariance decay, \( r(k) \approx Ck^{-\alpha} \) as \( k \to \infty \) for some \( C > 0 \) and \( 0 < \alpha < 1 \), whereby \( \sum_{k=1}^{\infty} r(k) = \infty \) holds (cf. Beran, 1994). Processes exhibiting long-range dependence (LRD) often have application, for example, in astronomy, hydrology and economics (cf. Beran, 1994; Montanari, 2003; Henry and Zaffaroni, 2003). The behavior of statistical methods can change dramatically between SRD and LRD cases, which complicates the development of appropriate resampling methods. For instance, Lahiri (1993) showed that the block bootstrap, which is generally valid under weak dependence (Künsch, 1989; Liu and Singh, 1993), is invalid for estimating the distribution of a sample mean from a class of long-memory processes.

The dissertation consists of three chapters (i.e., three manuscripts), one for each of the three methods listed above. We begin by considering weakly dependent processes and then transition to works involving LRD, where the approach to handling the time dependence in the methods also transitions as indicated in the following table.

Manuscript 1 considers a new blockwise empirical likelihood (BEL) method for stationary,
Table 1.1 Main themes of each manuscript

<table>
<thead>
<tr>
<th>Manuscript</th>
<th>Method</th>
<th>Methodological Approach</th>
<th>Process Dependence Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EL</td>
<td>Data block-based</td>
<td>SRD</td>
</tr>
<tr>
<td>2</td>
<td>Bootstrap</td>
<td>Data block-based</td>
<td>LRD, but including SRD</td>
</tr>
<tr>
<td>3</td>
<td>Bootstrap</td>
<td>Data (Fourier) transformation</td>
<td>LRD, but including SRD</td>
</tr>
</tbody>
</table>

weakly dependent time processes, called a progressive block empirical likelihood (PBEL). Unlike the standard BEL originally proposed by Kitamura (1997), the PBEL method does not require any block length selections. Because the performance of the standard BEL can depend critically on the block length choice, the PBEL method in contrast enjoys a type of robustness against block selection issues.

Manuscripts 2 and 3 consider different bootstrap problems for stationary, linear time series which could exhibit LRD. Manuscript 2 investigates the large-sample properties of a block bootstrap method for estimating the distribution of sample means. The results establish the validity of the block bootstrap under either LRD or SRD. Additionally, for estimating the variance of a sample mean under LRD, explicit expressions are provided for the large-sample bias and variance of block bootstrap estimators along with formulas for the theoretically optimal block sizes under LRD. Perhaps surprisingly, optimal blocks become shorter in length as the strength of the LRD increases.

Manuscript 3 develops a frequency domain bootstrap (FDB) method for a problem involving Whittle estimation (Whittle, 1953), which is a popular technique for fitting parametric spectral density models to time series. For linear LRD time processes, the resulting Whittle estimators are known to have normal limit laws. However, convergence to normality can be slow under LRD and the finite-sample distributions of Whittle estimators tend to be asymmetric. As a remedy, the FDB method can be used for calibrating confidence intervals in place of a normal approximation. Theoretical results establish the validity of the FDB for approximating the distribution of Whittle estimators for a broad class of time processes and spectral density models under LRD. The same results apply to SRD processes as well. Simulations show that the FDB method has better coverage accuracy than normal approximations under LRD.
References


CHAPTER 2. A PROGRESSIVE BLOCK EMPIRICAL LIKELIHOOD METHOD FOR TIME SERIES

A paper to be submitted to Journal of American Statistical Association

Young Min Kim, Soumendra N. Lahiri, Daniel J. Nordman

Abstract

This paper develops a new blockwise empirical likelihood (BEL) method for stationary, weakly dependent time processes, called the progressive block empirical likelihood (PBEL). In contrast to the standard version of BEL, which uses data blocks of constant length for a given sample size and whose performance can depend crucially on the block length selection, this new approach involves data blocking scheme where blocks increase in length by an arithmetic progression. Consequently, no block length selections are required for the PBEL method, which implies a certain type of robustness for this version of BEL. For inference of smooth functions of the process mean, theoretical results establish the chi-square limit of the log-likelihood ratio based on PBEL, which can be used to calibrate confidence regions. Simulation evidence indicates that the method can perform comparably to the standard BEL in coverage accuracy (when the latter uses a “good” block choice) and can exhibit more stability, all without the need to select a block length.

Key Words: Arithmetic progression; Block bootstrap; Stationarity; Weak Dependence
2.1 Introduction

Empirical likelihood (EL) is a nonparametric methodology, introduced by Owen (1988, 1990), for producing likelihood-type inference without specification of a joint (parametric) distribution for the data. While EL for independent data has been studied in a variety of contexts (cf. Owen, 2001), our interest in this manuscript concerns EL for stationary, weakly dependent time series. In this setting, Kitamura (1997) first introduced the so-called blockwise empirical likelihood (BEL) method, which creates an EL ratio for inference by using data blocks (i.e., consecutive blocks of observations in time) to capture the underlying dependence structure. Such data-blocking has also played an important role in extending bootstrap and subsampling methods to time series, such as the moving block bootstrap of Hall (1985), Künsch (1989) and Liu and Singh (1992), and time subsampling methods of Carlstein (1986), Politis and Romano (1993), and Politis, Romano and Wolf (1999); see Lahiri (2003) for an overview of block resampling methods for time series. The BEL has been shown to apply for time series inference in a wide range of problems (cf. Lin and Zhang, 2001; Bravo, 2005, 2009; Zhang, 2006; Nordman, Sibbertsen and Lahiri, 2007; Chen and Wong, 2009; Nordman, 2009; Wu and Cao, 2011; Lei and Qin, 2011). Much like the moving block bootstrap, the standard implementation of BEL typically involves data blocks of constant length for an observed time series, and therefore requires a corresponding block length selection. However, the performance of BEL often depends critically on the choice of block length. As a small illustration, Table 2.1 shows the effect of block length choice on the resulting BEL confidence intervals (CIs) for the mean of several MA(2) processes, based on a sample size \( n = 75 \) and either overlapping (OL) or non-overlapping (NOL) blocks. One observes that coverage accuracy can change intricately, depending on the block size \( b \) and underlying process, and that optimal block sizes may also vary by process and type of blocking scheme. Hence, the resulting CIs can lead to very different conclusions about the underlying mean parameter even when the block sizes do not differ by much (e.g., 5 or 10). The problem is compounded further by the fact that little is presently known about optimal block selection for the BEL method. As a result, the applicability of the BEL method in practice and the conclusions drawn from it are both subject to the variability
and/or unstability coming from the choice of the block size by the user.

Table 2.1 Coverage percentages for 90% BEL CIs for the mean of processes
\( X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} \), \((Z_t \text{ iid standard normal})\), with \( n = 75 \) and
OL/NOL blocks of size \( b \) (from 4000 simulations). Coverage rates closest to
nominal are indicated with associated \( b \) in (\( \cdot \)).

<table>
<thead>
<tr>
<th>( \theta_1 )</th>
<th>( \theta_2 )</th>
<th>BEL, OL blocks</th>
<th>BEL, NOL blocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.7</td>
<td>96.3 90.1 86.1 90.1 (10)</td>
<td>95.3 81.8 72.6 89.9 (5)</td>
</tr>
<tr>
<td>0.4</td>
<td>0.7</td>
<td>75.2 83.1 78.4 84.2 (6)</td>
<td>74.9 79.5 69.4 83.2 (5)</td>
</tr>
<tr>
<td>0.7</td>
<td>-0.3</td>
<td>90.5 86.8 82.7 89.8 (3)</td>
<td>90.3 82.2 71.5 90.3 (2)</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>77.7 83.9 79.5 84.5 (6)</td>
<td>77.3 79.6 71.2 83.9 (5)</td>
</tr>
</tbody>
</table>

As a remedy, in this paper we propose a **progressive** block empirical likelihood (PBEL) which
uses an alternative data-blocking device and does not require a block length choice. Instead,
for a given sample size, PBEL method uses data blocks which increase in length through an
arithmetic progression. That is, in contrast to data blocks of constant length (for a given
sample size) in the standard version of BEL, the resulting block sizes in the PBEL method are
non-constant and are allowed to grow progressively larger until the given time sample is “used
up.” More specifically, we define the empirical likelihood of a parameter by using successive
(disjoint) blocks of observations of lengths 2, 4, \ldots, 2N such that these cover all of the given
sample (upto a boundary block). As a consequence, given the sample size, the blocks in the
construction of the PBEL are automatically well defined and it does not involve any block
selection steps in the usual sense.

To investigate theoretical properties of the PBEL methodology, we consider the prototypical
problem of inference about the mean parameter and, more generally, inference about a
parameter defined as a smooth function of the mean for a stationary, weakly dependent time
series. We show that under suitable regularity conditions, a version of Wilks’ (1938) theorem
holds in both the problems. Specifically, we show that like the traditional log-likelihood ratio
statistic in a parametric inference problem with independent and identically distributed (iid)
random variables, twice the negative log-EL ratio based on the PBEL method has a limiting
chi-square distribution in each case, which can be used for calibrating confidence regions and
tests. It is well known (cf. Kitamura, 1997) that the log-EL ratio based on the BEL requires an explicit scale adjustment, depending on the choice of the fixed block length, to ensure a chi-square limit. In contrast, the blocking mechanism of the PBEL method automatically takes care of the scaling issue and does not require any explicit adjustments. Simulation evidence indicates that in finite samples, the PBEL can perform comparably to the standard version of BEL when the latter is used with a “good” choice of the block size. Consequently, the PBEL tends to work as well as the BEL at its optimal level, but can have a practical advantage in being robust to issues of block selection.

The rest of manuscript is organized as follows. In Section 2.2, we describe the PBEL method and its associated data blocking scheme. The limiting distributional results are established to justify the method for confidence region estimation. In Section 2.3, we present a simulation study of the coverage accuracy of the PBEL method and provide some comparisons to the standard BEL. Two real data examples are used to illustrate the PBEL methodology in Section 2.4. Section 2.5 provides some concluding remarks and proofs of the main results are deferred to an Appendix (Section 2.6).

### 2.2 Progression block empirical likelihood

#### 2.2.1 Description

Suppose we have an observed data stretch $X_1, \ldots, X_n$ from a strictly stationary process $\{X_t : t \in \mathbb{Z}\}$ taking values in $\mathbb{R}^d$. To explain the PBEL method, we first consider problem of inference on the process mean $EX_t = \mu \in \mathbb{R}^d$. We return to inference on a wider class of “smooth model” parameters in Section 2.2.2.

For comparative purposes, it is initially helpful to recall the standard BEL formulation (Kitamura, 1997). This involves choosing an integer block length $1 \leq b \leq n$ and forming a collection of length $b$ data blocks, which could possibly be maximally overlapping (OL) as given by $\{(X_{i}, \ldots, X_{i+b-1}) : i = 1, \ldots, K\}$ with $K = n - b + 1$, or non-overlapping (NOL) as given by $\{(X_{b(i-1)+1}, \ldots, X_{ib}) : i = 1, \ldots, K\}$ with $K = \lfloor n/b \rfloor$. In either case, all blocks have constant length $b$ for a given sample size $n$. For inference on the mean parameter $\mu$, each block
in the OL collection \(i = 1, \ldots, K\), contributes a centered block sum \(B_{i,\mu} \equiv \sum_{j=1}^{i+b-1} (X_j - \mu)\) (or \(B_{i,\mu} \equiv \sum_{j=b(i-1)+1}^{bi} (X_j - \mu)\) with NOL blocks) for defining a BEL function for \(\mu\) given as

\[
L_{\text{BEL},n}(\mu) = \sup \left\{ \prod_{i=1}^{K} p_i : p_i \geq 0, \sum_{i=1}^{K} p_i = 1, \sum_{i=1}^{K} p_i B_{i,\mu} = 0 \right\}.
\]

(2.1)

and corresponding BEL ratio \(R_{\text{BEL},n}(\mu) = L_{n}(\mu)/K^{-K}\). The function \(L_{\text{BEL},n}(\mu)\) quantifies the plausibility of a value \(\mu\) by maximizing a multinomial likelihood from probabilities \(\{p_i\}_{i=1}^{K}\) assigned to the centered block sums \(B_{i,\mu}\) under a zero-expectation linear constraint. Without the mean constraint in (2.1), the multinomial product is maximized when each \(p_i = 1/K\) (i.e., the empirical distribution on blocks), leading to the ratio \(R_{\text{BEL},n}(\mu)\). When \(0_d\) is in the interior convex hull of \(\{B_{i,\mu}\}_{i=1}^{K}\), then the expansion \(L_{\text{BEL},n}(\mu) = \prod_{i=1}^{K} p_{i,\mu} > 0\) holds with \(p_{i,\mu} = K^{-1}(1 + \lambda_{\text{BEL},n,\mu} B_{i,\mu})^{-1} \in (0, 1), \ i = 1, \ldots, K\), and a Lagrange multiplier \(\lambda_{\text{BEL},n,\mu} \in \mathbb{R}^{d}\) satisfying

\[
0_d = \sum_{i=1}^{K} \frac{B_{i,\mu}}{K(1 + \lambda_{\text{BEL},n,\mu} B_{i,\mu})};
\]

see Owen (1990) for more computational details with EL. Under certain mixing and moment conditions, and if the block size grows with the sample size \(n\) but at a smaller rate (i.e., \(b^{-1} + b^2/n \to 0\) as \(n \to \infty\)), the log-EL ratio of the standard BEL has chi-square limit

\[
-2 \frac{n}{bK} \log R_{\text{BEL},n}(\mu_0) \xrightarrow{d} \chi^2_d,
\]

at the true mean parameter \(\mu_0\). Above \(n/(bK)\) represents a necessary block adjustment factor for the distributional limit with BEL (cf. Kitamura, 1997). As mentioned in Section 2.1, not much is currently known about the best block length \(b\) selections for optimal coverage accuracy with BEL. In practice, one may typically borrow from the block bootstrap literature, where optimal orders for block sizes (across differing problems for distributional estimation) vary in powers of the sample size such as \(n^{1/3}\) or \(n^{1/4}\); see Hall, Horowitz and Jing (1995) and Lahiri (2003, ch. 5) for more details on optimal block selections for time series block bootstrap methods.

We next present the PBEL method, which uses a collection of NOL data blocks given by \(\{(X_{(i-1)i+1}, \ldots, X_{(i+1)}): i = 1, \ldots, N-1\} \cup \{(X_{(N-1)}, \ldots, X_n)\}\), and \(N\) denotes the number of blocks. In this case, the blocks do not have constant length but rather steadily increase
in length through an arithmetic progression (i.e., 2, 4, 6, 8, \ldots) of sizes. In large samples, the number \( N \) of blocks will be close to \( \sqrt{n} \), with the corresponding length of the largest block being approximately \( 2\sqrt{n} \). While the last block could be dropped or defined in many ways without changing the asymptotic results, for a concrete rule, we let \( \ell_1 = \lceil (\sqrt{4n+1} - 1) / 2 \rceil \) and \( \ell_2 = \lfloor (\sqrt{4n+1} - 1) / 2 \rfloor \) to define \( N \) as \( \ell_1 \) if \( n - \ell_1(\ell_1+1) < \ell_2(\ell_2+1) - n \) and \( \ell_2 \) otherwise; this aims at a block number to closely divide the sample into a NOL blocks with a constant progression. To assess the likelihood of a given value of \( \mu \), we then create centered block sums \( S_{i,\mu} = \sum_{j=(i-1)i+1}^{\min\{i(i+1),n\}} (X_j - \mu) \) and define a PBEL function \( L_n(\mu) \) and ratio \( R_n(\mu) \) as

\[
L_n(\mu) = \sup \left\{ \prod_{i=1}^{N} p_i : p_i \geq 0, \sum_{i=1}^{N} p_i = 1, \sum_{i=1}^{N} p_i S_{i,\mu} = 0_d \right\}, \quad R_n(\mu) = \frac{L_n(\mu)}{N^{-N}}. \tag{2.3}
\]

The computation of \( L_n(\mu) \) is the essentially same as for the standard BEL, where \( 0_d \) in the convex hull of \( \{S_{i,\mu}\}_{i=1}^{N} \) implies that \( L_n(\mu) = \prod_{i=1}^{N} N^{-1}(1 + \lambda_{n,\mu} S_{i,\mu})^{-1} > 0 \) for a Lagrange multiplier \( \lambda_{n,\mu} \in \mathbb{R}^d \) satisfying

\[
0_d = \sum_{i=1}^{N} \frac{S_{i,\mu}}{N(1 + \lambda_{n,\mu} S_{i,\mu})}.
\]

The next section establishes that, under mild dependence assumptions, the log-EL ratio from the PBEL method also has a chi-square limit, which can be used for tests or inverted to create confidence regions for \( EX_t = \mu \).

We end this section with some qualifying remarks on the PBEL formulation. The PBEL again does not appeal to a particular block choice. The advantage of the PBEL is, generally, that this approach can perform comparably to the standard BEL when the later employs a good block selection and much better when the standard BEL employs a bad block choice; simulations in Section 2.3 provide some illustration of these features. In this sense, by avoiding the usual block selection issues, the PBEL has a type of stability in its performance. A second important point is that the PBEL uses purely NOL blocks. In contrast, the standard BEL can use OL or NOL blocks for the distributional limit in (2.2). If one attempts to use OL progressively increasing blocks in the PBEL formulation, the resulting asymptotics break down and become highly non-standard. While it is difficult to quantify the asymptotic effect of OL blocks in PBEL approach, simulations indicate that the associated limiting distribution of
the log-PBEL ratio is not chi-square, implying that the simple chi-square calibration property typically associated with EL methods will consequently be lost.

2.2.2 Main distributional results

2.2.2.1 Inference on mean parameters

To provide the limit distribution of the log-PBEL ratio (2.3) for the process mean, we require additional notation. Let \( r(k) = \text{Cov}(X_t, X_{t+k}), k \in \mathbb{Z} \), denote the autocovariance function of the stationary process \( \{X_t\} \) and set \( \Sigma_\infty = \sum_{k=-\infty}^{\infty} r(k) \). Define the process strong mixing coefficient as \( \alpha(k) = \sup\{\left| P(A \cap B) - P(A)P(B) \right| : A \in \mathcal{F}_0, B \in \mathcal{F}_k^\infty \} \), where \( \mathcal{F}_{-\infty}, \mathcal{F}_k^\infty \) are the \( \sigma \)-algebras generated by \( \{X_t : t \leq 0\} \) and \( \{X_t : t \geq k\} \), respectively (cf. Doukhan, 1994). The mixing and moment assumptions in Theorem ?? are mild and also standard for block bootstrap methods as well (Künsch, 1989); these imply conditions of weak time dependence so that, for instance, \( \Sigma_\infty \) is finitely defined.

**Theorem 1.** Suppose that \( E(\|X_t\|^{6+\delta}) < \infty \) and \( \sum_{k=1}^{\infty} k^2 \alpha(k)^{\delta/(\delta+6)} < \infty \) for some \( \delta > 0 \). Let \( EX_t = \mu_0 \in \mathbb{R}^d \) denote the true mean and suppose \( \Sigma_\infty \) is positive definite. Then, as \( n \to \infty \),

\[
-2 \log R_n(\mu) \overset{d}{\rightarrow} \chi_d^2.
\]

**Remark 1:** If \( \Sigma_\infty \) is not positive, the above result holds with \( \text{rank}(\Sigma_\infty) \) degrees of freedom.

From Theorem 1, an approximate \((100 \times \alpha)\%\) confidence region for \( \mu \) is then

\[
\{\mu \in \mathbb{R}^d : -2 \log R_n(\mu) \leq \chi_{d,1-\alpha}^2\},
\]

based on a lower \( 1 - \alpha \) chi-square quantile calibration.

We may make a few additional comments on the limit result above. Due to its blocking mechanism, the PBEL method has a distributional result for the log EL-ratio which closely matches that found for mean inference with iid data (cf. Owen, 1990) where no block adjustments occur (i.e., for iid data, the block size is \( b = 1 \) for which \( n/(bK) = 1 \) in (2.2)). We also note that the proof of Theorem 1 indicates that the Lagrange multiplier \( \lambda_{n,\mu_0} \) in the PBEL method, evaluated at the true mean \( \mu_0 \), exhibits a convergence rate of \( O_p(n^{-1/2}) \). Interestingly,
this is the same rate commonly associated with iid versions of EL (cf. Owen, 1990; Qin and Lawless, 1994). In contrast, the OL or NOL block version of standard BEL has a Lagrange multiplier $\lambda_{\text{BEL},n,\mu_0}$ with a slower convergence rate $O_p(bn^{-1/2})$ where $b \to \infty$ as $n \to \infty$. Hence, despite similar chi-square limits, some asymptotic differences exist in the underlying mechanics of PBEL compared to standard BEL.

### 2.2.2.2 Inference under smooth function model

We next extend the PBEL method, considering inference on a broad class of parameters under the so-called “smooth function model” of Bhattacharya and Ghosh (1978) and Hall (1992). If $E X_t = \mu_0 \in \mathbb{R}^d$ again denotes the true mean of the process, we may also target inference on a vector-valued parameter defined as

$$\theta_0 = H(\mu_0) \in \mathbb{R}^p,$$

based on a smooth function $H(\mu) = (H_1(\mu), \ldots, H_p(\mu))'$ of the mean parameter $\mu$, where $H_i : \mathbb{R}^d \to \mathbb{R}$ for $i = 1, \ldots, p$ and $p \leq d$. This framework permits a wide range of parameters to be considered through appropriate functions, such as sums, differences, products or ratios, involving the $m$-dimensional moment structure (for a fixed $m$) of a time series. For instance, if data arise from a univariate stationary series $U_1, \ldots, U_n$, we can define a multivariate series $X_t$ based on transformations of $(U_t, U_{t+m-1})$ (for a fixed lag $m$) and estimate parameters for the process $\{U_t\}$ based on appropriate functions $H$ of the mean of $X_t$. The autocovariance $\theta$ of $\{U_t\}$ at a lag $m$, for example, can be translated into (2.4) by $X_t = (U_t, U_t U_{t+m})' \in \mathbb{R}^2$ and $H(x_1, x_2) = x_2 - x_1^2$. Künsch (1989) and Lahiri (2003, Ch. 4) provide further examples of smooth function parameters, and Hall and La Scala (1990), and Kitamura (1997) have considered EL for similar parameters based on independent and time series data, respectively.

To frame a result for the parameter $\theta$, define the PBEL ratio

$$R_n(\theta) = \sup \{R_n(\mu) : \mu \in \mathbb{R}, H(\mu) = \theta\}$$

using (2.3). The following can be used to (chi-square) calibrate confidence regions for $\theta$. 

Theorem 2. In addition to the assumptions of Theorem ??, suppose $H(\cdot)$ from (2.4) is continuously differentiable in a neighborhood of $\mu_0$ and that $\nabla_{\mu_0}$ has rank $p \leq d$, where $\nabla_{\mu} \equiv [\partial H_i(\mu)/\partial \mu_j]_{i=1,\ldots,p; j=1,\ldots,d}$ denotes the $p \times d$ matrix of first-order partial derivatives of $H$. Then, at the true parameter $\theta_0 = H(\mu_0)$, as $n \to \infty$

$$-2 \log R_n(\theta_0) \xrightarrow{d} \chi^2_p.$$ 

While we have focused the development of the PBEL method on the problem of inference on the mean of a stationary, weakly dependent time series, with extension to smooth function model parameters, the same method and blocking technique also apply to the framework of general estimating functions with stationary time series, similarly treated by Kitamura (1997) for the standard BEL for mixing time processes and by Qin and Lawless (1994) for iid data. In Section 2.4, we provide some data examples to illustrate the PBEL method for inference under the smooth function model along with an extension to a case of general estimating functions.

The next section examines the PBEL method through numerical studies.

2.3 Numerical studies

This section investigates performance of PBEL method for interval estimation through a simulation study involving weakly dependent time processes that exhibit differing positive or negative correlation structures with varying strengths. In particular, we considered ten real-valued time processes, with autoregressive (AR) or moving average (MA) components, defined as follows.

M1 : AR(1) process with parameter 0.9,
M2 : AR(1) process with parameter $-0.9$,
M3 : AR(1) process with parameter 0.7,
M4 : AR(1) process with parameter $-0.7$,
M5 : AR(2) process with parameters $(0.5, -0.5)$,
M6 : MA(1) process with parameter 0.7,
M7 : MA(2) process with parameter $(0.7, -0.3)$,
M8 : ARMA(2,1) process with AR parameters $(0.5, -0.5)$ and MA parameter 0.7,
M9 : ARMA(1,2) process with AR parameter 0.7 and MA parameters $(0.7, -0.3)$,
M10 : ARMA(2,2) process with AR parameters $(0.5, -0.5)$ and MA parameters $(0.7, -0.3)$. 

In these models, we considered iid, mean-zero innovations from either standard normal, uniform \((−\sqrt{3}, \sqrt{3})\) or \(\chi^2_1\) distributions, which produced qualitatively similar results; hence, in the following, we shall present results from standard normal innovations.

We considered PBEL intervals for the process mean \(\mathbb{E}X_t = 0\) for variety of sample sizes \(n = \{50, 100, 200, 600, 1200\}\) and a nominal coverage level of 90\% (in all except Table 2.2 to follow which gives coverage rates of both 90\% and 95\% CIs based on the PBEL); repeating the simulations with a 95\% level produced similar results in other cases. For comparison, we also included standard BEL intervals \(\{\mu \in \mathbb{R} : -2b^{-1}\log R_{\text{BEL},n}(\mu) \leq \chi^2_{1,1-\alpha}\}\) based on OL blocks (cf. Sec. 2.2.1) and tapered blockwise empirical likelihood (TBEL) intervals with OL blocks. (The TBEL is similar to the standard BEL in construction but uses a trapezoidal taper \(w(\cdot)\) to weight the observations in each length \(b\) block \((w([1-0.5]/b)X_i, \ldots, X_{i+b-1}w(\lfloor b-0.5/b\rfloor),\) where observations at the ends of blocks receive smaller weights; see Nordman (2009) for details). We also considered BEL intervals with NOL blocks, which typically performed slightly worse than the OL block versions and so are not presented in detail here; however, Table 2.1 provides a subset of the results comparing BEL with OL/NOL blocks, showing that the performance of the NOL version is about the same or slightly worse than that of the OL BEL version.

Because the standard BEL and TBEL require block selections, for which optimal choices are unknown, we employed six different block sizes \(b = Cn^{1/3},\ C = \{0.5, 1, 1.5, 2, 3, 5\}\). The block order \(n^{1/3}\) is chosen based on its consideration by Kitamura (1997, p. 2093) for standard BEL. These block selection choices also borrow from the block bootstrap literature, for which it is not uncommon to take a known optimal block order (e.g., \(n^{1/3}\) or \(n^{1/4}\), cf. Ch. 5 of Lahiri, 2003) for the bootstrap and adjust it by a constant factor (often \(C = 1\) or \(2\)) in implementation. For each sample size, model and interval method, we approximated coverage probabilities for the mean parameter based on 4000 simulation runs.

For compactness in presenting the simulation results, Table 2.2 displays the full coverage results for the PBEL method (including coverage for 95\% intervals) and Figure 2.1 shows differences between the nominal 90\% level and the observed coverage probabilities for PBEL/BEL/TBEL methods for sample sizes \(n = \{50, 200, 1200\}\), all block sizes \(b\) (for BEL/TBEL methods), and all ten process models; note that positive differences indicate un-
dercoverage in this figure. Figure 2.1 indicates that coverage performance of PBEL is fairly comparable to that of BEL and TBEL methods in this study, when the later methods employ “good” block sizes. As might be expected, the PBEL is not typically the absolute top performer against all methods considered because BEL approaches employ a variety of tuning parameters (block lengths). On the other hand, Figure 2.1 also shows that BEL and TBEL methods can be very sensitive to the block choice. As a result, these methods can perform much worse than the PBEL with an improper block choice b. This is particularly evident under the model (M1) for strong positive dependence, where the progressively increasing blocks appear generally better. As the sample size increases, performance differences among the methods tend to narrow.

2.4 Data examples

Here we aim to illustrate the PBEL method with two data examples. Section 2.4.1 considers inference in the mean and smooth model parameter settings of Section 2.2.2. In Section 2.4.2, an extension of the PBEL method to general estimating functions is illustrated.

2.4.1 U.S. unemployment rates

Figure 2.2 shows the average annual unemployment rates (given as a percentage of the civilian work force of age 16 years or over) in the U.S. in the years 1948-2011; the data are available from the U.S. Bureau of Labor Statistics. Assuming these data are a realization of a stationary process \( \{X_t\} \) (we provide some justification of this in what follows), we aim here to illustrate interval estimates for the mean \( \text{EX}_t = \mu \) annual unemployment rate as well as for the parameter \( \theta = r(2)/r(1) \), where \( r(k) = \text{Cov}(X_0, X_k) \), which we will show fits into the smooth model framework described in Section 2.2.2.2.

To obtain approximate 95% CIs for the mean parameter, we applied PBEL method as well as the BEL approach with OL or NOL blocks (denoted as BEL and NBEL, respectively) and the TBEL method. Intervals from BEL/NBEL/TBEL methods were computed over block sizes \( b = 2, 4, 8 \), corresponding here to \( Cn^{1/3} \) with \( C = 0.5, 1, 2 \). Table 2.4 displays the resulting intervals. The PBEL interval emerges as a type of compromise between the CIs of other EL methods using block sizes \( b = 4 \) or 8. Figure 2.3 illustrates the shapes of the corresponding
likelihood ratios for PBEL and BEL methods, which show the CIs are fairly symmetric in this case and that the BEL ratio has increasingly sharper tails for large block lengths. This visual illustration of the likelihood ratios is suggestive of how the BEL method can be sensitive to the block length choice.

The parameter \( \theta = r(2)/r(1) \) fits into the smooth function model by defining \( Y_t = (\sum_{i=0}^{2} X_{t+i}/3, \sum_{i=0}^{1} X_{t+i}X_{t+1+i}/2, X_tX_{t+3}) \), for \( t = 1, \ldots, 62 \), and noting that \( \theta = H(\text{E}(Y_t)) \) for \( H(x_1, x_2, x_3) = (x_3 - x_1^2)/(x_2 - x_1^2) \). Applying the EL methods to \( \{Y_t\} \), we obtained approximate 95% CIs for \( \theta \) listed in Table 2.5.

Regarding the unemployment rates in Figure 2.2, model diagnostics indicate that an ARMA(1,1) model provides a reasonable fit for these data. Sample autocovariances are significantly non-zero and decay slowly at small lags, while the partial autocovariance is significantly non-zero at lag 1, suggesting a potential autoregressive component. Model selection criteria (e.g., AICC) support an ARMA(1,1) and residual diagnostics (e.g., sample autocovariances, Ljung-Box statistic) from the fitted model agree with white noise. Under an ARMA(1,1) model, the parameter \( \theta \) here represents the corresponding AR model coefficient. Hence, for comparison to the EL intervals in Table 2.5, we computed approximate 95% CIs for \( \theta \) based on parametric fits in the ARMA(1,1) model by Hannan-Rissanen or (Gaussian) maximum likelihood estimation (cf. Brockwell and Davis, 2002), which were (0.367, 0.827) and (0.434, 0.898), respectively. The PBEL interval agrees with the parametric intervals, though with a slightly elongated lower endpoint. In this case, the PBEL interval appears again to be a blending of the other EL CIs and, as such, this interval turned out to be slightly more asymmetric than the other intervals, as seen in the likelihood plots of Figure 2.4.

2.4.2 Records of hemispheric temperatures

As mentioned at the end of Section 2.2, the PBEL blocking scheme and likelihood formulation also applies to inference with general estimating functions (cf. Qin and Lawless, 1994; Kitamura, 1997) and stationary time series. In this section, we consider a small illustration of estimating functions in a regression setting.

We shall consider a portion of the so-called “global and hemispheric temperature anomaly
time series” available from the Climatic Research Unit (U.K.); see Jones et al. 2011 for more
details. The data, consisting of adjusted monthly temperature averages from 1850-2010, rep-resent a product of combining gridded surface air temperatures from global land station records,as well as marine data records, after correcting for non-climatic (e.g., instrumental) errors. For
scientific reasons, the temperatures are then represented as anomalies (not actual temperatures), meaning that the data represent deviations from a mean temperature computed over a reference period (1961-1990) (cf. Brohan et al. 2006); in computing average monthly temperatures, Jones et al. 1999 discuss how this adjustment reduces technical problems (e.g., due to differing station elevations or reporting discrepancies among countries). See Tingley (2011) for
a statistical development in calculating climate anomalies.

Figure 2.5 shows annual average temperature anomalies for months December-January-February (DJF) and June-July-August (JJA) over the years 1850-2009, in both northern and southern hemispheres; DJF values are means of average temperature anomaly of December of the current year and January and February of the next year. To illustrate inference with general estimating functions, we consider fitting a simple linear model $Y_t = \beta X_t + \epsilon_t$ for predicting DJF temperature anomalies $\{Y_t\}$ from JJA ones $\{X_t\}$ (i.e., predicting winter/summer averages from summer/winter averages, depending on the hemisphere). For this, we consider an estimating function $g(Y_t, X_t, \beta) = X_t(Y_t - \beta X_t)$, supposing $E[g(Y_t, X_t, \beta_0)] = 0$ at the true parameter value. While the series in Figure 2.5 appear non-stationary, it is plausible that the series $Z_t = g(Y_t, X_t, \beta_0)$ is stationary, as plots of $g(Y_t, X_t, \hat{\beta})$, $t = 1850, \ldots, 2009$, using the ordinary least squares (OLS) estimator $\hat{\beta}$, suggest a stationary series with mean zero. From the overall variability explained in this OLS fit, JJA temperature anomalies appear to be better predictors of DJF anomalies in the southern hemisphere (with adjusted R-squared 0.930) than in the northern (with adjusted R-squared 0.623), which should then intuitively be reflected in the precision of interval estimates for $\beta$ by hemisphere. We applied the PBEL method, using $g(Y_t, X_t, \beta)$ to replace $X_t - \mu$ in EL function (2.3), and computed an approximate 90% CI for $\beta$, listed in Table 2.6. Table 2.6 also includes CIs from other blockwise EL methods.

In this estimating function framework, the PBEL intervals also seem to be a mixture of CIs from other blockwise EL methods based on different block sizes, with shorter intervals for
the southern hemisphere as expected. For comparison against the EL intervals, moving block bootstrap CIs for $\beta$ are $(1.017, 1.262)$ in northern or $(0.883, 0.952)$ in southern hemispheres based resampling blocks from $\{(Y_t, X_t)\}$ of size 5 and computing bootstrap percentile intervals for $\beta$ based on bootstrap OLS estimators $\hat{\beta}_n^*$. Among the EL intervals, the PBEL intervals tend to agree fairly well with the block bootstrap intervals, which are included here as a check because the block bootstrap is known to be valid for regression inference with potentially non-stationary data (cf. Fitzenberger, 1997) and uses/resamples data blocks which are fundamentally different than the data blocks used in the blockwise EL methods. That is, the latter use data blocks based on $g(Y_t, X_t, \beta) = X_t(Y_t - \beta X_t)$ to construct an EL function for inference about the regressor slope $\beta$, while the block bootstrap reconstructs versions $(Y_t^*, X_t^*)$ of the original response/regressor series by resampling blocks of paired values $(Y_t, X_t)$ and then computing subsequent bootstrap OLS estimators $\hat{\beta}_n^*$ from the resampled data. The fact that the PBEL intervals are supported by a very form of block resampling provides some evidence that the method is valid and reasonable for these temperature anomaly data in the context of general estimation functions.

2.5 Conclusion remarks

We have introduced alternative version of blockwise empirical likelihood (BEL) that uses a non-standard blocking device, where data blocks do not have constant lengths for a given sample size, but rather increase in length through an arithmetic progression. Hence, no block selections are required and the blocks involved are implicitly well-defined in this progressive block empirical likelihood (PBEL). Under conditions entailing weak dependence for stationary time series, the log-ratio statistic from the PBEL method has been shown to have a chi-square limit distribution, with some features resembling those found in EL with iid data (i.e., behavior of the log-ratio and Lagrange multiplier). While we have focused on inference on process means, or smooth functions of these, the blocking device presented here would remain valid for inference in other blockwise empirical likelihood scenarios involving estimating equations. Simulations have shown that the PBEL performs comparably to BEL when the later is based on a good block choice and often much better when an inadequate block selection is used. Because block
selection is a difficult and theoretically unresolved issue with the standard BEL, this means that the PBEL can have a practical advantage in being robust against against a choice of block size for capturing time dependence.

### 2.6 Proofs of main results

To establish Theorem 1, we require some additional notation along with several initial, technical results stated in Lemma 1 below. Let \( E(X_t) = \mu_0 \in \mathbb{R}^d \) again denote the true process mean and recall \( \Sigma_{\infty} \equiv \sum_{k=-\infty}^{\infty} r(k) \) is the sum of process autocovariances \( r(k) = \text{Cov}(X_0, X_k), \ k \in \mathbb{Z} \), where \( \mathbb{Z} \) denotes the set of integers. Recall \( n \) and \( N \), respectively, denote the sample size and the number of available progressive blocks, which satisfy \( \sqrt{n}/N \to 1 \) as \( n \to \infty \) by construction.

For the progressive (centered) block sums \( S_{i,\mu_0} = \sum_{j=\min(i(i+1)/2),n} S_{i,\mu_0} \), \( i = 1, \ldots, N \), define \( Z_n \equiv \max_{1 \leq i \leq N} \| S_{i,\mu_0} \| \) as well as a block-based estimator \( \hat{\Sigma}_n \equiv n^{-1} \sum_{i=1}^{N} S_{i,\mu_0} S_{i,\mu_0}^T \) of \( \Sigma_{\infty} \).

Let \( \text{CH}_n \) denote the interior of the convex hull of points \( \{ S_{i,\mu_0} \}_{i=1}^{N} \subset \mathbb{R}^d \) and \( 0_d \) denote the zero vector in \( \mathbb{R}^d \). In the following, unless indicated otherwise, limits \( \to \) denote convergence as \( n \to \infty \).

**Lemma 1.** Under the assumptions in Theorem 1, (a) \( n^{1/2}(\bar{X}_n - \mu_0) \overset{d}{\to} Z \sim \text{Normal}(0, \Sigma_\infty) \); (b) \( E(\|S_{i,\mu_0}\|^4) \leq C \) for a constant \( C > 0 \) not depending on \( n \) or \( i = 1, \ldots, N \); (c) \( Z_n = o_P(n^{1/4}) \); (d) \( \| \hat{\Sigma}_n - \Sigma_\infty \| = o_P(1) \); and (e) \( P(0_d \in \text{CH}_n) \to 1 \).

**Proof of Lemma 1.** Part (a) follows from the central limit theorem for mixing processes (cf. Ch. 16.3, Athreya and Lahiri, 2006), while part (b) is a moment bound on sums under Theorem 1 assumptions (cf. sec. 1.4.1, Doukhan, 1994). To show part (c), we use Jensen’s inequality and part (b) to bound

\[
E(Z_n) \leq E \left[ \left( \sum_{i=1}^{N} \| S_{i,\mu_0} \|^4 \right)^{1/4} \right] \leq \left[ \sum_{i=1}^{N} E(\| S_{i,\mu_0} \|^4) \right]^{1/4} = O(N^{3/4}) = o(n^{1/2}),
\]

using \( N/\sqrt{n} \to 1 \); part (c) then follows.

To establish part (d), note that the mixing/moment conditions imply the absolute summa-
bility of covariances and the fourth order cummulants \((cu)\)

\[
a_1 \equiv \sum_{k=1}^{\infty} k \|r(k)\| < \infty, \quad a_2 \equiv \sum_{t_1,t_2,t_3 \in \mathbb{Z}} \|cu(X_0,X_{t_1},X_{t_2},X_{t_3})\| < \infty,
\]

using Davydov’s inequality (cf. Doukhan, 1994, sec. 1.2.2; Athreya and Lahiri, 2006, Ch. 16.2).

Then, for each block \(i = 1, \ldots, N - 1\), we can expand

\[
\frac{1}{2i} E[S_{i,\mu_0}S_{i,\mu_0}^T] = \sum_{k=-2i}^{2i} \left(1 - \frac{|k|}{2i}\right) r(k) = \Sigma_\infty + R_i,
\]

where the remainder satisfies \(\|R_i\| \leq \sum_{|k| > 2i} \|r(k)\| + (2i)^{-1} \sum_{k \in \mathbb{Z}} \|k\| \|r(k)\| \leq i^{-1}a_1\) and

\[
E[S_{N,\mu_0}S_{N,\mu_0}^T] = \Sigma_\infty + O(N) \text{ by Lemma 1(b). From this, we may write}
\]

\[
E[\Sigma_n] = \frac{1}{n} \sum_{i=1}^{N} 2i \frac{E[S_{i,\mu_0}S_{i,\mu_0}^T]}{2i} = \Sigma_\infty \sum_{i=1}^{N} \frac{2i}{n} + O\left(\frac{(N+1)N}{n}\right) = \Sigma_\infty \frac{(N+1)N}{n} + o(1) \rightarrow \Sigma_\infty.
\]

Also, for any fixed \(v \in \mathbb{R}^d\) with \(\|v\| = 1\),

\[
\text{Var}(v^T \Sigma_n v) \leq \frac{1}{n^2} \sum_{i=1}^{N} \|S_{i,\mu_0}\|^4 + \frac{2}{n^2} \sum_{1 \leq i < j \leq N} \text{Cov}[(v^T S_{i,\mu_0})^2,(v^T S_{j,\mu_0})^2] = O(n^{-1/2}),
\]

which follows from bounding both sums above by \(O(N^3/n^2)\), using part (b) for the first sum

and using that \(\text{Cov}[(v^T S_{i,\mu_0})^2,(v^T S_{j,\mu_0})^2], 1 \leq i < j \leq N\), is bounded by

\[
2[\text{Cov}(v^T S_{i,\mu_0},v^T S_{j,\mu_0})]^2 + |cu(v^T S_{i,\mu_0},v^T S_{i,\mu_0},v^T S_{j,\mu_0},v^T S_{j,\mu_0})| \leq 2(a_1)^2 + ia_2
\]

to handle the second sum. Hence, \(\hat{\Sigma}_n\) is MSE-consistent for \(\Sigma_\infty\), proving part (d).

For part (e), fix \(v \in S^d = \{x \in \mathbb{R}^d : \|x\| = 1\}\) in the unit sphere and define a type of subsampling estimator \(\hat{P}_{n,v} = N^{-1} \sum_{i=1}^{N} \mathbb{I}[(2i)^{-1/2}v^T S_{i,\mu_0} < 0]\) of the normal probability \(P(v^TZ < 0), Z \sim N(0, \Sigma_\infty)\), where \(\mathbb{I}[]\) denotes the indicator function. By the distributional convergence in part (a), we have

\[
p_i \equiv P((2i)^{-1/2}v^T S_{i,\mu_0} < 0) \rightarrow P(v^TZ < 0)
\]

as \(i \rightarrow \infty\). Hence, \(E[\hat{P}_{n,v}] = \frac{1}{N} \sum_{i=1}^{N} p_i \rightarrow P(v^TZ < 0)\) as \(n \rightarrow \infty\) as a Cesaro mean. Additionally, Davydov’s inequality with the mixing coefficient imply that \(\text{Var}[\hat{P}_{n,v}] \leq N^{-1}(1 + 4 \sum_{k=1}^{\infty} \alpha(k)^{\delta/(\delta+6)}) = o(1)\), so that \(\hat{P}_{n,v} \xrightarrow{P} P(v^TZ < 0)\) for any \(v \in S^d\).
Then, for any integer \( m \geq 1 \), there exists a finite set \( C_m = \{ v_1, \ldots, v_{k_m} \} \subset S^d \) where open balls of radius \( 1/m \) around \( v \in C_m \) cover \( S^d \), and one may choose \( C_m \subset C_{m+1} \). Since \( \hat{P}_{n,v} \Rightarrow P(v^T Z < 0) \) for \( v \in \bigcup_{m=1}^{\infty} C_m \), the latter being a countable set, for any subsequence \( \{ n_j \} \) of \( \{ n \} \), there exists a further subsequence \( \{ n_k \} \subset \{ n_j \} \) where, \( \hat{P}_{n_k,v} \rightarrow P(v^T Z < 0) \) holds for all \( v \in \bigcup_{m=1}^{\infty} C_m \), almost surely (a.s.). This in turn implies \( \sup_{v \in S^d} |\hat{P}_{n_k,v} - P(v^T Z < 0)| \rightarrow 0 \) a.s., which is equivalent to \( \sup_{v \in S^d} |\hat{P}_{n,v} - P(v^T Z < 0)| = o_P(1) \). The positivity of \( \Sigma_{\infty} \) implies that, for some \( C > 0 \), \( \inf_{v \in S^d} P(v^T Z < 0) > C \) holds (cf. Lemma 2, Owen, 1990), so that \( P(\inf_{v \in S^d} \hat{P}_{n,v} > C/2) \rightarrow 1 \).

Finally, if \( 0_d \not\in CH_{\sigma}^{n} \), then \( \tilde{v}^T S_{i,\mu_0} \geq 0 \) holds for some \( \| \tilde{v} \| = 1 \in \mathbb{R}^d \) and all \( i = 1, \ldots, N \) by the supporting/separating hyperplane theorem, which implies that \( \hat{P}_{n,\tilde{v}} = 0 \). Hence, \( P(0_d \not\in CH_{\sigma}^{n}) \leq P(\inf_{v \in S^d} \hat{P}_{n,v} \leq C/2) \rightarrow 0 \), proving part (e). \( \square \)

### 2.6.1 Proof of Theorem 1

Assuming \( 0_d \in CH_{\sigma}^{n} \) (which happens with arbitrarily large probability as \( n \rightarrow \infty \) by Lemma 1(e)), \( R_n(\mu_0) \) is finitely positive and equals \( R_n(\mu_0) = \prod_{i=1}^{N}(1 + \gamma_i)^{-1} \) (Owen, 1990, p. 100), where \( \gamma_i = S_{i,\mu_0}^T \lambda_n \) and \( \lambda_n \in \mathbb{R}^d \) satisfies

\[
0_d = \frac{1}{n} \sum_{i=1}^{N} S_{i,\mu_0} = (\tilde{X}_n - \mu_0) - \frac{1}{n} \sum_{i=1}^{N} S_{i,\mu_0} S_{i,\mu_0}^T \lambda_n \quad \text{(2.5)}
\]

Writing \( \lambda_n = \| \lambda_n \| v_n \) for some \( v_n \in \mathbb{R}^d \), \( \| v_n \| = 1 \) and then multiplying (2.5) by \( -v_n \), we may obtain \( \| \tilde{X}_n - \mu_0 \| \geq (1 + \| \lambda_n \| Z_n)^{-1} \| \lambda_n \| \| v^T \Sigma v \| \). Applying Lemma 1(a),(c),(d) and letting \( \| \Sigma_{\infty} \|_2 > 0 \) denote the spectral norm of \( \Sigma_{\infty} \), this implies that \( n^{1/2} \| \tilde{X}_n - \mu_0 \| \leq n^{1/2} \| \lambda_n \| \| \Sigma_{\infty} \|_2 + o_P(1) \) holds with arbitrarily large probability as \( n \rightarrow \infty \), or that \( \lambda_n = O_P(n^{-1/2}) \). By Lemma 1(c), we also then have that \( \max_{1 \leq i \leq N} |\gamma_i| \leq Z_n \| \lambda_n \| = o_P(1) \). With probability approaching 1 as \( n \rightarrow \infty \), we may expand (2.5) to produce

\[
\lambda_n = \Sigma_n^{-1}(\tilde{X}_n - \mu_0 + \beta_n), \quad \beta_n \equiv n^{-1} \sum_{i=1}^{N} \gamma_i^2 S_{i,\mu_0}^T / (1 + \gamma_i), \quad \text{(2.6)}
\]

(used Lemma 1(d) above) and bound

\[
\| \beta_n \| \leq B_n \| \lambda_n \|^2 / (1 - \| \lambda_n \| Z_n) = O_P(n^{-3/4}),
\]
using $B_n = n^{-1} \sum_{i=1}^N \|S_{i,\mu_0}\|^3 = O_p(n^{1/2})$ because $E[B_n] \leq n^{-1} C \sum_{i=1}^N i^{3/2} = O(n^{1/4})$ holds by Holder’s inequality and Lemma 1(b).

When $\max_{1 \leq i \leq N} |\gamma_i| \leq Z_n \|\lambda_n\| < 1$, Taylor’s expansion gives $\log(1 + \gamma_i) = \gamma_i - \gamma_i^2/2 + \eta_i$ with $|\eta_i| \leq \|\lambda_n\|^3 \|S_{i,\mu_0}\|^3/(1 - \|\lambda\|Z_n)^3$, $i = 1, \ldots, N$. Note that $\sum_{i=1}^N |\eta_i| \leq \|\lambda_n\|^3 nB_n/(1 - \|\lambda_n\|Z_n)^3 = O_P(n^{3/2})O_P(n^{5/4})O_P(1) = O_P(n^{-1/4})$. Using this with (2.6), $\|\beta_n\| = O_P(n^{-3/4})$,

$$
\sum_{i=1}^N S_{i,\mu_0} = n(\bar{X}_n - \mu_0) \quad \text{and} \quad -2 \log R_n(\mu_0) = 2 \sum_{i=1}^n \log(1 + \gamma_i) = \sum_{i=1}^N (2\gamma_i - \gamma_i^2 + 2\eta_i) \quad \text{for} \quad \gamma_i = S_{i,\mu_0}^T\lambda_n,
$$

we have the following expansion (holding with arbitrarily high probability for large $n$)

$$
-2 \log R_n(\mu_0) = n(\bar{X}_n - \mu_0)^T \hat{\Sigma}_n^{-1}(\bar{X}_n - \mu_0) - n\beta_n^T \hat{\Sigma}_n^{-1} \beta_n + 2 \sum_{i=1}^N \eta_i
$$

$$
= n(\bar{X}_n - \mu_0)^T \hat{\Sigma}_n^{-1}(\bar{X}_n - \mu_0) + O_P(n^{-1/4}).
$$

Lemma 1(a) and (d) with Slutsky’s theorem complete the proof. □

2.6.2 Proof of Theorem 2

We sketch the proof, modifying arguments given in Hall and La Scala (1990, Theorem 1). If we define a set $\mathcal{M}_n \equiv \{\mu \in \mathbb{R}^d : R_n(\mu) \geq R_n(\mu_0)\}$ of mean values with an PBEL ratio at least as great as the true mean $\mu_0$, then it can be shown that $\sup_{\mu \in \mathcal{M}_n} \|\mu - \mu_0\| = O_P(n^{-1/2})$ (following from the arguments below, i.e., for $\mu = \mu_0 + n^{-1/2} \mathcal{Z}^1_w$ with $w \in \mathbb{R}^d$, $-2 \log R_n(\mu)$ has a non-central chi-square limit with non-centrality parameter $\|w\|$). Hence, it suffices to establish a limit distribution based on $R_n, C(\theta_0) = \sup \{R_n(\mu) : H(\mu) = \theta_0, \|\mu - \mu_0\| \leq Cn^{-1/2}\}$ for a fixed constant $C > 0$, as $R_n, C(\theta_0)$ and $R_n(\theta_0)$ will asymptotically match with arbitrarily high probability for large $C$.

Defining a subsampling-type estimator $\hat{P}_{\mu, n, v} = N^{-1} \sum_{i=1}^N I[(2i)^{-1/2} v^T S_{i,\mu_0} < 0]$ for $v \in \mathbb{S}^d$ as in the proof of Lemma 1(e), it can be shown that $\sup_{v \in \mathbb{S}^d, \|\mu - \mu_0\| \leq Cn^{-1/2}} |\hat{P}_{\mu, n, v} - P(v'Z < 0)|$ is bounded by

$$
\sup_{v \in \mathbb{S}^d} \frac{1}{N} \sum_{i=1}^N I[-C(2i/n)^{1/2} \leq (2i)^{-1/2} v^T S_{i,\mu_0} < 0] + \sup_{v \in \mathbb{S}^d} |\hat{P}_{\mu_0, n, v} - P(v'Z < 0)| = o_p(1);
$$
this implies \( P(0_d \text{ is in the interior convex hull of } \{S_{i,\mu}\}_{i=1}^N \text{ for any } \|\mu - \mu_0\| \leq Cn^{-1/2}) \to 1 \) as \( n \to \infty \). With arbitrarily high probability for large \( n \), we may then assume an expansion
\[
\log R_n(\mu) = -\sum_{i=1}^n \log(1 + \gamma_{i,\mu}) \text{ holds for each } \mu \in \mathbb{R}^d, \|\mu - \mu_0\| \leq Cn^{-1/2}, \text{ where } \gamma_{i,\mu} = \lambda_{n,\mu}^T S_{i,\mu} \in (-1, 1) \text{ and } \lambda_{n,\mu} \in \mathbb{R}^d \text{ satisfy }
\]
\[
0_d = (\bar{X}_n - \mu) - \frac{1}{n} \sum_{i=1}^N S_{i,\mu} S_{i,\mu}^T \lambda_{n,\mu}.
\]

analogously to (2.5). Defining \( \bar{Z}_n = \sup_{\|\mu - \mu_0\| \leq Cn^{-1/2}} \max_{1 \leq i \leq n} \|S_{i,\mu}\| \), it follows that \( \bar{Z}_n = o_p(n^{1/2}) \) by Lemma 1(c) and, defining \( \hat{\Sigma}_{n,\mu} = n^{-1} \sum_{i=1}^n S_{i,\mu} S_{i,\mu}^T \), Lemma 1(d) then establishes
\[
\sup_{\|\mu - \mu_0\| \leq Cn^{-1/2}} \|\hat{\Sigma}_{n,\mu} - \Sigma_\infty\| = o_p(1) + O_p(n^{-3/2} \bar{Z}_n N^2) = o_p(1).
\]
As in the proof of Theorem 1, one may then determine \( \sup_{\|\mu - \mu_0\| \leq Cn^{-1/2}} \|\lambda_{n,\mu}\| = O_p(n^{-1/2}) \) and \( \lambda_{n,\mu} = \hat{\Sigma}_{n,\mu}^{-1}(\bar{X}_n - \mu + \beta_{n,\mu}) \), where \( \sup_{\|\mu - \mu_0\| \leq Cn^{-1/2}} \|\beta_{n,\mu}\| = O_p(n^{-3/4}) \). It also similarly follows that
\[
-2 \log R_n(\mu) = n(\bar{X} - \mu)^T \Sigma_\infty^{-1}(\bar{X} - \mu) + E_{n,\mu} \text{ where } \sup_{\|\mu - \mu_0\| \leq Cn^{-1/2}} |E_{n,\mu}| = o_p(1).
\]

Letting \( \nabla_{\mu} \equiv [\partial H_i(\mu)/\partial \mu_j]_{i=1,...,p;j=1,...,d} \) and noting a Taylor expansion in a neighborhood of \( \mu_0 \) gives \( H(\mu_0 + \nu) - \theta_0 = D_\nu \) for \( D_\nu = \int_0^1 \nabla_{\mu_0 + t \nu} dt \to \nabla_{\mu_0} \) as \( ||\nu|| \to 0 \), we may use Lemma 1(a) to write
\[
-2 \log R_{n,C}(\theta_0) = \inf\left\{-2 \log R_n(\mu) : H(\mu) = \theta_0, \|\mu - \mu_0\| \leq Cn^{-1/2}\right\}
\]
\[
= \inf\left\{n(\bar{X} - \mu_0 + \nu)^T \Sigma_\infty^{-1}(\bar{X} - \mu_0 + \nu) : D_\nu \nu = 0_p, \|\nu\| \leq Cn^{-1/2}\right\} + o_p(1)
\]
\[
\xrightarrow{d} \inf\left\{(\Sigma_\infty^{-1/2} Z + \nu)^T (\Sigma_\infty^{-1/2} Z + \nu) : \nu \in \mathbb{R}^d, \nabla_{\mu_0} \Sigma_\infty^{1/2} \nu = 0_p \right\}
\]
\[
= (\Sigma_\infty^{-1/2} Z)^T P_{\Sigma_\infty^{-1/2} \nabla_{\mu_0}^T} (\Sigma_\infty^{-1/2} Z) \xrightarrow{d} \chi_p^2,
\]
where \( P_{\Sigma_\infty^{-1/2} \nabla_{\mu_0}^T} \) denotes the projection matrix for the column space in \( \mathbb{R}^d \) spanned by \( \Sigma_\infty^{1/2} \nabla_{\mu_0}^T \).

Since \( \Sigma_\infty^{-1/2} Z \) is distributed as a vector of \( d \) independent standard normals and \( P_{\Sigma_\infty^{-1/2} \nabla_{\mu_0}^T} \) is idempotent with rank \( p \), the chi-square distributional limit follows.
References


Table 2.2 Empirical coverage probabilities for 90% confidence intervals for the process mean based on BEL, with either overlapping (OL) or non-overlapping (NOL) blocks. Results are presented over three data-generating models, and various block sizes $b$, and sample sizes $n$.

<table>
<thead>
<tr>
<th>Model</th>
<th>Block Size</th>
<th>OL 50</th>
<th>NOL 50</th>
<th>OL 100</th>
<th>NOL 100</th>
<th>OL 200</th>
<th>NOL 200</th>
<th>OL 600</th>
<th>NOL 600</th>
<th>OL 1200</th>
<th>NOL 1200</th>
</tr>
</thead>
<tbody>
<tr>
<td>M3 0.5</td>
<td>$n^{1/3}$</td>
<td>49.40</td>
<td>49.40</td>
<td>61.10</td>
<td>61.00</td>
<td>63.88</td>
<td>63.78</td>
<td>74.38</td>
<td>74.08</td>
<td>77.55</td>
<td>77.63</td>
</tr>
<tr>
<td>M3 1.0</td>
<td>$n^{1/3}$</td>
<td>65.65</td>
<td>65.03</td>
<td>71.75</td>
<td>71.23</td>
<td>77.40</td>
<td>77.48</td>
<td>82.38</td>
<td>82.10</td>
<td>83.95</td>
<td>83.78</td>
</tr>
<tr>
<td>M3 1.5</td>
<td>$n^{1/3}$</td>
<td>70.63</td>
<td>70.53</td>
<td>77.53</td>
<td>76.83</td>
<td>81.63</td>
<td>81.35</td>
<td>85.10</td>
<td>84.88</td>
<td>86.33</td>
<td>86.00</td>
</tr>
<tr>
<td>M3 2.0</td>
<td>$n^{1/3}$</td>
<td>72.58</td>
<td>67.10</td>
<td>79.35</td>
<td>77.83</td>
<td>83.55</td>
<td>82.75</td>
<td>86.65</td>
<td>85.95</td>
<td>87.15</td>
<td>86.70</td>
</tr>
<tr>
<td>M3 3.0</td>
<td>$n^{1/3}$</td>
<td>72.45</td>
<td>69.28</td>
<td>80.03</td>
<td>76.35</td>
<td>84.95</td>
<td>82.88</td>
<td>87.70</td>
<td>86.55</td>
<td>87.73</td>
<td>87.23</td>
</tr>
<tr>
<td>M3 5.0</td>
<td>$n^{1/3}$</td>
<td>65.58</td>
<td>38.55</td>
<td>77.75</td>
<td>73.15</td>
<td>83.93</td>
<td>79.75</td>
<td>88.20</td>
<td>86.10</td>
<td>87.80</td>
<td>86.38</td>
</tr>
<tr>
<td>M7 0.5</td>
<td>$n^{1/3}$</td>
<td>74.28</td>
<td>74.28</td>
<td>84.53</td>
<td>84.58</td>
<td>84.98</td>
<td>84.93</td>
<td>87.40</td>
<td>87.25</td>
<td>87.50</td>
<td>87.35</td>
</tr>
<tr>
<td>M7 1.0</td>
<td>$n^{1/3}$</td>
<td>84.88</td>
<td>83.85</td>
<td>87.33</td>
<td>84.00</td>
<td>88.03</td>
<td>84.83</td>
<td>88.70</td>
<td>84.05</td>
<td>88.48</td>
<td>83.95</td>
</tr>
<tr>
<td>M7 1.5</td>
<td>$n^{1/3}$</td>
<td>84.70</td>
<td>83.10</td>
<td>88.08</td>
<td>86.38</td>
<td>88.53</td>
<td>87.50</td>
<td>89.13</td>
<td>88.80</td>
<td>89.03</td>
<td>88.60</td>
</tr>
<tr>
<td>M7 2.0</td>
<td>$n^{1/3}$</td>
<td>84.03</td>
<td>77.35</td>
<td>88.08</td>
<td>85.28</td>
<td>88.80</td>
<td>87.33</td>
<td>89.33</td>
<td>88.28</td>
<td>89.13</td>
<td>88.20</td>
</tr>
<tr>
<td>M7 3.0</td>
<td>$n^{1/3}$</td>
<td>81.00</td>
<td>74.05</td>
<td>86.70</td>
<td>80.38</td>
<td>88.33</td>
<td>85.28</td>
<td>88.88</td>
<td>87.98</td>
<td>89.08</td>
<td>88.48</td>
</tr>
<tr>
<td>M7 5.0</td>
<td>$n^{1/3}$</td>
<td>73.33</td>
<td>40.10</td>
<td>83.38</td>
<td>75.98</td>
<td>86.20</td>
<td>80.78</td>
<td>87.93</td>
<td>86.28</td>
<td>89.05</td>
<td>86.83</td>
</tr>
<tr>
<td>M9 0.5</td>
<td>$n^{1/3}$</td>
<td>46.05</td>
<td>46.05</td>
<td>61.05</td>
<td>60.88</td>
<td>60.78</td>
<td>60.83</td>
<td>74.28</td>
<td>74.00</td>
<td>75.58</td>
<td>75.50</td>
</tr>
<tr>
<td>M9 1.0</td>
<td>$n^{1/3}$</td>
<td>65.75</td>
<td>65.13</td>
<td>72.13</td>
<td>72.08</td>
<td>75.48</td>
<td>75.40</td>
<td>82.35</td>
<td>82.15</td>
<td>82.50</td>
<td>82.48</td>
</tr>
<tr>
<td>M9 1.5</td>
<td>$n^{1/3}$</td>
<td>72.08</td>
<td>70.98</td>
<td>78.23</td>
<td>77.13</td>
<td>80.58</td>
<td>79.90</td>
<td>85.25</td>
<td>84.70</td>
<td>85.15</td>
<td>84.95</td>
</tr>
<tr>
<td>M9 2.0</td>
<td>$n^{1/3}$</td>
<td>73.88</td>
<td>69.00</td>
<td>79.93</td>
<td>78.20</td>
<td>82.65</td>
<td>82.23</td>
<td>86.25</td>
<td>85.83</td>
<td>86.18</td>
<td>85.88</td>
</tr>
<tr>
<td>M9 3.0</td>
<td>$n^{1/3}$</td>
<td>73.25</td>
<td>69.73</td>
<td>80.28</td>
<td>76.05</td>
<td>83.83</td>
<td>81.45</td>
<td>86.73</td>
<td>86.45</td>
<td>87.33</td>
<td>86.48</td>
</tr>
<tr>
<td>M9 5.0</td>
<td>$n^{1/3}$</td>
<td>66.28</td>
<td>40.10</td>
<td>78.13</td>
<td>73.03</td>
<td>83.03</td>
<td>78.13</td>
<td>86.85</td>
<td>85.48</td>
<td>87.40</td>
<td>86.20</td>
</tr>
</tbody>
</table>
Table 2.3 Coverage probabilities for nominal 90% and 95% progressive block empirical likelihood (PBEL) confidence intervals for the process mean of ten time series processes over variance sample sizes \(n\) (based on 4000 simulations).

<table>
<thead>
<tr>
<th>Model / (n)</th>
<th>90% Coverage Probability (%)</th>
<th>95% Coverage Probability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>M1</td>
<td>53.88</td>
<td>66.23</td>
</tr>
<tr>
<td>M2</td>
<td>90.93</td>
<td>91.53</td>
</tr>
<tr>
<td>M3</td>
<td>71.15</td>
<td>77.65</td>
</tr>
<tr>
<td>M4</td>
<td>87.68</td>
<td>88.18</td>
</tr>
<tr>
<td>M5</td>
<td>76.78</td>
<td>81.83</td>
</tr>
<tr>
<td>M6</td>
<td>80.55</td>
<td>84.68</td>
</tr>
<tr>
<td>M7</td>
<td>79.33</td>
<td>83.73</td>
</tr>
<tr>
<td>M8</td>
<td>77.25</td>
<td>81.50</td>
</tr>
<tr>
<td>M9</td>
<td>71.60</td>
<td>77.83</td>
</tr>
<tr>
<td>M10</td>
<td>77.15</td>
<td>81.33</td>
</tr>
</tbody>
</table>

Table 2.4 Approximate 95% CIs (Lower,Upper) for the mean annual unemployment rate along with point estimates (Est) as maximizers of respective EL functions.

<table>
<thead>
<tr>
<th>Block (b) =</th>
<th>PBEL</th>
<th>NBEL</th>
<th>BEL</th>
<th>TBEL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Est</td>
<td>5.767</td>
<td>5.767</td>
<td>5.767</td>
<td>5.767</td>
</tr>
<tr>
<td>Lower</td>
<td>5.078</td>
<td>5.265</td>
<td>5.182</td>
<td>5.042</td>
</tr>
</tbody>
</table>

Table 2.5 Approximate 95% intervals (Lower,Upper) for the parameter \(\theta = r(2)/r(1)\) (corresponding to the AR coefficient in an ARMA(1,1) model) along with point estimates (Est) as maximizers of respective EL functions.

<table>
<thead>
<tr>
<th>Block (b) =</th>
<th>PBEL</th>
<th>NBEL</th>
<th>BEL</th>
<th>TBEL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Est</td>
<td>0.571</td>
<td>0.571</td>
<td>0.661</td>
<td>0.606</td>
</tr>
<tr>
<td>Lower</td>
<td>0.159</td>
<td>0.179</td>
<td>0.331</td>
<td>0.332</td>
</tr>
<tr>
<td>Upper</td>
<td>0.841</td>
<td>1.059</td>
<td>1.001</td>
<td>0.701</td>
</tr>
</tbody>
</table>
Table 2.6  Estimated regression coefficient and approximate 90% CIs for $\beta$ (in linear prediction of DJF temperature anomalies from JJA records over 1850-2009) with different EL methods and block sizes $b = 5, 10, 15$ (corresponding to $Cn^{1/3}$ for $C = 1, 2, 3$).

<table>
<thead>
<tr>
<th>Block $b =$</th>
<th>PBEL</th>
<th>NBEL</th>
<th>BEL</th>
<th>TBEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Est</td>
<td>1.100</td>
<td>1.100</td>
<td>1.100</td>
<td>1.100</td>
</tr>
<tr>
<td></td>
<td>1.100</td>
<td>1.100</td>
<td>1.100</td>
<td>1.100</td>
</tr>
<tr>
<td>Northern</td>
<td>1.002</td>
<td>0.989</td>
<td>0.992</td>
<td>0.998</td>
</tr>
<tr>
<td>Lower</td>
<td>1.012</td>
<td>1.018</td>
<td>1.027</td>
<td>1.020</td>
</tr>
<tr>
<td>Upper</td>
<td>1.222</td>
<td>1.244</td>
<td>1.293</td>
<td>1.239</td>
</tr>
<tr>
<td></td>
<td>1.219</td>
<td>1.244</td>
<td>1.293</td>
<td>1.239</td>
</tr>
<tr>
<td>Est</td>
<td>0.921</td>
<td>0.921</td>
<td>0.921</td>
<td>0.922</td>
</tr>
<tr>
<td>Southern</td>
<td>0.892</td>
<td>0.888</td>
<td>0.885</td>
<td>0.881</td>
</tr>
<tr>
<td>Lower</td>
<td>0.884</td>
<td>0.876</td>
<td>0.872</td>
<td>0.886</td>
</tr>
<tr>
<td>Upper</td>
<td>0.959</td>
<td>0.961</td>
<td>0.969</td>
<td>0.965</td>
</tr>
</tbody>
</table>

Note: The table is truncated for brevity. Full table available upon request.
Figure 2.1  Plot of differences between the nominal level and actual coverage rates for 90% CIs for the process mean; positive differences indicate undercoverage. Results are presented for three sample sizes \( n \) and ten data-generating models (denoted M1-M10). Coverage differences for the PBEL method are indicated by \( o \); coverage differences for BEL and TBEL methods based on block sizes \( b = Cn^{1/3} \) (\( C = 0.5, 1.0, 1.5, 2.0, 3.0, 5.0 \)) are indicated by “EL1-EL6” and “TEL1-TEL6.”
Figure 2.2  Plot of U.S. annual average unemployment rates from 1948-2011.

Figure 2.3  Plots of EL ratios for the mean $\mu$ annual unemployment rate. Solid horizontal lines in each plot indicate the calibration cut-off for defining 95% CIs.
Figure 2.4  Plots of EL ratios for $\theta = r(2)/r(1)$ from unemployment data. Solid horizontal lines in each plot indicate the calibration cut-off for defining 95% CIs.

Figure 2.5  Plots of average annual seasonal temperature anomalies for northern and southern hemispheres over the years 1850-2009, for seasons/months DJF (−) and JJA (···).
CHAPTER 3. PROPERTIES OF A BLOCK BOOTSTRAP UNDER LONG-RANGE DEPENDENCE

A paper to be published to Sankhya in November, 2011

Young Min Kim and Daniel J. Nordman

Abstract

The block bootstrap has been largely developed for weakly dependent time processes and, in this context, much research has focused on the large-sample properties of block bootstrap inference about sample means. This work validates the block bootstrap for distribution estimation with stationary, linear processes exhibiting strong dependence. For estimating the sample mean’s variance under long-memory, explicit expressions are also provided for the bias and variance of moving and non-overlapping block bootstrap estimators. These differ critically from the weak dependence setting and optimal blocks decrease in size as the strong dependence increases. The findings in distribution and variance estimation are then illustrated using simulation.

Key Words: Block size; Confidence interval; Sample average; Variance estimation

3.1 Introduction

Block bootstrap methods for time series involve resampling data blocks to capture time dependence, which provided a breakthrough in bootstrap formulation following Singh’s (1981)
observation that Efron’s (1979) iid bootstrap (individual data resampling) could be invalid under dependence. While other time series bootstraps have become available, such as the sieve bootstrap (Bühlmann, 1997) and frequency-domain bootstrap (Kreiss and Paparoditis, 2003), the appeal of the block bootstrap has been its general validity over a wide range of time processes. Additionally, the type of block resampling flexibly allows for different block bootstraps, such as the moving block bootstrap (MBB) of Künsch (1989) and Liu and Singh (1992) and the non-overlapping block bootstrap (NBB) of Carlstein (1986) among several others. However, most developments for the block bootstrap have treated only weakly dependent data. One exception is due to Lahiri (1993) who showed that the MBB could fail in approximating sample means for a category of strongly or long-range dependent (LRD) processes generated by transformations of Gaussian series. This finding appears to have largely deflated confidence in the block bootstrap for long-range dependence (LRD).

Our goal here is to establish block bootstrap inference about the sample mean for a different, but practically broad, class of stationary linear processes exhibiting LRD which includes popular models for LRD such as fractional Gaussian processes (Mandelbrot and Van Ness, 1968) and fractional autoregressive integrated moving averages (Adenstedt, 1974; Granger and Joyeux, 1980; and Hosking, 1981). For these processes, we show MBB and NBB methods to be consistent for distribution estimation under mild and flexible conditions entailing LRD. In particular, these conditions permit general filter coefficients for defining the linear process and also allow for weak dependence, establishing the block bootstrap without the more usual mixing assumptions in this case. The former implies that the block bootstrap may be more widely valid under LRD than the sieve bootstrap, which has recently been justified for causal linear LRD series (Kapetanios and Psaradakis, 2006; Poskitt, 2007).

We also develop the large sample properties of block bootstrap variance estimators for the sample mean under LRD. While a great deal of research has focused on this problem in the weak dependence case (Künsch, 1989; Hall, Horowitz and Jing, 1995; Lahiri, 1999; Politis and White, 2004), little has been known about block bootstrap performance under strong dependence. We provide detailed expressions for the bias and variance of MBB and NBB estimators under LRD, and these findings are somewhat surprising. It turns out that, in contrast to weak dependence
(cf. Künsch, 1989), the MBB may not be asymptotically superior to the NBB under LRD and, rather counter-intuitively, optimal blocks for bootstrap variance estimation should decrease in magnitude as the underlying dependence increases.

We end this section by mentioning other resampling works under LRD. Section 3.2 describes the LRD linear processes and provides results validating MBB and NBB estimators of the sample mean’s distribution. Section 3.3 outlines and summarizes a simulation study of these bootstrap estimators and corresponding confidence intervals for the process mean parameter. Section 3.4 gives the theoretical bias, variance and optimal block sizes of MBB and NBB variance estimators and provides some simulation evidence as well. Section 3.5 provides some concluding remarks and proofs of the main results appear in the Appendix.

In addition to sieve bootstrap mentioned above, other bootstrap formulations have been examined under strong dependence. Hidalgo (2003) proposed a frequency-domain bootstrap for estimating regression coefficients in certain regression models with causal linear LRD processes, where the data transformation aims to weaken the dependence structure. Andrews, Lieberman, and Marmer (2006) established error rates for parametric bootstrap estimation with stationary LRD Gaussian processes. For the same LRD, transformed-Gaussian series for which Lahiri (1993) showed the MBB to be invalid, Hall, Jing, and Lahiri (1998) developed a subsampling method for consistently estimating the distribution of the studentized sample mean and Nordman and Lahiri (2005) extended the validity of this approach for stationary linear processes. McElroy and Politis (2007, Sec. 4) also suggest subsampling for estimating the limit distribution of self-normalized sample means for certain infinite-variance, long-memory series. See Lahiri (2003) for a summary of these and other bootstrap methods for time series.

### 3.2 Block bootstrap distribution estimation under LRD

#### 3.2.1 Target processes

To characterize the linear processes targeted for inference, suppose \( \{X_t\} \) is a stationary time series with mean \( \mathbb{E}X_t = \mu \in \mathbb{R} \) constructed as

\[
X_t = \mu + \sum_{j \in \mathbb{Z}} b_j \varepsilon_{t-j}
\]  

(3.1)
where \( \{ \varepsilon_t \} \) are iid variables with mean \( \mathbb{E} \varepsilon_t = 0 \) and variance \( \mathbb{E} \varepsilon_t^2 < \infty \) and the real-valued \( \{ b_j \} \) sequence of constants satisfies \( \sum_{j \in \mathbb{Z}} b_j^2 < \infty \). We may define \( \{ X_t \} \) as exhibiting strong or long-range dependence (LRD) if the covariances \( r(k) = \text{Cov}(X_0, X_k) \) satisfy a slow decay condition

\[
r(k) \sim \sigma^2 k^{-\theta}, \quad k \to \infty
\]

for some \( \theta \in (0, 1) \) and \( \sigma^2 > 0 \); this is one common formulation of strong dependence where the partial covariance sum \( \sum_{k=1}^{n} |r(k)| = O(n^{1-\theta}) \) diverges as \( n \to \infty \) (Beran, 1994, p. 42; Robinson, 1995a, p. 1634). In contrast, weakly dependent series have covariances \( r(k) \) decaying rapidly to zero as the lag \( k \to \infty \) increases so that \( \sum_{k=1}^{\infty} |r(k)| < \infty \) holds. For studying the block bootstrap and drawing connections to the weak dependence case, it will be most convenient to characterize LRD through the behavior of a sample mean’s variance.

Suppose \( X_1, \ldots, X_n \) is an observed time stretch from (3.1) with sample mean \( \bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i \). Setting \( \sigma^2_{n,\theta} \equiv n^\theta \text{Var}(\bar{X}_n) \), the LRD covariances (3.2) imply that

\[
\lim_{n \to \infty} \sigma^2_{n,\theta} = \sigma^2_{\infty,\theta} > 0,
\]

holds for a constant \( \sigma^2_{\infty,\theta} \) depending on \( \theta \in (0, 1) \). That is, under LRD, the variance \( \text{Var}((\bar{X}_n) \) of the sample mean decays at a slower rate \( O(n^{-\theta}) \) as \( n \to \infty \) than the typical \( O(n^{-1}) \) rate under weak dependence. For later reference, note that allowing \( \theta = 1 \) in \( \sigma^2_{n,\theta} \) and (3.3) is appropriate for prescribing sample mean behavior under weak dependence.

The next section describes and justifies block bootstrap estimation of the distribution of the scaled sample mean \( n^{\theta/2}(\bar{X}_n - \mu) \), or a version \( n^{\hat{\theta}/2}(\bar{X}_n - \mu) \) involving an appropriate estimator of the long-memory parameter \( \theta \), which can be used for calibrating confidence intervals for the mean \( \mathbb{E}X_t = \mu \) under strong dependence.

### 3.2.2 Block bootstrap distributional results

Let \( \ell < n \) denote an integer block length and let \( B(i) = (X_i, \ldots, X_{i+\ell-1}) \) denote a data block with starting point \( 1 \leq i \leq n - \ell + 1 \). A block bootstrap “rendition” of the original series \( X_1, \ldots, X_n \) follows by independently resampling \( b = \lfloor n/\ell \rfloor \) blocks, with replacement, from a collection of length \( \ell \) blocks and concatenating these. Resampling from the collection \( \{ B(i) :
may be written as using (3.4). The NBB variance estimator is analogously defined by non-overlapping blocks and
of the MBB estimator of \( \sigma_n \) estimating the distribution of inflation factor ensures that the MBB version \((\theta/\ell, \theta)\)
(\( b\ell \)) of the original sample mean is approximately \( \sigma_n \bar{X} \). Let \( \bar{X}_N^* = \sum_{i=1}^N X_i^*/N \) denote the mean of a
MBB or NBB sample and let \( P_*, E_*, \text{Var}_* \) denote probability, expectation and variance of the bootstrap distribution given the data.

Approximating the distribution of \( n^{\theta/2}(\bar{X}_n - \mu) \) with the bootstrap counterpart \( N^{\theta/2}(\bar{X}_N^* - E_\mu \bar{X}_N^*) \) appears natural and is valid under weak dependence (i.e., setting \( \theta = 1 \)), but wrong under LRD as pointed out by Künsch (1989, Remark 3.2) and Lahiri (1993). Recalling \( b = \lfloor n/\ell \rfloor \), the bootstrap sample mean should be “inflated” by a factor \( b^{(1-\theta)/2} \) under LRD, producing \( b^{(1-\theta)/2} N^{\theta/2} \bar{X}_N^* = (b\theta)^{1/2} \bar{X}_N^* \) as the correct version of \( n^{\theta/2} \bar{X}_n \) (i.e., no inflation under weak dependence \( \theta = 1 \)). Without re-scaling, the MBB sample mean has variance

\[
\text{Var}_*(N^{\theta/2} \bar{X}_N^*) = \frac{b^{\theta-1}}{n-\ell+1} \sum_{i=1}^{n-\ell+1} \ell^\theta (\bar{B}_i - \hat{\mu}_{n,\text{MBB}})^2, \quad \hat{\mu}_{n,\text{MBB}} = \sum_{i=1}^{n-\ell+1} \bar{B}_i/(n-\ell+1) \quad (3.4)
\]

corresponding essentially to a sample variance of block averages \( \bar{B}_i = \sum_{j=i}^{i+\ell-1} X_i/\ell, \ i = 1, \ldots, n-\ell+1 \). That is, each term in bootstrap variance (3.4) estimates the scaled variance \( b^{\theta-1}\ell^\theta \text{Var}(\bar{X}_\ell) \approx b^{(1-\theta)/2} \sigma_{\infty,\theta}^2 \) of a length \( \ell \) sample mean, while the variance \( \sigma_{n,\theta}^2 = \text{Var}(n^{\theta/2} \bar{X}_n) \) of the original sample mean is approximately \( \sigma_{\infty,\theta}^2 \) in large samples from (3.3). Therefore, the inflation factor ensures that the MBB version \((b\theta)^{1/2} \bar{X}_N^* \) has the “right” variance for approximating the distribution of \( n^{\theta/2} \bar{X}_n \) under LRD; the same holds for the NBB. The correct form of the MBB estimator of \( \sigma_{n,\theta}^2 \) then becomes \( \hat{\sigma}_{\ell,\theta,\text{MBB}}^2 = \text{Var}_*((b\theta)^{1/2} \bar{X}_N^*) = b^{1-\theta} \text{Var}_*(N^{\theta/2} \bar{X}_N^*) \) using (3.4). The NBB variance estimator is analogously defined by non-overlapping blocks and may be written as

\[
\hat{\sigma}_{\ell,\theta,\text{NBB}}^2 = \frac{\ell^\theta}{b} \sum_{i=1}^b (\bar{B}_{1+(i-1)\ell} - \hat{\mu}_{n,\text{NBB}})^2, \quad \hat{\mu}_{n,\text{NBB}} = \sum_{i=1}^b \bar{B}_{1+(i-1)\ell}/b.
\]

Theorem 1 establishes the validity of block bootstrap distribution estimation for linear LRD processes, without assuming a specific form for the long-memory covariances. Instead,
LRD is prescribed through the behavior of sample mean’s variance \( (3.3) \) and a condition that covariances between block averages \( \bar{B}_i = \sum_{j=i}^{j+i-1} X_j / \ell, \ i \geq 1 \) are negligible across large lags:

\[
\text{for any } \epsilon \in (0, 1), \quad \lim_{n \to \infty} \max_{n \epsilon \leq i \leq n} |\ell^\theta \text{Cov}(\bar{B}_1, \bar{B}_i)| \to 0 \quad \text{if } \ell^{-1} + \ell/n \to 0; \quad (3.5)
\]

Together \( (3.3) \) and \( (3.5) \) are weaker than (implied by) the covariance formulation \( (3.2) \) of LRD.

**Theorem 1.** Assume that the linear process \( \{X_t\} \) satisfies \( (3.1) \) with \( E\varepsilon_t^2 < \infty \). Suppose, for some \( \theta \in (0, 1] \), that the variance \( \sigma_{n, \theta}^2 \equiv n^\theta \text{Var}(\bar{X}_n) \) satisfies \( (3.3) \) and that \( (3.5) \) and \( \sum_{k=1}^{n} |r(k)| = O(n^{1-\theta}) \) hold. If \( \ell^{-1} + \ell/n \to 0 \) as \( n \to \infty \), then

(i) for either MBB or NBB methods,

\[
\sup_{x \in \mathbb{R}} \left| P_x \left( b^{1/2} \frac{\ell^\theta/2}{\ell} (\bar{X}_n - E\bar{X}_n) \leq x \right) - P(n^{\theta/2}(\bar{X}_n - \mu) \leq x) \right| \xrightarrow{P} 0;
\]

(ii) \( |\hat{\sigma}_n^2 - \sigma_{n, \theta}^2| \xrightarrow{P} 0 \), where \( \hat{\sigma}_n^2 \) denotes either estimator \( \hat{\sigma}_{\ell, \theta, \text{MBB}}^2 \) or \( \hat{\sigma}_{\ell, \theta, \text{NBB}}^2 \) of \( \sigma_{n, \theta}^2 \).

(iii) Both (i)-(ii) hold upon replacing \( \theta \) with an estimator \( \hat{\theta} \), based on \( X_1, \ldots, X_n \), which satisfies \( |\hat{\theta} - \theta| \log n \xrightarrow{P} 0 \).

While Lahiri (1993) proved that the MBB can indeed fail for certain LRD processes, Theorem 1 shows that the block bootstrap remains consistent over a practical class of linear LRD processes, which need not be causal \( (b_j = 0 \text{ for } j < 0 \text{ in } (3.1)) \) as assumed for the sieve bootstrap (cf. Poskitt, 2007).

As noted by Lahiri (1993), the MBB fails for the LRD processes considered in that work precisely whenever these produce non-normal limits for \( \bar{X}_n \) (i.e., by non-linear transformations of Gaussian processes). The difference here is that, for linear processes \( (3.1) \) where \( (3.3) \) holds, the scaled sample mean has a normal limit,

\[
n^{\theta/2}(\bar{X}_n - \mu) \overset{d}{\to} N(0, \sigma_{\infty, \theta}^2) \quad (3.6)
\]
as \( n \to \infty \) (cf. Davydov, 1970; Ibragimov and Linnik, 1971, Theorem 18.6.5), and the bootstrap sample mean, as a sum of independently resampled block averages, has a large-sample normal distribution with a variance matching that of \( n^{\theta/2}(\bar{X}_n - \mu) \) by Theorem 1(ii). In the next section, we provide a numerical study of block bootstrap distribution estimation and block sizes. Section 3.4 then re-visits the topic of variance estimation in more detail.
Remark 1. Theorem 1 allows $\theta = 1$ as a characterization of weak dependence, where a positive limit for $n \text{Var}(\bar{X}_n)$ in (3.3) is supposed along with summable covariances $\sum_{k=1}^{\infty} |r(k)| < \infty$; the latter automatically implies (3.5) for $\theta = 1$. By phrasing Theorem 1 through the variance behavior of sample averages (i.e., covariance sums), the block bootstrap distributional result holds naturally between strong and weak dependence settings and requires no mixing conditions for weak dependence, as often assumed for the MBB (cf. Lahiri, 2003, Ch. 4).

Remark 2. If the sample mean of a stationary process satisfies (3.3) and (3.6) for some $\theta \in (0, 1]$ and if the bootstrap variance estimator is consistent $|\hat{\sigma}_{\ell, \theta}^2 - \sigma_{n, \theta}^2| \xrightarrow{p} 0$ and the expected bootstrap mean $E_\ast \bar{X}_N^\ast$ converges to the process mean $\mu$ such that $\ell^{\theta/2} |E_\ast \bar{X}_N^\ast - \mu| \xrightarrow{p} 0$ along a block sequence $\ell$, then the bootstrap result in Theorem 1(i) holds with no further assumptions on the time process or block sizes.

Remark 3. Several estimators of the long-memory parameter $\theta$ exist and satisfy the Theorem 1(iii) condition under mild assumptions. One common choice $\hat{\theta}_n = 1 - 2\hat{d}_{m,n}$ may be based on log-periodogram regression (Geweke and Porter-Hudak, 1983) against Fourier frequencies $\lambda_i = 2\pi i/n$, $i = 1, \ldots, m$, where $\hat{d}_{m,n} = \sum_{i=1}^{m} (g_i - \bar{g}) \log I_n(\lambda_i)/\sum_{i=1}^{m} (g_i - \bar{g})^2$ with $\bar{g} = \sum_{i=1}^{m} g_i/m$, $g_i = -2 \log |1 - e^{-\lambda_i\sqrt{-1}}|$ and $I_n(\lambda_i) = |\sum_{t=1}^{n} X_t e^{-t\lambda_i \sqrt{-1}}|^2/(2\pi n)$. A bandwidth $m = O(n^{4/5})$ is optimal for mean-squared error in estimation so that $|\hat{\theta}_n - \theta| = O_p(n^{-2/5})$ in probability (Robinson, 1995b; Hurvich, Deo and Brodsky, 1998). Local Whittle estimation is another option (Künsch, 1987), having similar consistency and optimal bandwidth orders (Robinson, 1995a; Andrews and Sun, 2004). Moulines and Soulier (2003) describe these and other estimators of $\theta$. 
3.3 Simulation study of bootstrap confidence intervals

3.3.1 Simulation design

This section investigates the performance of the block bootstrap distributional approximations for the calibration of confidence intervals under LRD. For comparison to the block bootstrap, we also examine an autoregressive sieve bootstrap for stationary linear time processes, proposed by Kreiss (1992) and Bühlmann (1997) under weak dependence and extended to long-memory series by Kapetanios and Psaradakis (2006) and Poskitt (2007). We briefly sketch this bootstrap method for completeness. A realization \(X_1, \ldots, X_n\) from the original series (having mean \(E X_t = \mu\) and covariances \(r(k), k \geq 0\)) is first approximated by a stationary autoregressive process \(\{Y_t\}\) of order \(p \equiv p_n\), fit to minimize the distance \(E[(X_t - \mu) - \sum_{i=1}^{p} \beta_i(X_{t-i} - \mu)]^2\) over \((\beta_1, \ldots, \beta_p)\). The series \(\{Y_t\}\) is then defined as

\[Y_t = \mu + \sum_{i=1}^{p} \beta_i(Y_{t-i} - \mu) + \tilde{\varepsilon}_t\]

where \(\beta \equiv (\beta_1, \ldots, \beta_p)' = \Gamma_p^{-1}r_p\) are the coefficients of the best-linear predictor of \(X_t - \mu\) in terms of \((X_{t-1} - \mu, \ldots, X_{t-p} - \mu)\), \(\{\tilde{\varepsilon}_t\}\) are iid variables in the approximation having mean \(E \tilde{\varepsilon}_t = 0\) and variance \(E \tilde{\varepsilon}_t^2 = r(0) - \beta'\Gamma_p\beta\), and \(\Gamma_p\) denotes the \(p \times p\) matrix with \(r(i - j)\) as the \((i, j)\)th entry with \(r_p = (r(1), \ldots, r(p))'\). The main idea to produce a length \(n\) bootstrap rendition of \(\{Y_t\}\) to mimic the joint distribution of \(\{X_1, \ldots, X_n\}\), where the initial (sieve) approximation through \(\{Y_t\}\) should intuitively improve as the order \(p\) increases suitably with the sample size \(n\). Based on observed data \(X_1, \ldots, X_n\), let \((\hat{\beta}_1n, \ldots, \hat{\beta}_pn)\) denote estimates from solving the sample version of Yule-Walker equations (Brockwell and Davis, 1991, Ch. 8) and define residuals \(\hat{\varepsilon}_t = (X_t - \bar{X}_n) - \sum_{i=1}^{p} \hat{\beta}_i n(X_{t-i} - \bar{X}_n), p + 1 \leq t \leq n\). The sieve bootstrap observations are generated recursively as

\[(Y_t^* - \bar{X}_n) = \sum_{i=1}^{p} (Y_{t-i}^* - \bar{X}_n) + \tilde{\varepsilon}_t^*, \quad t \geq p + 1\]

where each \(\tilde{\varepsilon}_t^*\) is resampled independently from the centered residuals \(\{\hat{\varepsilon}_t - (n-p)^{-1}\sum_{i=p+1}^{n} \hat{\varepsilon}_i : p + 1 \leq t \leq n\} \) and we define \(Y_1^* = \cdots = Y_p^* = \bar{X}_n\). Running the recursion \(Y_1^*, \ldots, Y_{n+q}\) with a burn-in of length \(q \geq 1\), the last \(n\) observations are taken to produce a sieve bootstrap
sample mean \( n^{\theta/2}(\bar{X}_n^* - \bar{X}_n) \), which validly approximates the distribution of \( n^{\theta/2}(\bar{X}_n - \mu) \) for linear, LRD processes under certain conditions (cf. Theorem 2, Kapetanios and Psaradakis, 2006). Under weak dependence, results in Bühlmann (1997) and Choi and Hall (2000) indicate that the sieve bootstrap can provide more accurate distribution estimation than the block bootstrap, but requires stronger assumptions on the underlying process (e.g., causal linear and often invertible) in comparison.

In the following simulation study, we consider data from (mean-zero) fractional autoregressive integrated moving average (FARIMA\((0,d,0)\)) processes

\[
X_t = \sum_{j=0}^{\infty} \frac{\Gamma(j + d)}{\Gamma(j + 1)\Gamma(d)} \varepsilon_{t-j}\tag{3.7}
\]

of Granger and Joyeux (1980) and Hosking (1981), where \( \Gamma(\cdot) \) denotes the gamma function, \( \{\varepsilon_t\} \) are iid mean-zero variables, and the long-memory parameter is given by \( \theta = 1 - 2d \in (0,1) \). The resulting process covariances satisfy (3.2) with \( \sigma = \Gamma(1 - 2d)/[\Gamma(1 - d)\Gamma(d)] \) in this case (Beran, 1994, p. 64) and both sieve and block bootstrap methods are applicable. We shall consider \( \{\varepsilon_t\} \) as iid standard normals and focus on a variety of long-memory parameters \( \theta \in \{0.1, 0.3, 0.5, 0.7, 0.9\} \) and sample sizes \( n \in \{250, 500, 1000, 5000\} \), which are expected to more critically impact resampling performance than the innovation type or a further filter (cf. Nordman and Lahiri, 2005).

For each FARIMA process and sample size, we computed 90\% confidence intervals for the process mean \( \mu = 0 \) using the sample mean \( \bar{X}_n \pm n^{-\theta/2}q_{0.9}^* \) and bootstrap quantile \( q_{0.9}^* \) defined as \( P_*(b^{1/2}(\theta/2)|\bar{X}_N^* - E_\theta \bar{X}_N^*| \leq q_{0.9}^*) = 0.9 \) for the block bootstrap (NBB or MBB) and \( P_*(n^{\theta/2}|\bar{X}_n^* - \bar{X}_n| \leq q_{0.9}^*) = 0.9 \) for the sieve bootstrap. The quantile \( q_{0.9}^* \) was approximated by 1000 bootstrap sample mean renditions for each data simulation. Bootstrap intervals were also computed without knowledge of the long-memory parameter, by replacing \( \theta \) with an estimate \( \hat{\theta}_n = 1 - 2d_{m,n} \) based on log-periodogram regression (see Remark 2) with two bandwidths \( m = Cn^{1/5}, C \in \{1/2, 1\} \) of optimal order (Hurvich, Deo and Brodsky, 1998).

For the MBB and NBB methods, block lengths were chosen as \( \ell = Cn^{1/2}, C \in \{1/2, 1, 2\} \). Hall, Jing and Lahiri (1998) proposed blocks of this form for a subsampling method with LRD, under the intuition that, to capture data dependence under long-memory, blocks should have
larger order than size $O(n^\kappa)$ blocks, $\kappa \leq 1/3$, known to be optimal for block bootstrap estimation under weak dependence (Hall, Horowitz and Jing, 1995; Lahiri, 2003, Ch. 5). Additionally, results in Section 3.4 show that, at least for block bootstrap variance estimation under LRD, optimal blocks have order $O(n^{\min\{\theta, 1/(3-2\theta)\}})$ with powers increasing from 0 to 1 in $\theta \in (0, 1)$ (with an exception at $\theta = 0.5$ where the order is $(n/\log n)^{1/2}$, not $n^{1/2}$). While optimal blocks for variance and distribution may not coincide, the order $O(n^{1/2})$ seems to emerge as a “middle of the road” block choice under long-memory.

The sieve bootstrap used a burn-in of 300 along with either a fixed $p_n = \lfloor 2(\log n)^2 \rfloor$ or an estimated order $p_n = \lfloor 10 + 2\hat{h} \rfloor$, where $\hat{h}$ minimized an information criterion from Bühlmann (1997) and Poskitt (2007) over $1 \leq h \leq 10\log_{10} n$. The fixed order agrees with conditions set by Kapetanios and Psaradakis (2006) and Poskitt (2007) under LRD and the estimated formulation of $p_n$ is along lines suggested by Choi and Hall (2000).

### 3.3.2 Summary of results

Table 3.1 provides empirical coverages and average lengths for 90% MBB intervals, where the long-memory parameter $\theta$ was taken as given or estimated with log-periodogram regression (only results for bandwidth $m = n^{4/5}$ are shown as performance with $m = n^{4/5}/2$ was similar). For block lengths $\ell$ as scalar multiples of $n^{1/2}$, the coverage accuracies improve as the dependence decreases $\theta \uparrow 1$ or as sample sizes increase. For comparison against these blocks, a block choice $\ell_\theta = n^{\min\{\theta, 1/(3-2\theta)\}}$ was also included (i.e., optimal order for MBB variance estimation in Section 3.4). To interpret the results, note the power $\min\{\theta, 1/(3-2\theta)\}$ of $n$ defining $\ell_\theta$ is less or greater than $1/2$ depending on $\theta < 0.5$ or $\theta > 0.5$.

For $\theta = 0.1$ or $0.3$, $\ell_\theta$ appeared to perform better than $\ell = n^{1/2}/2$ or $n^{1/2}$, particularly for $\theta = 0.1$. This suggests, perhaps counter-intuitively, that relatively small blocks are favorable under stronger dependence. The MBB coverages in Table 3.1 indicate that blocks should become longer as the dependence weakens ($\theta$ increases), but the optimal block sizes $\ell_\theta$ for variance estimation were seemingly too large in the weaker dependence cases $\theta = 0.7$ and $\theta = 0.9$ where MBB performance deteriorated using $\ell_\theta$. In the extreme case $\theta = 0.1$, coverages were particularly poor and exhibited slow convergence, where the average lengths of MBB
intervals did not shrink over the increasing sample sizes in Table 3.1 (these decrease for sample sizes over 10,000).

Throughout Table 3.2, differences in coverages between estimated and known value of $\theta$ were relatively small, with the largest discrepancies occurring under the strongest dependence $\theta = 0.1$. We note that, to automate the simulation, regression estimates of the long-memory parameter were set to 0 or 1 whenever $\hat{\theta} < 0$ or $\hat{\theta} > 1$ occurred. This behavior most often occurred for the near-boundary values of $\theta = 0.1$ or $0.9$ and small sample sizes (with instances as high as 16% of simulations for $n = 250$). However, removing such cases from the simulations also did not significantly change empirical coverages.

Coverage probabilities for 90% sieve bootstrap intervals for the mean are presented in Table 3.2 which were typically well below nominal, but improving with increasing sample size or weakening dependence (i.e., larger $\theta$). Both fixed and estimated autoregressive orders performed similarly, where estimated orders in Table 3.2 increased as the dependence increased. Compared to the MBB, the sieve bootstrap appeared to exhibit extreme bias in distribution estimation, as illustrated in Figure 3.1. This figure displays the distribution of the sample mean $n^{\theta/2}(\bar{X}_n - \mu)$ for $n = 500$ as well as sieve/MBB bootstrap estimates based on five different data simulations. While both bootstraps exhibit clear bias in the LRD case $\theta = 0.1$, the sieve bootstrap more severely underestimates quantiles in the target distribution. The explanation for this may lie in using the “standard” formulation of the sieve bootstrap based Yule-Walker estimates to fit autoregressive coefficients (cf. Bühlmann, 1997; Poskitt, 2007; Kapetanios and Psaradakis, 2006). Poskitt (2007, Sec. 5.1) mentions potential bias in Yule-Walker estimates under strong dependence and our simulations seem to illustrate this effect for sieve bootstrap estimation (for the sample mean at least). We also re-ran our simulations using Burg’s algorithm to fit autoregressive coefficients, but the coverage probabilities did not essentially change for the processes and long-memory parameters considered. Hence, this numerical study does not suggest what the best implementation of the sieve bootstrap may be, but further autoregressive fitting techniques may potentially be used (cf. Poskitt 2007, Sec. 5.1). Additionally, as the dependence weakens ($\theta \to 1$) and moves into short-memory, the performance of sieve bootstrap does catch up to (and perhaps even surpasses) that of the MBB.
as illustrated in Figure 3.2.

We do not report NBB empirical coverages since performance differences between MBB and NBB confidence intervals were small. In fact, coverages for both methods were within 2% for block lengths \( \ell = Cn^{1/2} \), \( C \in \{1/2, 1, 2\} \) with all \( \theta \) values (estimated with either bandwidth or not) and sample sizes in Table 3.1. Discrepancies in coverage only emerged for relatively large blocks \( \ell_\theta \) when \( \theta = 0.7 \) or 0.9 where, with much fewer blocks available, the NBB exhibited more extreme undercoverage than the MBB (about 10% less than MBB coverages in Table 3.1). We add that, although theoretical optimal block sizes are unknown for distribution estimation under LRD and most likely vary with both the underlying dependence parameter \( \theta \) and type of block bootstrap, the middle of the road block order \( O(n^{1/2}) \) appeared to perform adequately across both MBB/NBB methods and dependence structures of varying strength, except under extremely strong dependence (\( \theta = 0.1 \)).

### 3.4 Block bootstrap variance estimation under LRD

As described in Section 3.2, the behavior of the block bootstrap variance estimator plays a role in formulating this bootstrap under LRD (i.e., re-scaling) and, for the linear LRD processes considered here, Theorem 1 established the consistency of such estimators for the sample mean’s variance \( \sigma^2_{n,\theta} \equiv n^\theta \text{Var}(\bar{X}_n) \). Section 3.4.1 goes a step further in providing expressions for the large sample bias and variance of MBB and NBB estimators of \( \sigma^2_{n,\theta} \) under LRD. The bias terms differ largely from the more frequently studied weak dependence case, and there is some interesting continuity in the relative variances of MBB and NBB estimators between weak and strong dependence settings. Our exposition in this section is mostly of theoretical interest, but was originally motivated to understand one aspect in the block bootstrap appearing in the simulations of Section 3.2.2. Namely, why block bootstrap approximations seemed to, rather non-intuitively, improve under the strongest forms of dependence (small \( \theta \)) by employing relatively short block sizes (e.g., compare Table 3.1 values for \( \ell = \ell_\theta \) against larger \( \ell = n^{1/2} \) when \( \theta = 0.1 \)). Section 3.4.2 provides expressions for optimal block sizes for MBB and NBB variance estimation, which support this behavior.
3.4.1 Large-sample bias and variance properties

Assuming \( \theta \in (0, 1) \) is known, consider estimating \( \sigma_{n, \theta}^2 \equiv n^\theta \text{Var}(\bar{X}_n) \) using a block bootstrap estimator \( \hat{\sigma}_{\ell, \theta, MBB}^2 \) or \( \hat{\sigma}_{\ell, \theta, NBB}^2 \), as defined Section 3.2.2. Bias expansions require a more detailed form of the LRD covariances than (3.2) and we suppose that

\[
r(k) = \sigma^2 k^{-\theta} + r_1(k), \quad k > 0, \tag{3.8}
\]

holds for some \( \theta \in (0, 1) \) and \( \sigma^2 > 0 \), where \( r(0) = r_1(0) \) and \( \sum_{k=0}^{\infty} |r_1(k)| < \infty \). This covariance form matches one of Giraitis, Robinson and Surgailis (1999, p. 5) who examined variance-type estimators of the long-memory parameter \( \theta \). As those authors note, (3.8) holds for fractional autoregressive integrated moving average and fractional noise models of long-memory. As a function of \( \theta \in (0, 1) \) and the block bootstrap type, define also a proportionality constant

\[
V_\theta \equiv \begin{cases} 
1 + \frac{(2 - \theta)^2(2\theta^2 + 3\theta - 1)}{4(1 - 2\theta)(3 - 2\theta)} - \frac{\Gamma^2(3 - \theta)}{\Gamma(4 - 2\theta)} & \text{if } 0 < \theta < 1/2, \text{ MBB or NBB} \\
9/32 & \text{if } \theta = 1/2, \text{ MBB or NBB} \\
\sum_{x=-\infty}^{\infty} g_\theta^2(x) & \text{if } 1/2 < \theta < 1, \text{ NBB} \\
\int_{-\infty}^{\infty} g_\theta^2(x)dx & \text{if } 1/2 < \theta < 1, \text{ MBB}
\end{cases}
\]

where \( \Gamma(\cdot) \) denotes the gamma function and \( g_\theta(x) \equiv (|x + 1|^{2-\theta} - 2|x|^{2-\theta} + |x - 1|^{2-\theta})/2, x \in \mathbb{R} \). In the definition of \( V_\theta \), note \( g_\theta^2(x) \) is summable/integrable when \( \theta \in (1/2, 1) \) since \( g_\theta(x) \sim (2 - \theta)(1 - \theta)x^{-\theta}/2 \) as \( x \to \infty \).

In Theorem 2, the bias and variance of block bootstrap variance estimators depend on the limiting variance of the sample mean given by \( \sigma_{\infty, \theta}^2 = 2\sigma^2/\{(1 - \theta)(2 - \theta)\} \) in (3.3), under the LRD covariances (3.2) or (3.8) (cf. Beran, 1994, p. 45).

**Theorem 2.** Let \( \hat{\sigma}_{\ell, \theta}^2 \) denote either \( \hat{\sigma}_{\ell, \theta, MBB}^2 \) or \( \hat{\sigma}_{\ell, \theta, NBB}^2 \), \( \theta \in (0, 1) \). If (3.1) and (3.8) hold and \( \ell^{-1} + \ell/n \to 0 \) as \( n \to \infty \), then

(i) the bias of \( \hat{\sigma}_{\ell, \theta}^2 \) in estimating \( \sigma_{n, \theta}^2 = n^\theta \text{Var}(\bar{X}_n) \) is given by

\[
E\hat{\sigma}_{\ell, \theta}^2 - \sigma_{n, \theta}^2 = B_\theta \frac{\ell_1 - \theta}{\ell_1 - \theta}(1 + o(1)) - \sigma_{\infty, \theta}^2 \frac{\ell_1 - \theta}{n^{\theta}}(1 + o(1))
\]
assuming $B_\theta \neq 0$, for $B_\theta \equiv \sum_{k \in \mathbb{Z}} r_1(k) - 2\sigma^2 I_\theta$, $I_\theta \equiv \lim_{\ell \to \infty} (\int_0^\ell x^{-\theta} dx - \sum_{x=1}^\ell x^{-\theta}) \in (0, \infty)$.

(ii) if $E \varepsilon^6_t < \infty$ holds additionally in (3.1) and $\tilde{V}_\theta \equiv 2\sigma^4_{\infty, \theta} V_\theta$, the variance of $\hat{\sigma}^2_{\ell, \theta}$ is given by

$$\text{Var}(\hat{\sigma}^2_{\ell, \theta}) = \begin{cases} 
\tilde{V}_\theta \left( \frac{\ell}{n} \right)^{2\theta} (1 + o(1)) & \text{if } 0 < \theta < 1/2, \\
\tilde{V}_\theta \frac{\ell}{n} \log \left( \frac{n}{\ell} \right) (1 + o(1)) & \text{if } \theta = 1/2, \\
\tilde{V}_\theta \frac{\ell}{n} (1 + o(1)) & \text{if } 1/2 < \theta < 1.
\end{cases}$$

**Remark 4.** As part of a technique for estimating $\theta$ under LRD, Giraitis, Robinson and Surgailis (1999) derived variance and bias expansions for a block-based estimator, which appears to resemble the NBB estimator presented here. In contrast to their results, Theorem 2 does not require Gaussian processes, has exact form for the proportionality constants $V_\theta$ (compared to less explicit moments of limiting stochastic processes), and also applies to the MBB estimator. However, the variance orders in Theorem 2 do match their findings.

The long-memory parameter $\theta \in (0, 1)$ determines the orders of both the bias and variance of bootstrap estimators. By Theorem 2(i), both MBB and NBB estimators have the same principal bias under LRD; this property holds as well under weak dependence but with a very different bias expression. Under weak dependence $\sum_{k=0}^\infty |r(k)| < \infty$, we may substitute $\theta = 1$ to obtain the proper bootstrap estimator $\hat{\sigma}^2_{\ell, \theta=1}$ of $\sigma^2_{n, \theta=1} = \text{Var}(n^{1/2} \bar{X}_n)$ (i.e., defined with $\theta = 1$ in (3.3)) which has bias given by

$$E\hat{\sigma}^2_{\ell, \theta=1} - \sigma^2_{n, \theta=1} = \frac{B_1}{\ell} (1 + o(1)) - \sigma^2_{\infty, 1} \frac{\ell}{n} (1 + o(1)),$$

for $B_1 \equiv \sum_{k=-\infty}^\infty |k|r(k)$ and $\sigma^2_{\infty, 1} \equiv \sum_{k=-\infty}^\infty r(k) > 0$ denoting limiting variance of $n^{1/2} \bar{X}_n$ in this case (cf. Lahiri, 1999). Note that the MBB/NBB bias under weak dependence will not seamlessly follow by setting $\theta = 1$ in Theorem 2 due to an abrupt change in the form of the first bias term in Theorem 2(i).

To interpret the variance results in Theorem 2, we recall expressions for the variance of block bootstrap estimators under weak dependence given by

$$\text{Var}(\hat{\sigma}^2_{\ell, \theta=1, MBB}) = \frac{2}{3} \text{Var}(\hat{\sigma}^2_{\ell, 1, NBB})(1 + o(1)), \quad \text{Var}(\hat{\sigma}^2_{\ell, 1, NBB}) = 2\sigma^4_{\infty, 1} \frac{\ell}{n} (1 + o(1)),$$
under appropriate mixing/moment conditions (Künsch, 1989; Lahiri, 2003, Ch. 5). That is, under weak dependence and due to the larger block collection used in MBB resampling, the variance of the MBB estimator is well-known to be $2/3$ that of the NBB version. In contrast, Theorem 2 indicates that large-sample variance of MBB and NBB estimators will match under the strongest forms $\theta \in (0, 1/2]$ of LRD. In this situation, the dependence among observations is so strong that, despite more available blocks, the MBB completely loses any variance advantage over the NBB.

However, unlike bootstrap biases, the variances of bootstrap estimators do exhibit a type of continuity from strong to weak dependence. It turns out that the limiting variance ratio

$$
\Delta_\theta \equiv \lim_{n \to \infty} \frac{\text{Var}(\hat{\sigma}_{\ell,\theta,\text{MBB}})}{\text{Var}(\hat{\sigma}_{\ell,\theta,\text{NBB}})} = \begin{cases} 
1 & \text{if } 0 < \theta \leq 1/2, \\
\int_{-\infty}^{\infty} g_{\theta}(x)dx / \sum_{x=\infty}^{x=-\infty} g_{\theta}(x) & \text{if } 1/2 < \theta < 1,
\end{cases}
$$

(3.9)

is a continuous function of the dependence parameter $\theta \in (0, 1)$ and strictly decreasing on $(1/2, 1)$ with $\lim_{\theta \to 1} \Delta_\theta = 2/3$; see Figure 3.3 for illustration. In other words, as the dependence weakens and $\theta \uparrow 1$, the MBB recovers its $2/3$ variance reduction under short-memory. Since MBB/NBB estimators $\hat{\sigma}_{\ell,\theta}^2$ have the right form for weak dependence by substituting $\theta = 1$, we may meaningfully define the variance ratio $\Delta_1 \equiv 2/3$ at $\theta = 1$ to reinforce this continuity in relative variances (3.9) between strong and weak dependence cases.

### 3.4.2 Optimized block sizes and mean-squared error

The optimal block sizes $\ell_{\theta}^{\text{opt}}$ for minimizing the large sample mean squared error (MSE) $E(\hat{\sigma}_{\ell,\theta}^2 - \sigma_n^2)^2$ of a bootstrap variance estimator are provided in Corollary 1. To capture the dependence structure, one might intuitively expect the order of $\ell_{\theta}^{\text{opt}}$ to increase with the dependence strength, but the opposite holds true and large block orders are required as dependence weakens.

**Corollary 1.** Let $\theta \in (0, 1)$ . Under the assumptions of Theorem 2(ii) and as $n \to \infty$,
(i) the optimal block size $\ell_{\theta}^{\text{opt}}$ for a block bootstrap variance estimator $\hat{\sigma}_{\ell,\theta}^2$ is given by

$$\ell_{\theta}^{\text{opt}} = \begin{cases} 
  a_{\theta} n^\theta & \text{if } 0 < \theta < 1/2, \\
  (a_{\theta} n / \log n)^{1/2} & \text{if } \theta = 1/2, \\
  \{a_{\theta} n\}^{1/(3-2\theta)} & \text{if } 1/2 < \theta < 1,
\end{cases}$$

where $a_{\theta} \equiv (1 - \theta) B_\theta^2 / \{\sigma_{\infty,\theta}^4 V_\theta\}$ for $1/2 \leq \theta < 1$ and, for $0 < \theta < 1/2$, $a_{\theta} \equiv \{|B_\theta| (1 + 8\theta (1 - \theta) V_\theta)^{1/2} - B_\theta (1 - 2\theta)\} / \{2\theta \sigma_{\infty,\theta}^2 [1 + 2 V_\theta]\}$.

(ii) at the optimal block $\ell = \ell_{\theta}^{\text{opt}}$, the MSE $E(\hat{\sigma}_{\ell,\theta}^2 - \sigma_{n,\theta}^2)^2$ has order given by $O(n^{-2\theta(1-\theta)})$, $O(\lfloor n/\log n \rfloor^{-1/2})$, $O(n^{-2(1-\theta)/(3-2\theta)})$ in the cases $0 < \theta < 1/2$, $\theta = 1/2$, and $1/2 < \theta < 1$, respectively.

Remark 5. Corollary 1 implies that the theoretically optimal blocks $\ell_{\theta}^{\text{opt}}$ for MBB and NBB variance estimators are the same for $0 < \theta \leq 1/2$. When $1/2 < \theta < 1$, the optimal block for the MBB is larger than that of the NBB by a factor $\Delta_{\theta}^{-1/(3-2\theta)}$, using the limiting variance ratio (3.9). For comparison, in the weak dependence case, theoretical MBB blocks should a larger by a factor $(3/2)^{1/3}$ for variance estimation (Künsch, 1989).

Remark 6. The results in Corollary 1 assume the long-memory parameter $\theta$ is known. If we replace $\hat{\sigma}_{\ell,\theta}$ with a bootstrap variance $\hat{\sigma}_{\ell,\hat{\theta}}^2$ based on an appropriate estimator $\hat{\theta}$, the optimal block expressions $\ell_{\theta}^{\text{opt}}$ and optimized MSE orders will typically not change. For example, with estimators of $\theta$ described in Remark 2, it suffices to assume that $m^{1/2}(\hat{\theta} - \theta)$ has enough bounded moments with a bandwidth form $m = C n^\kappa$, $\kappa > 1/2$, which are not overly restrictive and the bandwidth condition is consistent with practical implementation (cf. Robinson, 1995ab).

To explain the behavior in optimal block sizes, note that the block bootstrap bias in Theorem 2(i) becomes worse as $\theta \uparrow 1$, thus requiring larger block orders to reduce this bias. The best order of bootstrap MSE occurs for $\theta = 1/2$ and convergence rates worsen as $\theta$ approaches 0 or 1 (i.e., as the dependence strength relatively increases or decreases).

To illustrate the theoretical results, Figure 3.4 shows MSE curves, as a function of block
length $\ell$, for MBB/NBB estimators $\hat{\sigma}^2_{\ell,\theta}$ for LRD (mean-zero) Gaussian series with covariance function $r(k) = |k - 1|^{2-\theta} + |k + 1|^{2-\theta} - 2|k|^{2-\theta}$, $k > 0$ and $r(0) = 5$, satisfying (3.8) with $\sigma^2 = (1 - \theta)(2 - \theta)$. Three long-memory parameters $\theta = 0.2, 0.5, 0.8$ and four sample sizes are considered. As the sample size increases, Figure 3.4 shows that the MSEs for the MBB and NBB estimators closely match when $\theta = 0.2$ and 0.5, which agrees with Theorem 2 indicating that both estimators should have the same essential large-sample bias and variance. In the case $\theta = 0.8$, the ratio of the empirical MSEs for MBB/NBB estimators more slowly converge to their theoretical ratio $\Delta_{\theta,\ell}^{2(1-\theta)/(3-2\theta)} \approx 0.94$. Empirically determined blocks for minimizing MSE in Figure 3.4, say $\ell_{\theta,\text{emp}}$, generally agree with theoretical blocks $\ell_{\theta,\text{opt}}$, especially for smaller values of $\theta$. For example, pairs $(\ell_{\theta,\text{opt}}, \ell_{\theta,\text{emp}})$ are (3.6, 3), (9.4, 7), (19.4, 6) for $\theta = 0.2, 0.5, 0.8$ respectively with $n = 100$ and (5.7, 6), (24.2, 25), (100.4, 37) for $n = 1000$. For weaker dependence $\theta = 0.8$, the finite-sample best blocks $\ell_{\theta,\text{emp}}$ are slow to match the large-sample version $\ell_{\theta,\text{opt}}$ as the sample size grows; this feature is not necessarily negative because the MSE curves are relatively flat over a large block range between $\ell_{\theta,\text{emp}}$ and $\ell_{\theta,\text{opt}}$, implying that block bootstrap performance is less sensitive to block choice in this case.

Data-driven block choices are possible for variance estimation, based on the large-sample theoretical expressions for $\ell_{\theta,\text{opt}}$ in Corollary 1 and general estimation approaches existing in the literature. One possibility is a plug-in estimator which substitutes estimates for unknown quantities in $\ell_{\theta,\text{opt}}$. Lahiri, Furukawa and Lee (2007) describe a general route for this (for example, estimating the bias $B_{\theta}$ term in $\ell_{\theta,\text{opt}}$ as $\hat{B}_{\theta,\ell} = \ell^{1-\theta}(1 - 2^{-1 + \theta})^{-1}(\hat{\sigma}^2_{\theta,\ell} - \hat{\sigma}^2_{\theta,2\ell})$ with estimators at two block lengths $\ell$ and $2\ell$, which is asymptotically unbiased and consistent by Theorem 2). Since the order of $\ell_{\theta,\text{opt}}$ is known, the subsampling method of Hall, Horowitz and Jing (1995) could also be applied to choose a block size, which essentially estimates the scaling term $a_{\theta}$ in $\ell_{\theta,\text{opt}}$ directly. In either case, estimation of the long-memory parameter $\theta$ can be performed along lines described in Remark 2.

We add finally that the results on bootstrap block selection under LRD do not apply to order selection for the sieve bootstrap. In fact, simulations indicate that the best autoregressive order for the sieve bootstrap appears to increase as dependence strength increases ($\theta \to 0$) under LRD, which is the opposite of optimal MBB blocks. To illustrate, Figure 3.5 shows MSE curves
for the sieve bootstrap with data generated from FARIMA models (3.7), for which the method
is valid. For a sample size \( n = 1000 \) and parameters \( \theta = 0.2, 0.5, 0.8 \), the optimal autoregressive
orders are 95, 69, 29, respectively, and the MSEs are generally smaller under weaker dependence
(e.g., \( \theta = 0.8 \)).

3.5 Concluding remarks

This paper has developed block bootstrap estimation of the sampling distribution of the sample
mean for a large class of linear time series exhibiting long-memory. The assumptions on the
linear process essentially accommodate weak dependence as well, showing that there is some
continuity in the validity of block bootstrap between weak and strong dependence settings. For
variance estimation, the relative efficiencies of moving and non-overlapping block bootstrap es-
timators under strong dependence are drastically different than the weak dependence case, since
both bootstraps may have the same large-sample variance when the underlying dependence is
strong enough.

While the performance of the methods depends on a block choice and optimal blocks for
variance estimation are shown to decrease as the strength of the underlying dependence in-
creases, blocks of size \( O(n^{1/2}) \) may be a compromise for use in practice. Some simulations
suggest that this block size performs reasonably over a range of dependence structures for
interval estimation as well. In addition to variance estimation, optimal block choices for esti-
mation distribution might be possible using recent Edgeworth expansions for the sample mean
with linear, long-range dependent series (Fay, Mouline and Soulier, 2004). Under strong de-
pendence, we conjecture that these blocks might also increase as the underlying dependence
decreases, which is consistent with our simulation evidence.

3.6 Proofs of main results

In the following, let \( C \) denote a generic constant, not dependent on the data, \( n \) or any subscripts.
3.6.1 Proof of Theorem 1

We consider only the MBB case; the NBB version is similar. Let \( n_\ell = n - \ell + 1 \) denote the available number of blocks for MBB resampling. Recall \( \mathbb{E}X_\ell = \mu \) and the MBB sample mean has expectation

\[
\hat{\mu}_n \equiv \mathbb{E}_s(\bar{X}_N^*) = \frac{1}{n_\ell} \left( n\bar{X}_n - \frac{1}{\ell} \sum_{i=1}^{\ell} (\ell - i)(X_i + X_{n-i+1}) \right)
\]  

(3.10)
as in (3.4). To show Theorem 1, we essentially require the limits in (3.3) and (3.6) along with two results stated in Lemma 1. A proof of Lemma 1 is provided later.

**Lemma 1.** Under the assumptions of Theorem 1, as \( n \to \infty \), (a) \( \ell^6\mathbb{E}(\hat{\mu}_n - \mu)^2 \to 0 \); and (b) \( b\ell^4\text{Var}_s(\bar{X}_N^*) - \sigma_{\infty,\theta}^2 \xrightarrow{p} 0 \).

Theorem 1(ii) follows directly from Lemma 1(b) and (3.3) recalling \( \hat{\sigma}_{\ell,\theta}^2 = b\ell^4\text{Var}_s(\bar{X}_N^*) \).

Let \( \bar{B}_1^*, \ldots, \bar{B}_b^* \) denote the sample averages of the \( b = [n/\ell] \) resampled MBB blocks; these are independent and identically distributed (iid) with a uniform distribution on \( \bar{B}_i, i = 1, \ldots, n_\ell \) (the collection of the length \( \ell \) block averages as defined in (3.4)). Then, \( T_n^* = \ell^2(\bar{X}_N^* - E_s\bar{X}_N^*) = \sum_{i=1}^{b} W_i^* \) is a sum of iid variables \( W_i^* = b^{-1/2}\ell^{3/2}[\bar{B}_i^* - E_s\bar{B}_i^*], E_s\bar{B}_i^* = \hat{\mu}_n \). To ease notation, we assume \( E(X_\ell) = \mu = 0 \) without loss of generality. Let \( M = \sigma_{\infty,\theta}^2/16 \) and define

\[
\hat{\Delta}_n \equiv bE_s|W_1^*|^2\mathbb{I}(|W_1^*| > 2M) = \frac{\ell^3}{n_\ell} \sum_{i=1}^{n_\ell} [\bar{B}_i - \hat{\mu}_n]^2\mathbb{I}(\ell^{3/2}|\bar{B}_i - \hat{\mu}_n| > 2Mb^{1/2})
\]

\[
\leq 4\ell^3\hat{\mu}_n^2 + 4\ell^3\sum_{i=1}^{n_\ell} \bar{B}_i^2\mathbb{I}(\ell^{3/2}|\bar{B}_i| > Mb^{1/2})
\]

using the indicator \( \mathbb{I}(\cdot) \) function. Then, \( \hat{\Delta}_n \xrightarrow{p} 0 \) follows from \( \mathbb{E}\hat{\Delta}_n \to 0 \), which holds by Lemma 1(a) and the fact that \( \ell^2\mathbb{E}\bar{B}_1^2\mathbb{I}(\ell^{3/2}|\bar{B}_1| > Mb^{1/2}) \to 0 \) under the extended dominated convergence theorem with \( \sigma_{\ell,\theta}^2 = \ell^2\mathbb{E}\bar{B}_1^2 \to \sigma_{\infty,\theta}^2 \) by (3.3) and \( \ell^{3/2}\bar{B}_1 \xrightarrow{d} N(0, \sigma_{\infty,\theta}^2) \) by (3.6).

From this and Lemma 1(a), noting \( \hat{\sigma}_{\ell,\theta}^2 = \text{Var}_s(T_n^*) \), it holds that, given any subsequence \( \{n_k\} \), there exists a further subsequence \( \{n_j\} \subset \{n_k\} \) where \( \hat{\Delta}_{n_j} \to 0 \) and \( \text{Var}_s(T_{n_j}^*) \to \sigma_{\infty,\theta}^2 \) with probability 1 (wp1). Now Lindeberg’s condition holds and \( T_{n_j}^* \xrightarrow{d} N(0, \sigma_{\infty,\theta}^2) \) wp1 and, by Polya’s theorem, \( \sup_{x \in \mathbb{R}} |P_s(T_{n_j}^* \leq x) - \Phi(x/\sigma_{\infty,\theta})| \to 0 \) wp1, where \( \Phi(\cdot) \) denotes the standard normal distribution function. Theorem 1(i) now follows from (3.6). (These arguments are a
modication of Lahiri (2003, Sec. 3.2.2). Part (iii) of Theorem 1 then follows from $\ell^{\theta-\eta} \overset{p}{\to} 1$ and Slutsky’s theorem. □

**Proof of Lemma 1.** For $S_{t,\mu} \equiv \sum_{i=1}^{\ell} (1-\ell^{-1}i)(X_i-\mu)$, $\text{Var}(S_{t,\mu})$ is bounded by $\ell \sum_{k=0}^{\ell} |r(k)| = O(\ell^2 \theta)$ and hence $\ell^\theta \hat{\mu}_n^2 \leq C(\ell/n)^{\theta} \to 0$ follows in Lemma 1(a) from (3.3), (3.10) and $\ell = o(n)$.

To show Lemma 1(b), write $\hat{\sigma}^2_{\ell,\theta} = S_n - \ell^\theta \hat{\mu}_n^2$ with $S_n \equiv \ell^\theta \sum_{i=1}^{n} \hat{B}_i^2 / n_{\ell}$ based on block means from (3.4). By Lemma 1(a), it suffices to show $S_n \overset{p}{\to} \sigma^2_{\infty,\theta}$. For $m \geq 1$, let observations $X_{t,m}$, $t \in \mathbb{Z}$, be defined by replacing errors $\{\varepsilon_t\}$ in (3.1) with truncated and centered versions $\{\varepsilon_{t,m}\}$, for $\varepsilon_{t,m} = \varepsilon_t I(|\varepsilon_t| \leq m) - E\varepsilon_t I(|\varepsilon_t| > m)$ with $E\varepsilon_{0,m}^2 \in (0,\infty)$. Let $\bar{B}_{t,m}$, $i = 1, \ldots, n_{\ell}$, be versions of sample block averages $\bar{B}_i$, $i = 1, \ldots, n_{\ell}$, upon replacing $\{X_t\}_{t=1}^{n}$ with $\{X_{t,m}\}_{t=1}^{n}$ and let $S_{n,m} \equiv \ell^\theta \sum_{i=1}^{n_{\ell}} \bar{B}_{t,m}^2 / n_{\ell}$. For any fixed $m$, we will show $S_{n,m} \overset{p}{\to} [E\varepsilon_{0,m}^2/E\varepsilon_{0}^2] \sigma^2_{\infty,\theta}$ as $n \to \infty$.

From this, it follows that $S_n \overset{p}{\to} \sigma^2_{\infty,\theta}$ using $\lim_{m \to \infty} E\varepsilon_{0,m}^2 - E\varepsilon_{0}^2$ along with

$$
\sup_{n \geq 1} E|S_n - S_{n,m}| \leq \sup_{n \geq 1} 2\ell^\theta \{E|\bar{B}_1 - \bar{B}_{1,m}|^2[E\bar{B}_1^2 + E\bar{B}_{1,m}^2]\}^{1/2} \leq C \sqrt{E\varepsilon_{0,m}^2 I(|\varepsilon_t| \geq m)} \to 0
$$
as $m \to \infty$ under (3.3) and $\sigma^2_{\ell,\theta} = \ell^\theta E\bar{B}_1^2 = [E\varepsilon_{0,m}^2/E\varepsilon_{0}^2] \ell^\theta E\bar{B}_{1,m}^2$.

Fix $m \geq 1$. Because $E S_{n,m} = [E\varepsilon_{0,m}^2/E\varepsilon_{0}^2] \sigma^2_{\ell,\theta} \to [E\varepsilon_{0,m}^2/E\varepsilon_{0}^2] \sigma^2_{\infty,\theta}$ as $n \to \infty$ under (3.3), Lemma 1(b) will follow by showing $\lim_{n \to \infty} \text{Var}(S_{n,m}) = 0$. We may expand

$$
\text{Var}(S_{n,m}) = \left( \frac{E\varepsilon_{0,m}^2}{E\varepsilon_{0}^2} \right)^2 \frac{1}{n_{\ell}} \sum_{i=0}^{n_{\ell}-1} (I(i \neq 0) + 1)(n_{\ell} - i) \left(2 \left[ \ell^\theta \text{Cov}(\bar{B}_1, \bar{B}_{1+i}) \right]^2 + c_{u_{i+1,\ell}} \right)
$$

(3.11)

where $c_{u_{i+1,\ell}} \equiv \text{cumulant}(\ell^{\theta/2} \bar{B}_1, \ell^{\theta/2} \bar{B}_1, \ell^{\theta/2} \bar{B}_{1+i}, \ell^{\theta/2} \bar{B}_{1+i})$; the second parenthetical term in (3.11) is an expression for $\text{Cov}(\ell^{\theta/2} \bar{B}_1, \ell^{\theta/2} \bar{B}_{1+i})$. From (3.1), we may write $\ell^{\theta/2} \bar{B}_i = \sum_{j \in \mathbb{Z}} h_{j,i,\ell} \varepsilon_j$ with $h_{j,i,\ell} \equiv \ell^{\theta/2} \sum_{k=0}^{\ell-1} b_{i+k-j} / \ell$, $h_{j,i,\ell} = h_{j-i+1,1}$, for $i \geq 1$, $j \in \mathbb{Z}$. Since $\sigma^2_{\ell,\theta} = E\varepsilon_{0}^2 \sum_{j \in \mathbb{Z}} h_{j,1,\ell}^2$ implies $\sum_{j \in \mathbb{Z}} h_{j,1,\ell}^2 = O(1)$ and we may write $c_{u_{i+1,\ell}} = (E\varepsilon_{0}^4 - 3(E\varepsilon_{0}^2)^2) \sum_{j \in \mathbb{Z}} h_{j,1,\ell}^2 h_{j,i+1,\ell}^2$, we can bound

$$
\frac{1}{n_{\ell}} \sum_{i=0}^{n_{\ell}-1} |c_{u_{i+1,\ell}}| \leq C \frac{C_{\ell,\theta}}{n_{\ell}} \sum_{j \in \mathbb{Z}} h_{j,1,\ell}^2 \sum_{i=0}^{n_{\ell}} h_{j-i+1,1,\ell}^2 \leq C / n_{\ell}.
$$

(3.12)

By the Cauchy-Schwarz inequality, $|\ell^\theta \text{Cov}(\bar{B}_1, \bar{B}_{1+i})| \leq \sigma^2_{\ell,\theta}$ holds so that

$$
\frac{1}{n_{\ell}} \sum_{i=0}^{n_{\ell}-1} \left[ \ell^\theta \text{Cov}(\bar{B}_1, \bar{B}_{1+i}) \right]^2 \leq C \sigma^2_{\ell,\theta} + \max_{n_{\ell} \leq i \leq n} \left[ \ell^\theta \text{Cov}(\bar{B}_1, \bar{B}_i) \right]^2
$$
for any $\epsilon \in (0, 1)$. Then $\lim_{n \to \infty} \text{Var}(S_{n,m}) \leq 4\sigma_{\infty, \theta}^2\epsilon$ follows for arbitrary $\epsilon > 0$ by (3.3), (3.5) and (3.12), implying $\text{Var}(S_{n,m}) \to 0$ and proving Lemma 1(b). □

3.6.2 Proof of Theorem 2.

We will mainly treat the MBB case and describe modifications for the NBB. Using notation from the proof of Lemma 1 and assuming $E X_t = \mu = 0$, write $\tilde{\sigma}^2_{\ell, \theta, MBB} = S_n - \ell \hat{\mu}_n^2$; the NBB estimator defined in Section 3.2.2 can be analogously written $\hat{\sigma}^2_{\ell, \theta, NBB} = S_{n,NBB} - \ell \hat{\mu}_{n,NBB}^2$, where $S_{n,NBB} = \sum_{i=1}^b \ell \theta \bar{B}_{1+(i-1)\ell}/b$.

The expectation of the MBB estimator is $E \hat{\sigma}^2_{\ell, \theta, MBB} = \sigma^2_{\ell, \theta} - \ell \theta E \hat{\mu}_n^2$. Direct computation, using (3.10) and the covariance form (3.8), yields $n^\theta E \hat{\mu}_n^2 \to \sigma^2_{\infty, \theta}$ while $\sigma^2_{\ell, \theta} = \sigma^2_{\infty, \theta} + B_\theta n^{-1+\theta}(1 + o(1))$ holds under (3.8) (cf. Giraitis, Robinson and Surgailis, 1999, p. 14). The bias of the MBB estimator then follows in Theorem 2(i); the NBB bias is derived similarly.

To determine

$$\text{Var}(\hat{\sigma}^2_{\ell, \theta, MBB}) = \text{Var}(S_n) + \ell^{2\theta} \text{Var}(\hat{\mu}_n^2) - 2\text{Cov}(S_n, \ell \hat{\mu}_n^2),$$

we first find an expression for $\ell^{2\theta} \text{Var}(\hat{\mu}_n^2)$ and then for $\text{Var}(S_n)$. When $\theta \geq 1/2$, these two expressions are enough to see the main form of (3.13), which is determined only by $\text{Var}(S_n)$. However, when $\theta < 1/2$, every term in (3.13) contributes significantly to the variance. Note that the expansion in (3.11) prescribes $\text{Var}(S_n)$ by letting $m \to \infty$ there.

The $\ell^{2\theta} \text{Var}(E \hat{\mu}_n^2)$ component. The covariances (3.8) with (3.10) and (3.14) imply $n^{\theta} E (\bar{X}_n - \hat{\mu}_n)^2 \to 0$ so that $n^{\theta/2} \hat{\mu}_n \xrightarrow{d} N(0, \sigma^2_{\infty, \theta})$ holds by (3.6). From the process form (3.1) and $E \varepsilon_t^6 < \infty$, Davydov (1971, Theorem 3) yields

$$E (\bar{X}_n - \mu)^6 \leq C [\text{Var}(\bar{X}_n)]^3, \quad E (S_{\ell, \mu})^6 \leq C [\text{Var}(S_{\ell, \mu})]^3,$$

where $S_{\ell, \mu} \equiv \sum_{i=1}^\ell (1 - \ell^{-1})(X_i - \mu)$, so that $n^{3\theta} E \hat{\mu}_n^6 \leq C$ by (3.10) and $\{n^{2\theta} \hat{\mu}_n^4\}$ is uniformly integrable. Hence, $\text{Var}(n^{\theta} \hat{\mu}_n^2) \to 2\sigma^4_{\infty, \theta}$ so that

$$\ell^{2\theta} \text{Var}(\hat{\mu}_n^2) = 2\sigma^4_{\infty, \theta} \left( \frac{\ell}{n} \right)^{2\theta} (1 + o(1));$$

(3.15)
the same expansion holds for $\text{Var}(\hat{\mu}_{n,NBB}^2)$.

The $\text{Var}(S_n)$ component when $\theta > 1/2$. We first determine $\text{Var}(S_n)$ by supposing that $\theta \in (1/2, 1)$. Note that the covariance form (3.8) implies

$$|\ell^\theta \text{Cov}(B_1, B_{1+i})| \leq C(i/\ell)^{-\theta} + C\ell^{-1+\theta} \sum_{k=-\ell+1}^{\ell-1} |r_1(i+k)|, \quad i \geq 2\ell \quad (3.16)$$

holds and $|\ell^\theta \text{Cov}(B_1, B_{1+i})| \leq C, i \geq 1$. From (3.11), (3.12) and (3.16) and recalling $b = [n/\ell]$, we may write

$$\text{Var}(S_n) = \frac{4}{n_\ell} \sum_{i=1}^{n_\ell-1} \left[ \ell^\theta \text{Cov}(B_1, B_{1+i}) \right]^2 + O \left( \frac{\ell^{2\theta}}{n_\ell^2} \sum_{k=1}^{n_\ell} k^{1-2\theta} + \frac{\ell^{2\theta}}{\ell n_\ell^2} \sum_{k \in \mathbb{Z}} |r_1(k)| + \frac{\ell^2}{n^2} + \frac{1}{n} \right)$$

$$= \frac{4\ell}{n_\ell} \sum_{k=1}^{b} \frac{1}{\ell} \sum_{i=(k-1)\ell+1}^{k\ell} \left[ \ell^\theta \text{Cov}(\tilde{B}_1, \tilde{B}_{1+i}) \right]^2 + o(\ell/n). \quad (3.17)$$

For $x > 0$, define a function $g_n(x) = \ell^\theta \text{Cov}(\tilde{B}_1, \tilde{B}_{1+[\ell x]})(x \leq b)$. Fix $x > 0$ and let $k_x = \lceil x \rceil + 1$. Then, the process $Y_x(t) = \sum_{i=1}^{[k_x]} X_i/([\ell k_x]^{2-\theta} \sigma_{\infty,\theta}^2)^{1/2}, t \in [0, 1]$ converges in distribution to a Gaussian process with correlation function $C(s, t) = (t^{2-\theta} + s^{2-\theta} - |s - t|^{2-\theta})/2$ by Davydov (1971, Theorem 2 and Lemma 5). Since $\{\ell^\theta \tilde{B}_1 \tilde{B}_{1+[\ell x]}\}$ is uniformly integrable (by (3.14) and $E\sigma_0^6 < \infty$), we have for each $x > 0$ that

$$g_n(x) = \sigma_{\infty,\theta}^2 k_x^{2-\theta} \text{Cov} \left( Y_x \left( \frac{1}{k_x} \right), Y_x \left( \frac{\ell + |x|/k_x - 1}{\ell k_x} \right) - Y_x \left( \frac{|x|/k_x}{\ell k_x} \right) \right)$$

$$\to \sigma_{\infty,\theta}^2 k_x^{2-\theta} \left[ C \left( \frac{1}{k_x}, \frac{x+1}{k_x} \right) - C \left( \frac{1}{k_x}, \frac{x}{k_x} \right) \right] = \sigma_{\infty,\theta}^2 g_0(x).$$

By (3.16), there is a sequence $\tilde{g}_n(x), x > 0$ of integrable functions such that $\tilde{g}_n(x) = C_1$ for $x < 2$ and $C_1 \{x^{-\theta} + \ell^{-1+\theta} \sum_{k=-\ell}^{\ell} |r_1(|x|+k)| \}$ for $x \geq 2$ (for some fixed $C_1 > 0$) such that $g_n^2(x) \leq \tilde{g}_n^2(x)$ for all $x > 0$. Since $\tilde{g}_n(x) \to \tilde{g}(x) \equiv C_1 \{x^{-\theta} \mathbb{I}(x \geq 2) + \mathbb{I}(x < 2)\}$ for $x > 0$ while $\int_0^\infty \tilde{g}_n^2(x)dx - \int_0^\infty \tilde{g}^2(x)dx \leq C\ell^{-1+\theta} \sum Z |r_1(k)| \to 0$, the extended dominated convergence theorem can be applied

$$\int_0^{b} \frac{1}{\ell} \sum_{i=(k-1)\ell+1}^{k\ell} \left[ \ell^\theta \text{Cov}(\tilde{B}_1, \tilde{B}_{1+i}) \right]^2 = \int_0^\infty g_n^2(x)dx \to \sigma_{\infty,\theta}^4 \int_0^\infty g_0^2(x)dx.$$

This along with (3.15) and (3.17) prove the MBB variance result for $\theta \in (1/2, 1)$, using $2 \int_0^\infty g_0^2(x)dx = \int_{-\infty}^\infty g_0^2(x)dx$; in the same case, the main term in the NBB variance is sim-
The Cov($S_n$) component when $\theta \leq 1/2$. For $x > 0$ and $\theta \in (0, 1/2]$, define a function $d_{x, \theta} \equiv x^{2\theta}$ if $\theta < 1/2$ and $x \log(x^{-1})$ if $\theta = 1/2$. In the case $\theta \leq 1/2$, we use (3.16) to write

$$\text{Var}(S_n) = \frac{4\ell}{n\ell} \sum_{k=2}^{b-1} \left(1 - \frac{k\ell}{n\ell}\right) \frac{1}{\ell} \sum_{i=0}^{\ell-1} \left[\ell^\theta \text{Cov}(\bar{B}_1, \bar{B}_{1+i+\ell})\right]^2 + \delta_{1,n}$$

(3.18)

$$\delta_{1,n} = O\left(\frac{(1+2\theta)^n}{n^2} \sum_{k=1}^{\ell} k^{-2\theta} + \frac{(2\theta)^n}{n^2} \sum_{k \in \mathbb{Z}} |r_1(k)| + \frac{\ell}{n}\right) = o(d_{\ell/n, \theta})$$

For $k \geq 2$ and $0 \leq i \leq \ell - 1$, we use the covariance form (3.8) and Taylor expansion to obtain

$$\ell^\theta \text{Cov}(\bar{B}_1, \bar{B}_{1+i+\ell}) = \sigma^2 k^{-\theta} + \delta_n(k),$$

(3.19)

with a remainder $|\delta_n(k)| \leq C(k-1)^{-1-\theta} + C\ell^{-1+\theta} \sum_{d=-2\ell}^{\ell} |r_1(d + k\ell)|$. Then, (3.18) may be expanded as

$$\text{Var}(S_n) = \lambda_n + \delta_{2,n} + o(d_{\ell/n, \theta}), \quad \lambda_n \equiv \frac{4\ell}{n\ell} \sum_{k=2}^{b-1} \left(1 - \frac{k\ell}{n\ell}\right) \sigma^4 k^{-2\theta},$$

(3.20)

where $\delta_{2,n} = O\left(n_\ell^{-1} \ell \sum_{k=2}^{b-1} (k-1)^{-1-2\theta} + n_\ell^{-1} \ell^\theta \sum_{k \in \mathbb{Z}} |r_1(k)| + \ell/n\right) = o(d_{\ell/n, \theta})$. If $\theta = 1/2$, $\lambda_n/d_{\ell/n, \theta} \to 4\sigma^4$ and if $\theta < 1/2$, $\lambda_n/d_{\ell/n, \theta} \to 4\sigma^4 \int_0^1 (1-x)x^{-2\theta} = 2\sigma^4 (1-\theta)^{-1} (1-2\theta)^{-1}$. Note that, for $\theta = 1/2$, we now have the MBB variance as $\text{Var}({\hat{\sigma}}^2_{\ell/\theta,MBB}) = \text{Var}(S_n) + o(d_{\ell/n, \theta}) = 4\sigma^4 d_{\ell/n, \theta}(1 + o(1))$ since the term in (3.15) is $o(d_{\ell/n, \theta})$. Additionally, the NBB estimator has a variance component

$$\text{Var}(S_{n,NBB}) = \frac{4}{b} \sum_{k=2}^{b-1} \left(1 - \frac{k}{b}\right) \left[\ell^\theta \text{Cov}(\bar{B}_1, \bar{B}_{1+k\ell})\right]^2 + o(d_{\ell/n, \theta}) = \lambda_n + o(d_{\ell/n, \theta}),$$

essentially matching that of the MBB in this case.

The Cov($S_n, \ell^\theta \tilde{\mu}^2_n$) component when $\theta < 1/2$. For the case $\theta \in (0, 1/2)$, we may replace $\tilde{\mu}_n$ with $\bar{X}_n$ since $|\text{Cov}(S_n, \ell^\theta |\tilde{\mu}_n^2 - \bar{X}_n^2)| \leq C[d_{\ell/n, \theta}(\ell/n)^{2-\theta}]^{1/2} = o(d_{\ell/n, \theta})$ by (3.10) and Holder’s inequality based on (3.10), (3.14) and (3.20). We then expand

$$\text{Cov}(S_n, \ell^\theta \bar{X}_n^2) = \frac{1}{n\ell} \sum_{i=1}^{n\ell} \left(2 \left[\ell^\theta \text{Cov}(\bar{B}_1, \bar{X}_n)\right]^2 + (\ell/n)^\theta \tilde{c}_{i,\ell}\right)$$
where $c\mu_{i,\ell} \equiv \text{cumulant}(\ell^{\theta/2} \bar{B}_i, \ell^{\theta/2} \bar{B}_i, n^{\theta/2} \bar{X}_n, n^{\theta/2} \bar{X}_n)$ and, with notation and arguments as in (3.12), we have
\[
\frac{1}{n_\ell} \sum_{i=1}^{n_\ell} |c\mu_{i,\ell}| \leq C \frac{1}{n_\ell} \sum_{i=1}^{n_\ell} h_{j,1,n}^2 \sum_{i=1}^{n_\ell} h_{j-i,1,\ell}^2 \leq C \frac{1}{n_\ell}.
\]
Hence, we may write
\[
\text{Cov}(S_n, \ell^\theta \bar{X}_n^2) = \frac{2}{n_\ell} \sum_{k=2}^{b-2} \sum_{\ell=0}^{\ell-1} [\ell^\theta \text{Cov}(\bar{B}_1+1+k\ell, \bar{X}_n)]^2 + o(d_{\ell/n,\theta})
\]
using additionally $|\ell^\theta \text{Cov}(\bar{B}_1, \bar{X}_n)| \leq C$, $i \geq 1$. For each $0 \leq i \leq \ell - 1$, define $\bar{X}_{i,n} \equiv \sum_{j=i+1}^{i+b\ell} X_{j}/n$. The covariances (3.8) and the Cauchy-Schwarz inequality imply $|\ell^\theta \text{Cov}(\bar{B}_1+1+k\ell, \bar{X}_n-\bar{X}_{i,n})| \leq C(\ell/n)$ and we may again re-write
\[
\text{Cov}(S_n, \ell^\theta \bar{X}_n^2) = \frac{2\ell}{n_\ell} \sum_{k=2}^{b-2} [\ell^\theta \text{Cov}(\bar{B}_1+1+k\ell, \bar{X}_0,n)]^2 + o(d_{\ell/n,\theta})
\]
(3.21) since $\text{Cov}(\bar{B}_1+1+k\ell, \bar{X}_{i,n}) = \text{Cov}(\bar{B}_1+1+k\ell, \bar{X}_0,n)$ by stationarity, where $\bar{X}_{0,n} = (\ell/n) \sum_{k=0}^{b-1} \bar{B}_1+1+k\ell$.

Then, for $1 < k < b - 1$, we use (3.19) and sum to expand
\[
(n/\ell)\ell^\theta \text{Cov}(\bar{B}_1+1+k\ell, \bar{X}_0,n) = \sum_{j=0}^{k} \ell^\theta \text{Cov}(\bar{B}_1, \bar{B}_1+j\ell) + \sum_{j=1}^{b-1-k} \ell^\theta \text{Cov}(\bar{B}_1, \bar{B}_1+j\ell)
\]
\[
= \frac{\sigma^2}{1-\theta} \left( k^{-\theta} + (b-1-k)^{-\theta} \right) + \psi_n(k),
\]
where the difference term satisfies $\sup_n \sup_{k \geq 2} |\psi_n(k)| \leq C$ using that $\sup_{k \geq 2} |\sum_{j=2}^{k} j^{-\theta} - \int_{0}^{k} x^{-\theta} dx| \leq C$. Applying this in (3.21), we have
\[
\text{Cov}(S_n, \ell^\theta \bar{X}_n^2) = 2 \left( \frac{\sigma^2}{1-\theta} \right) \frac{n_\ell}{\ell} \sum_{k=2}^{b-1} \left( \frac{\ell}{n} \right)^2 \left( k^{-\theta} + (b-1-k)^{-\theta} \right)^2 + \delta_{3,n} + o(d_{\ell/n,\theta})
\]
(3.22)
where $|\delta_{3,n}| \leq C(\ell/n)^2 |1 + \sum_{k=1}^{b} k^{-\theta}| = o(d_{\ell/n,\theta})$. Recalling $d_{\ell/n,\theta} = (\ell/n)^{2\theta}$ for $\theta < 1/2$, the Riemann integral converges
\[
\frac{1}{d_{\ell/n,\theta}} \frac{\ell}{n_\ell} \sum_{k=2}^{b-1} \left( \frac{\ell}{n} \right)^{-2\theta} \left( k^{-\theta} + (b-1-k)^{-\theta} \right)^2 \to \int_{0}^{1} \left( x^{-\theta} + (1-x)^{-\theta} \right)^2 dx
\]
\[
= \frac{2}{3-2\theta} + \frac{2\Gamma^2(2-\theta)}{\Gamma(4-2\theta)}.
\]
Hence, combining this with (3.15), (3.20), (3.22) and $\sigma^2_{\infty,\theta} = 2\sigma^2/\{(1-\theta)(2-\theta)\}$ into (3.13) gives the MBB estimator's variance $\text{Var}(\hat{\sigma}_{\ell,\theta,MBB}^2) = \text{Var}(S_n - \ell^\theta \hat{\mu}_n^2)$ in the case $\theta < 1/2$. In this case, the NBB estimator has a variance component $\text{Cov}(S_{n,NBB}, \ell^\theta \hat{\mu}_{n,NBB}^2)$ which matches (3.21) so that $\text{Var}(\hat{\sigma}_{\ell,\theta,NBB}^2) = \text{Var}(\hat{\sigma}_{\ell,\theta,MBB}) + o(d_{\ell/n,\theta})$ follows for $\theta < 1/2$. □
References


Liu, R. Y. and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak de-


Table 3.1  Empirical coverage (%) and average length for 90% MBB confidence intervals for EXₜ = µ for FARIMA model parameters θ, sample sizes n and block lengths ℓ = n¹/²/2, n¹/₂, ℓθ, assuming θ given or estimated (based on 1000 simulations).

<table>
<thead>
<tr>
<th>θ</th>
<th>Coverage</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 250</td>
<td>n = 500</td>
<td>n = 1000</td>
<td>n = 5000</td>
<td>n¹/²</td>
<td>n¹/₂</td>
<td>ℓθ</td>
<td>n¹/²</td>
<td>n¹/₂</td>
<td>ℓθ</td>
<td>n¹/²</td>
</tr>
<tr>
<td>0.1</td>
<td>given</td>
<td>57.9</td>
<td>51.0</td>
<td>60.2</td>
<td>60.0</td>
<td>55.4</td>
<td>67.2</td>
<td>63.5</td>
<td>59.0</td>
<td>70.7</td>
<td>68.7</td>
</tr>
<tr>
<td></td>
<td>estim.</td>
<td>51.1</td>
<td>46.3</td>
<td>55.6</td>
<td>57.2</td>
<td>52.6</td>
<td>64.6</td>
<td>60.1</td>
<td>57.3</td>
<td>67.4</td>
<td>68.3</td>
</tr>
<tr>
<td>0.3</td>
<td>given</td>
<td>79.8</td>
<td>75.0</td>
<td>76.8</td>
<td>82.7</td>
<td>79.3</td>
<td>82.8</td>
<td>82.6</td>
<td>79.0</td>
<td>81.8</td>
<td>85.1</td>
</tr>
<tr>
<td></td>
<td>estim.</td>
<td>76.6</td>
<td>72.4</td>
<td>77.9</td>
<td>82.6</td>
<td>78.7</td>
<td>83.9</td>
<td>81.8</td>
<td>79.1</td>
<td>83.0</td>
<td>85.4</td>
</tr>
<tr>
<td>0.5</td>
<td>given</td>
<td>84.6</td>
<td>81.0</td>
<td>76.8</td>
<td>87.1</td>
<td>85.4</td>
<td>82.6</td>
<td>85.8</td>
<td>83.7</td>
<td>81.8</td>
<td>87.6</td>
</tr>
<tr>
<td></td>
<td>estim.</td>
<td>82.7</td>
<td>79.4</td>
<td>78.5</td>
<td>85.3</td>
<td>83.7</td>
<td>84.2</td>
<td>84.1</td>
<td>82.3</td>
<td>82.1</td>
<td>88.4</td>
</tr>
<tr>
<td>0.7</td>
<td>given</td>
<td>88.5</td>
<td>86.5</td>
<td>75.6</td>
<td>88.8</td>
<td>87.1</td>
<td>79.0</td>
<td>89.6</td>
<td>88.9</td>
<td>82.6</td>
<td>88.3</td>
</tr>
<tr>
<td></td>
<td>estim.</td>
<td>87.3</td>
<td>85.5</td>
<td>79.4</td>
<td>88.4</td>
<td>86.9</td>
<td>82.8</td>
<td>87.1</td>
<td>86.8</td>
<td>84.2</td>
<td>87.5</td>
</tr>
<tr>
<td>0.9</td>
<td>given</td>
<td>90.1</td>
<td>90.0</td>
<td>78.3</td>
<td>89.0</td>
<td>88.4</td>
<td>73.9</td>
<td>90.5</td>
<td>89.6</td>
<td>73.8</td>
<td>91.2</td>
</tr>
<tr>
<td></td>
<td>estim.</td>
<td>93.2</td>
<td>92.8</td>
<td>80.6</td>
<td>91.7</td>
<td>90.0</td>
<td>74.5</td>
<td>91.4</td>
<td>90.1</td>
<td>75.7</td>
<td>91.1</td>
</tr>
</tbody>
</table>

Average lengths

<table>
<thead>
<tr>
<th>θ</th>
<th>Average lengths</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 250</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>given</td>
<td>2.245</td>
<td>1.943</td>
<td>2.406</td>
<td>2.318</td>
<td>2.084</td>
<td>2.643</td>
<td>2.371</td>
<td>2.150</td>
<td>2.745</td>
<td>2.497</td>
</tr>
<tr>
<td></td>
<td>estim.</td>
<td>2.015</td>
<td>1.776</td>
<td>2.258</td>
<td>2.210</td>
<td>1.997</td>
<td>2.590</td>
<td>2.260</td>
<td>2.103</td>
<td>2.674</td>
<td>2.526</td>
</tr>
<tr>
<td>0.3</td>
<td>given</td>
<td>1.320</td>
<td>1.210</td>
<td>1.292</td>
<td>1.254</td>
<td>1.166</td>
<td>1.258</td>
<td>1.167</td>
<td>1.102</td>
<td>1.182</td>
<td>0.963</td>
</tr>
<tr>
<td></td>
<td>estim.</td>
<td>1.370</td>
<td>1.238</td>
<td>1.454</td>
<td>1.290</td>
<td>1.193</td>
<td>1.389</td>
<td>1.204</td>
<td>1.122</td>
<td>1.279</td>
<td>0.982</td>
</tr>
<tr>
<td>0.5</td>
<td>given</td>
<td>0.772</td>
<td>0.725</td>
<td>0.688</td>
<td>0.657</td>
<td>0.622</td>
<td>0.608</td>
<td>0.561</td>
<td>0.543</td>
<td>0.530</td>
<td>0.383</td>
</tr>
<tr>
<td></td>
<td>estim.</td>
<td>0.830</td>
<td>0.764</td>
<td>0.808</td>
<td>0.691</td>
<td>0.652</td>
<td>0.676</td>
<td>0.580</td>
<td>0.554</td>
<td>0.566</td>
<td>0.391</td>
</tr>
<tr>
<td>0.7</td>
<td>given</td>
<td>0.449</td>
<td>0.434</td>
<td>0.372</td>
<td>0.353</td>
<td>0.341</td>
<td>0.300</td>
<td>0.280</td>
<td>0.274</td>
<td>0.247</td>
<td>0.161</td>
</tr>
<tr>
<td></td>
<td>estim.</td>
<td>0.495</td>
<td>0.467</td>
<td>0.439</td>
<td>0.371</td>
<td>0.359</td>
<td>0.332</td>
<td>0.286</td>
<td>0.278</td>
<td>0.263</td>
<td>0.163</td>
</tr>
<tr>
<td>0.9</td>
<td>given</td>
<td>0.273</td>
<td>0.273</td>
<td>0.224</td>
<td>0.200</td>
<td>0.199</td>
<td>0.155</td>
<td>0.146</td>
<td>0.145</td>
<td>0.111</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>estim.</td>
<td>0.331</td>
<td>0.316</td>
<td>0.250</td>
<td>0.227</td>
<td>0.220</td>
<td>0.166</td>
<td>0.159</td>
<td>0.155</td>
<td>0.117</td>
<td>0.071</td>
</tr>
</tbody>
</table>
Table 3.2  Empirical coverage (%) for 90% sieve bootstrap intervals for $E X_t = \mu$ for FARIMA model parameters $\theta$, sample sizes $n$ and orders $p_n = \lfloor 2(\log n)^2 \rfloor$ or $\hat{p}_n$, assuming $\theta$ given or estimated (based on 1000 simulations). The average estimated order also appears for each $(\theta, n)$ pair.

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$n = 250$ $[2(\log n)^2]$ $\hat{p}_n$</th>
<th>$n = 500$ $[2(\log n)^2]$ $\hat{p}_n$</th>
<th>$n = 1000$ $[(\log n)^2]$ $\hat{p}_n$</th>
<th>average estimated order $\hat{\theta}_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>17.6 17.6 17.4  21.5 21.5 21.0 21.2 21.7 21.0</td>
<td>22.0 22.0 21.7 21.0</td>
<td>22.0 22.0 21.7 21.0</td>
<td>17.42 21.90 27.55</td>
</tr>
<tr>
<td>estim.</td>
<td>36.6 36.6 36.8 41.6 41.6 40.9 41.7 40.7 38.3</td>
<td>40.9 40.9 40.7 38.3</td>
<td>40.9 40.9 40.7 38.3</td>
<td>17.34 21.18 26.15</td>
</tr>
<tr>
<td>0.3</td>
<td>52.0 52.0 51.5 57.0 58.1 58.1 58.2 59.7 54.2 54.2</td>
<td>54.2 54.2 54.4 49.2</td>
<td>54.2 54.2 54.4 49.2</td>
<td>16.16 19.09 23.09</td>
</tr>
<tr>
<td>estim.</td>
<td>63.6 63.6 63.5 72.7 68.4 68.4 69.0 71.3 70.7 70.7</td>
<td>70.7 70.7 70.7 69.1</td>
<td>70.7 70.7 70.7 69.1</td>
<td>14.95 16.48 18.68</td>
</tr>
<tr>
<td>0.5</td>
<td>74.2 74.2 74.5 83.4 77.0 77.0 76.6 83.0 79.3 79.3</td>
<td>79.3 79.3 79.5 81.9</td>
<td>79.3 79.3 79.5 81.9</td>
<td>13.85 14.16 14.93</td>
</tr>
<tr>
<td>estim.</td>
<td>84.2 84.2 84.5 93.4 87.0 87.0 86.6 93.0 89.3 89.3</td>
<td>89.3 89.3 89.5 91.9</td>
<td>89.3 89.3 89.5 91.9</td>
<td>12.75 13.06 13.83</td>
</tr>
</tbody>
</table>

Figure 3.1  For size $n = 500$ samples from FARIMA models $\theta = 0.1, 0.5, 0.9$, the sampling distribution of $n^{\theta/2}(X_n - \mu)$ is shown (dotted line based on 10000 simulations) along with estimates of the block bootstrap (top, $\ell = n^{1/2}$) and sieve bootstrap (bottom, $p_n = \lfloor 2(\log n)^2 \rfloor$) based on five data realizations (solid lines).
Figure 3.2 Empirical coverage of 90% confidence intervals for $E X_t = \mu$ vs autoregressive order for sieve bootstrap (based on Burg’s or Yule-Walker estimation) and coverage vs block size for MBB, based on size $n = 1000$ samples from FARIMA models $\theta$ (from 1000 simulations).
Figure 3.3 Limiting ratio $\Delta_\theta$ of MBB/NBB variances versus the dependence parameter $\theta$. A straight line (dotted) between $(\theta, \Delta_\theta) = (1/2, 1)$ and $(1, 2/3)$ is included for comparison.

Figure 3.4 Scaled MSEs $E(\hat{\sigma}^2_{\ell,\theta} - \sigma^2_{n,\theta}/\sigma^4_{n,\theta}$ against block length $\ell$ for four sample sizes $n$ (based on 10000 simulations). Dots $\bullet$, $\circ$ and $\bullet$ denote the MBB for $\theta = 0.2, 0.5, 0.8$; the NBB is denoted by dotted, solid and dashed lines.
Figure 3.5 Scaled MSEs \( E(\hat{\sigma}^2_{p,\theta} - \sigma^2_{n,\theta})^2 / \sigma^4_{n,\theta} \) of sieve bootstrap variance estimators \( \hat{\sigma}^2_{p,\theta} \) against autoregressive order \( p \) for sample sizes \( n = 100, 1000 \) from FARIMA models (3.7) (based on 10000 simulations). Dots •, ○ and ● denote parameters \( \theta = 0.2, 0.5, 0.8 \).
CHAPTER 4. A FREQUENCY DOMAIN BOOTSTRAP FOR WHITTLE ESTIMATION UNDER LONG-RANGE DEPENDENCE

A paper to be submitted to Journal of Multivariate Analysis

Young Min Kim and Daniel J. Nordman

Abstract

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

Key Words: FARIMA, Interval estimation, Long memory, Spectral density, Periodogram
4.1 Introduction

In this paper, we consider a problem in parametric spectral density estimation for time series that could exhibit potentially 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

\[
\lim_{\lambda \to 0} |\lambda|^{2d} g(\lambda) = C
\]

(4.1)

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 \( d > 0 \). This dependence-type classification is common, in which long-range dependence (LRD) entails a pole of \( g \) at the origin (Hosking, 1981; Beran, 1994). Time series exhibiting LRD often have applications in astronomy, hydrology and economics (cf. Beran, 1994; Montanari, 2003; Henry and Zaffaroni, 2003), where correlations may decrease particularly slowly between observations over time. That is, LRD can be alternatively formulated in terms of slow decay of process autocovariances \( r(k) = \text{Cov}(X_t, X_{t+k}) \approx ak^{-1+2d} \) as \( |k| \to \infty \) for some \( a > 0 \); see Bingham et al. (1987, p. 240) or Robinson (1995a, p. 1634) for mild conditions under which this covariance behavior is equivalent to (4.1). Slow covariance decay implies \( \sum_{k=0}^{\infty} r(k) \) is not finitely summable unlike the usual SRD case and, consequently, statistics and associated methods with LRD processes often exhibit behaviors compared to weak dependence.

For instance, sample means \( \bar{X}_n \) have larger variances \( O(n^{-1+2d}) \) under LRD compared to the \( O(1/n) \) order under SRD and, for example, resampling methods developed for SRD can often fail under LRD without suitable modification (e.g., block bootstrap, Lahiri, 1993).

While some semiparametric approaches focus on estimating the long-memory exponent \( d \) of LRD processes (e.g., Geweke and Porter-Hudak, 1983; Robinson, 1995ab; Giraitis, Robinson and Surgailis, 1999; Moulines and Soulier, 2003, Andrews and Sun, 2004), we consider an inference scenario involving a parametric collection of spectral densities

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

(4.2)

defined by a kernel density \( f(\lambda; \theta) \) involving \( p \) parameters \( \theta = (\theta_1, \ldots, \theta_p)' \). 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 Mandelbrot and van Ness (1968) and the fractional autoregressive integrated moving average (FARIMA) models of Adenstedt (1974), Granger and Joyeux (1980) and Hosking (1981). For fitting such models to data, Whittle estimation is a common approach, which is often computationally easier with less distributional assumptions compared to maximum likelihood for dependent data (Whittle, 1953). For SRD linear processes, Walker (1964) and Hannan (1973) established the consistency and asymptotic normality of Whittle estimators of the parameters \( \theta \). Fox and Taqqu (1986) established the same properties of Whittle estimators with LRD Gaussian processes, while Giraitis and Surgailis (1990) extended these results to more general linear LRD time series. These results enable confidence regions for \( \theta \) 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. Mantegna and Stanley, 1994), which can be more asymmetric (Kang and Yoon, 2007). For example, Palma (2007) 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, Dahlhaus and Janas (1996) 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, including Whittle 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 Janas and Dahlhaus (1994) 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 fre-
frequency domain under SRD, such as nonparametric spectral density estimation (cf. Franke and Härdle, 1992; Nordgaard, 1992). 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., Woodroofe and van Ness, 1967) which is a critical component underlying the FDB of Janas and Dahlhaus (1994) and other frequency domain resampling methods under SRD (cf. Kreiss and Paparoditis, 2003; Jentsch and Kreiss, 2010). 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 (4.2). 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 still requires no assumptions about the full probability structure of the time process. A large simulation study to follow also suggests that the FDB method generally outperforms a normal approximation for estimating Whittle parameters.

The paper is organized as follows. We end this section by briefly summarizing other resampling literature under LRD. Section 4.2 describes the Whittle estimation problem and the associated FDB method. Section 4.3 gives the main distributional results on the consistency of FDB method for distribution estimation of Whittle estimators under LRD. We additionally establish the consistency and asymptotic normality of Whittle estimators under LRD or SRD, with slightly weaker assumptions than previous results from Fox and Taqqu (1986) and Giraitis and Surgailis (1990). Section 4.4 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, computationally simple non-studentized versions of FDB distribution estimators are often much better for interval estimation than studentized FDB versions or normal approximations. In Section 4.5, we provide some concluding remarks. The proofs of all results are deferred to an Appendix (Sections 4.6 and 4.7) and, to
ease the exposition, Section 4.8 includes a collection of numerical tables providing the complete simulation results, which are graphically summarized in Section 4.4.


4.2 Estimation Problem and Bootstrap Method

4.2.1 Whittle estimation with linear time processes

If the data $X_1, \ldots, X_n$ arise from a real-valued, strictly stationary time process, with mean $\mathbb{E}X_t = \mu$ and integrable spectral density $g$ as described in Section 1, then the process $\{X_t\}$ has a moving average representation

$$X_t = \mu + \sum_{j=-\infty}^{\infty} b_j \varepsilon_{t-j}$$

(4.3)
with respect to a process \( \{ \varepsilon_t \} \) of uncorrelated, mean-zero random variables with variance 
\( 0 < \mathbb{E}(\varepsilon_t^2) = \sigma_0^2 \), where the filter sequence \( \{ b_j \} \) satisfies \( \sum_{j=-\infty}^{\infty} b_j^2 < \infty \) with \( b_0 = 1 \) (Ibragimov and Linnik, 1971, Ch. 16.7). The spectral density can then be re-written as 
\[
g(\lambda) = (2\pi)^{-1} \sigma_0^2 \left| \sum_{j=-\infty}^{\infty} e^{ij\lambda} \right|^2, \quad \lambda \in \Pi = (-\pi, \pi], \text{ where } i = \sqrt{-1}.
\]

We suppose the process spectral density \( g \) belongs to a model class \( F \) from (4.2) so that
\[
g(\lambda) \equiv g(\lambda; \sigma_0^2, \theta_0) = \frac{\sigma_0^2}{2\pi} f(\lambda; \theta_0), \quad \lambda \in \Pi = (-\pi, \pi]
\]
holds at some true parameter values \((\sigma_0^2, \theta_0)\). For Whittle estimation, we make some common assumptions on the model class \( F \) (cf. Hannan, 1973; Fox and Taqqu, 1986). In (4.2), the spectral densities are assumed to be positive on \( \Pi \) and identifiable (i.e., \( \{ \lambda : g(\lambda; \sigma_1^2, \theta_1) \neq g(\lambda; \sigma_2^2, \theta_2) \} \) has positive Lebesgue measure for \((\sigma_1^2, \theta_1) \neq (\sigma_2^2, \theta_2)\)). Additionally, we suppose these densities satisfy Kolmogorov’s formula 
\[
(4\pi)^{-1} \int_{0}^{\pi} \log[g(\lambda; \sigma^2, \theta)] = \log[\sigma^2/(2\pi)] \quad \text{(i.e., treating } \sigma^2 \text{ as the innovation variance in a linear time series representation)};
\]
see the Whittle likelihood (4.6) below. We note that the space of parameters \((\sigma^2, \theta)\) for (4.2) is \((0, \infty) \times \Theta\), with \( \Theta \subset \mathbb{R}^p \). Here we do not assume that \( \Theta \) is necessarily compact, unlike the previously mentioned works with Whittle estimation. We simply assume \( \Theta \) contains some closed neighborhood around \( \theta_0 \). Common spectral density models of the form (4.2), and satisfying the conditions mentioned above, include the previously mentioned fractional Gaussian processes (Mandelbrot and van Ness, 1968) with spectral densities
\[
g(\lambda; H, \sigma^2) = \frac{4\sigma^2 \Gamma(2H - 1)}{(2\pi)^{2H+2}} \cos(\pi H - \pi/2) \sin^2(\lambda/2) \sum_{k=-\infty}^{\infty} \frac{|\lambda/(2\pi) + k|^{-1-2H}}{|k^{-2H}|}, \quad \lambda \in \Pi, \quad (4.4)
\]
for \( 1/2 < H < 1 \), and FARIMA processes (Adenstedt, 1974; Granger and Joyeux, 1980; Hosking, 1981) with spectral density
\[
g(\lambda; d, \rho, \varrho, \sigma^2) = \frac{\sigma^2}{2\pi} \left| 1 - e^{i\lambda} \right|^{-2d} \left| \sum_{j=0}^{p} \rho_j \left(e^{i\lambda}\right)^j \right|^2 \left| \sum_{j=0}^{q} \varrho_j \left(e^{i\lambda}\right)^j \right|^2, \quad \lambda \in \Pi, \quad (4.5)
\]
based on parameters \( 0 < d < 1/2, \rho = (\rho_1, \ldots, \rho_p), \varrho = (\varrho_1, \ldots, \varrho_q) \) with \( \rho_0 = \varrho_0 = 1 \). These models, based on their parameters, fulfill (4.1) with exponents \( 2H - 1 \) and \( 2d \), respectively.

For fitting the model to data, Whittle estimation (Whittle, 1953) seeks to determine the
parameter values at which the theoretical distance measure
\[ W(\sigma^2, \theta) = (4\pi)^{-1} \int_0^\pi \left\{ \log g(\lambda; \sigma^2, \theta) + \frac{g(\lambda)}{g(\lambda; \sigma^2, \theta)} \right\} d\lambda \]  
(4.6)

achieves its minimum (cf. Dzhaparidze, 1986). Under mild conditions, the true parameter values
\( (\sigma_0^2, \theta_0) \) are then determined as the unique solutions to the score equations
\[ \frac{\partial W(\sigma, \theta)}{\partial \sigma, \theta} = 0 \]  
and
\[ \int_0^\pi \frac{\partial f^{-1}(\lambda; \theta)}{\partial \theta} g(\lambda)d\lambda = 0_p \quad \text{and} \quad 2\int_0^\pi f^{-1}(\lambda; \theta)g(\lambda)d\lambda = \sigma^2, \]  
(4.7)

where \( f^{-1}(\lambda; \theta) = 1/f(\lambda; \theta) \). Parameter estimation involves solving an approximation to (4.7) whereby the periodogram \( I_n \) is substituted as an estimator of the true spectral density \( g \). Based on the data \( X_1, \ldots, X_n \), the periodogram is defined as
\[ I_n(\lambda) = \frac{1}{2\pi n} \left( \sum_{t=1}^n (X_t - \bar{X}_n)e^{-it\lambda} \right)^2, \lambda \in \Pi, \]
with \( \iota = \sqrt{-1} \) and \( \bar{X}_n = n^{-1} \sum_{t=1}^n X_t \).

We then formally define the Whittle estimators \( \hat{\theta}_n \) of \( \theta \in \Theta \) as a solution of the periodogram-based estimating functions
\[ \ell_n(\theta) = 0_p, \quad \ell_n(\theta) = \frac{2\pi n}{N} \sum_{m=1}^N \frac{\partial f^{-1}(\lambda_m; \theta)}{\partial \theta} I_n(\lambda_m), \]  
(4.8)

defined by a Riemann integral approximation to (4.7) that substitutes \( I_n \) evaluated at discrete Fourier frequencies \( \lambda_m = 2\pi m/n \) for \( m = 1, 2, \ldots, N = \lfloor (n-1)/2 \rfloor \). The corresponding Whittle estimator of \( \sigma^2 \) is then given by
\[ \hat{\sigma}_n^2 = 2 \cdot \frac{2\pi}{n} \sum_{m=1}^N f^{-1}(\lambda_m; \hat{\theta}_n)I_n(\lambda_m). \]  
(4.9)

Alternatively, one could analogously define estimators \( \tilde{\theta}_n, \tilde{\sigma}_n^2 \) with continuous integrals
\[ \int_0^\pi \frac{\partial f^{-1}(\lambda; \tilde{\theta}_n)}{\partial \theta} I_n(\lambda)d\lambda = 0_p, \quad \tilde{\sigma}_n^2 = 2 \int_0^\pi f^{-1}(\lambda; \tilde{\theta}_n)I_n(\lambda)d\lambda, \]
without changing the main large-sample results of the next section. We use the discrete integral version \( \ell_n(\theta) \), however, as this is computationally simpler in practice and more useful for motivating a FDB version Whittle estimators in the next section.
4.2.2 A frequency domain bootstrap method for Whittle estimation

The goal here is to prescribe a bootstrap estimator for the distribution $P(T_n \leq x)$, $x \in \mathbb{R}^p$, of the Whittle estimators $T_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$, centered at the true parameter values. We define a bootstrap versions $\hat{\theta}_n^*$ and $T_n^*$ according to the following resampling mechanism. In the following, let $P_*$ and $E_*$ denote the bootstrap distribution and expectation induced by the resampling scheme, conditional on the observed data $X_1, \ldots, X_n$.

**Bootstrap Procedure:**

1. Define spectral density estimates $\hat{g}_n(\lambda_m) \equiv g(\lambda_m; \hat{\sigma}_n^2, \hat{\theta}_n)$ at discrete Fourier frequencies $m = 1, 2, \ldots, N$ using Whittle estimates $(\hat{\theta}_n, \hat{\sigma}_n^2)$ from the original data.

2. Studentize periodogram ordinates $\hat{\varepsilon}_m = I_n(\lambda_m)/\hat{g}_n(\lambda_m)$, $m = 1, 2, \ldots, N$, and then re-scale as $\tilde{\varepsilon}_m = \hat{\varepsilon}_m / \hat{\varepsilon}$, where $\hat{\varepsilon} = N^{-1} \sum_{m=1}^{N} \hat{\varepsilon}_m$.

3. Randomly sample $\varepsilon_1^*, \ldots, \varepsilon_N^*$ from the empirical distribution of $\{\hat{\varepsilon}_m; m = 1, 2, \ldots, N\}$.

4. Define a bootstrap version of periodogram ordinates as $I_n^*(\lambda_m) = \varepsilon_m^* \hat{g}_n(\lambda_m)$ for $m = 1, 2, \ldots, N$.

5. Define bootstrap versions of Whittle estimators $\hat{\theta}_n^*$ as the solution of

$$\ell_n^*(\theta) - E_* \ell_n^*(\hat{\theta}_n) = 0_p, \quad (4.10)$$

where

$$\ell_n^*(\theta) \equiv \frac{2\pi}{n} \sum_{m=1}^{N} \frac{\partial f^{-1}(\lambda_m; \theta)}{\partial \theta} I_n^*(\lambda_m), \quad E_* \ell_n^*(\hat{\theta}_n) \equiv \frac{2\pi}{n} \sum_{m=1}^{N} \frac{\partial f^{-1}(\lambda_m; \hat{\theta}_n)}{\partial \theta} \hat{g}_n(\lambda_m).$$

6. Define $T_n^* = \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ as the bootstrap version of $T_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$ so that $P_*(T_n^* \leq x)$ is the bootstrap estimator of $P(T_n \leq x)$, $x \in \mathbb{R}^p$.

We make a few comments on the above FDB procedure. As mentioned in Section 4.1, the periodogram scaling scheme is different from the general one used in the original FDB of Dahlhaus and Janas (1996) under SRD. Again, for scaling, those authors used a nonparametric
(kernel) estimator of the spectral density having uniform consistency properties on the whole spectrum \((0, \pi]\). An analogous theory of such estimators does not seemingly exist under LRD. However, the Whittle estimation problem allows re-scaling the periodogram differently, in a way that is valid under either SRD or LRD. In Step 2, the rescaling of studentized periodogram ordinates serves to eliminate unnecessary bias in the FDB re-creation (Dahlhaus and Janas, 1996), as the studentized periodogram at a fixed frequency \(0 < \lambda < \pi\) has an asymptotic exponential distribution with mean parameter under SRD/LRD (cf. Yajima, 1989). Steps 5 and 6 aim to reproduce the structural relationship between the Whittle estimators \(\hat{\theta}_n\) and the true parameters \(\theta_0\) at the level of the bootstrap. As the Whittle estimator is a type of M-estimator, issues of “centering” are important in bootstrap estimators (cf. Lahiri, 2003, sec. 4.3). Namely, the Whittle estimators \(\hat{\theta}_n\) solve the estimating functions \(\ell_n(\theta) = 0\) from (4.8) and, at the true parameters \(\theta_0\), the expected value of these estimating functions is approximately zero by (4.7),

\[
E\ell_n(\theta_0) \approx \int_0^\pi \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta} g(\lambda) d\lambda = 0_p,
\]

which identifies \(\theta_0\). The FDB version \(\hat{\theta}_n^*\) solve a modified equation (4.10) where \(\ell_n^*(\theta) - E_*\ell_n^*(\hat{\theta}_n)\) has expectation zero at the original Whittle estimators \(\hat{\theta}_n\). In this way, the relationship between \(\hat{\theta}_n^*, \hat{\theta}_n\) mimics that of \(\hat{\theta}_n, \theta_0\), with the advantage that the Whittle estimator \(\hat{\theta}_n\) in hand becomes an appropriate centering of \(\hat{\theta}_n^*\) in defining \(T_n^*\). This idea follows a suggestion of Shorack (1982) in bootstrapping M-estimators in linear regression models, which has been shown to be valid for reproducing M-estimators with the time series block bootstrap (cf. Lahiri, 2003, Theorem 4.2). In contrast, Dahlhaus and Janas (1996) use a different approach by which \(\hat{\theta}_n^*\) would require centering by the solution to \(E_*\ell_n^*(\theta) = 0_p\), which has a slight disadvantage of requiring more root-finding steps.

The next sections establishes the validity of the FDB procedure for a broad class of linear time processes and spectral model families (4.2).
4.3 Main Distributional Results

For clarity, Section 4.3.1 describes the assumptions on the time process and the spectral model class (4.2) considered for Whittle estimation. Section 3.2 then provides the main distributional results on Whittle estimators and their FDB approximations.

4.3.1 Assumptions

We will refer the following conditions.

L The real-valued process \( \{X_t\} \) has a moving average representation as in (4.3) with respect to independent, identically distributed (iid) innovations \( \{\varepsilon_t\} \) with \( \mathbb{E}(\varepsilon_t) = 0, \mathbb{E}(\varepsilon_t^2) = \sigma_0^2 \) and \( \mathbb{E}(\varepsilon_t^8) < \infty \).

A.1 \( \int_0^\pi \log f(\lambda; \theta) d\lambda \equiv 0 \) is twice differentiable in \( \theta \in \Theta \) around the true value \( \theta_0 \) under the integral sign.

A.2 \( f(\lambda; \theta) \) is continuous at \( (\lambda, \theta) \in \Pi \times \Theta \), with \( \lambda \neq 0 \), and \( f^{-1}(\lambda; \theta) \) is continuous on \( \Pi \times \Theta \).

A.3 \( \partial f^{-1}(\lambda; \theta)/\partial \theta_j \) and \( \partial^2/\partial \theta_j \partial \theta_k f^{-1}(\lambda; \theta) \) are continuous on \( \Pi \times \Theta \).

There exists \( 0 < d < 1/2 \) and \( C(\theta_0) > 0 \) depending on the true value \( \theta_0 \) such that

A.4 \( \|\partial/\partial \theta_j f^{-1}(\lambda; \theta_0)\| \leq C(\theta_0)|\lambda|^{2d}, \|\partial^2/\partial \theta_j \partial \lambda f^{-1}(\lambda; \theta_0)\| \leq C(\theta_0)|\lambda|^{2d-1}, \)

and \( \|\partial^2/\partial \theta_j \partial \theta_k f^{-1}(\lambda; \theta)\| \leq C(\theta_0)|\lambda|^{2d} \) for \( \lambda \in \Pi \) with \( \lambda \neq 0 \);

A.5 \( |f(\lambda; \theta_0)| \leq C(\theta_0)|\lambda|^{-2d}, |f^{-1}(\lambda; \theta)| \leq C(\theta_0)|\lambda|^{2d} \) and for some \( \eta \in (0, 1/2) \) with \( 2d+\eta < 1 \), there exists \( \delta \equiv \delta(\eta) > 0 \) and \( C(\eta) > 0 \) such that

\[
\sup_{\lambda \neq 0, \lambda \in \Pi} \sup_{\theta \in \mathcal{B}} |\lambda|^{2d+\eta} f(\lambda; \theta) \leq C(\eta)
\]

where \( \mathcal{B} \equiv \{\theta \in \Theta : ||\theta - \theta_0|| \leq \delta\} \subset \Theta \).

A.6 The \( p \times p \) matrix

\[
D_0 = \frac{2\pi}{\sigma_0^2} \int_0^\pi \left( \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta} \right) \left( \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta} \right)^T g^2(\lambda) d\lambda
\]

is nonsingular, where \( g(\lambda) = (2\pi)^{-1}\sigma_0^2 f(\lambda; \theta_0) \).
Remark 1. The matrix $D_0$ of A.6 determines the asymptotic covariance of Whittle estimators. Under A.1, this can be rewritten as

$$D_0 = \int_0^\pi \frac{\partial^2 f^{-1}(\lambda; \theta_0)}{\partial \theta \partial \theta^T} g(\lambda) d\lambda;$$

see Fox and Taqqu (1986).

We comment briefly on the conditions above. Condition L entails that a wide class of linear processes may be considered, which need not be causal. In particular, these include processes which could exhibit either SRD (e.g., ARMA processes) or LRD (e.g., fractional Gaussian noise or FARIMA processes); see Beran (1994). Conditions A.1-A.6 are assumptions on the parametric family (4.2) of spectral densities and, for example, are satisfied for the fractional Gaussian (4.4) or FARIMA (4.5) densities under LRD (cf. Fox and Taqqu, 1986). These conditions are similar to, but slightly weaker than, those of Fox and Taqqu (1986) and Giraitis and Surgailis (1990), and allow both SRD $d = 0$ or LRD $d > 0$ processes. In condition A.5, the growth condition on the spectral density kernel $f(\lambda; \theta_0)$ allows for the SRD or LRD behavior in the true spectral density $g(\lambda)$ as described in (4.1).

4.3.2 Distributional results

In Theorem 1, we first provide the asymptotic distribution of Whittle estimators of $\theta \in \Theta$, which is an extension of the results of Fox and Taqqu (1987) and Giraitis and Surgailis (1990) for Gaussian or linear LRD processes. A main difference is that we do not assume the parameter space $\Theta$ is compact. The proof is provided in Section 4.7.

Theorem 1. Under assumptions L, A.1-A.6, as $n \to \infty$,

(i) $\hat{\theta}_n \xrightarrow{p} \theta_0$ and $\hat{\sigma}_n^2 \xrightarrow{p} \sigma_0^2$.

(ii) $\sqrt{n} (\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, \sigma_0^2 D_0^{-1})$.

Remark 2. The proof of Theorem 1 shows that there exists a sequence of Whittle estimators $\hat{\theta}_n$ where

$$P \left( \hat{\theta}_n \text{ is a root of } \ell_n(\theta) = 0 \& ||\hat{\theta}_n - \theta_0|| \leq C n^{1/2} \log^2 n \right) \to 1 \text{ as } n \to \infty,$$
for a constant $C > 0$. Usually, the solutions $\hat{\theta}_n$ are unique. Hence, consistent Whittle estimators are guaranteed to exist on a possibly non-compact parameter space.

Theorem 2 establishes validity of the FDB method for estimating the distribution of Whittle estimators for either SRD or LRD parametric spectral densities. The first part of the theorem guarantees that, in sufficiently large samples, the FDB can be implemented and bootstrap versions of Whittle estimators will exist. The second part of Theorem 2 establishes the consistency of the FDB distributional approximation.

**Theorem 2.** Let $T_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$ and $T^*_n = \sqrt{n}(\hat{\theta}^*_n - \hat{\theta}_n)$. Under assumptions L, A.1-A.6, as $n \to \infty$,

(i) $P_*(\hat{\theta}^*_n)$ solving (4.10) exists and $\|\hat{\theta}^*_n - \hat{\theta}_n\| \leq C_1 n^{-1/2} \log n \xrightarrow{p} 1$, for a constant $C_1 > 0$.

(ii) $\sup_{x \in \mathbb{R}} |P_*(T^*_n \leq x) - P(T_n \leq x)| \xrightarrow{p} 0$.

**Remark 3.** Theorem 2 also holds for studentized versions of Whittle estimators, such as $T_n = \hat{V}_{\theta_n}^{-1/2} \sqrt{n}(\hat{\theta}_n - \theta_0)$ and $T^*_n = \hat{V}_{\theta}^{-1/2} \sqrt{n}(\hat{\theta}_n - \theta_0)$, where

$$\hat{V}_{\theta_n} = 2\pi \left[ \frac{2\pi}{n} \sum_{j=1}^{N} \left( \frac{\partial f^{-1}(\lambda_j; \hat{\theta}_n)}{\partial \theta} \right) \left( \frac{\partial f^{-1}(\lambda_j; \hat{\theta}_n)}{\partial \theta} \right)^T f^2(\lambda_j; \hat{\theta}_n) \right]^{-1},$$

is an estimator of the limiting variance $\sigma_0^2 D_0^{-1}$ from Theorem 1 and $\hat{V}_{\theta_n}$ denotes a bootstrap version. The performances of different forms of studentization and non-studentization are considered in the simulation studies of Section 4.4.

With regard to the Whittle estimator $\hat{\sigma}_n^2$ from (4.9) of the innovation variance parameter $\sigma^2$, the FDB method here is generally not valid for distribution estimation. The situation is the same for the FDB of Dahlhaus and Janas (1996) under SRD. The issue is that the resampling mechanism treats the periodogram ordinates $\{I_n(\lambda_m) : m = 1, \ldots, N\}$ as independent across all discrete Fourier frequencies, which is not generally true and can distort the ability of the FDB to correctly capture the variance of estimators in the frequency domain. For Whittle estimators of
the parameters $\theta$, the FDB approximation is valid, as these parameters are identified by integral (4.7) of the spectral density $g$ having an expectation of zero. In contrast, the parameter $\sigma^2$ is determined by non-zero spectral mean (4.7), so that the FDB approximation generally fails because there are process moments in the variance of $\hat{\sigma}^2$ that FDB does not capture, unless the fourth order cumulant of the innovation process is zero in condition L (e.g., independent Gaussian innovations). See Dahlhaus and Janas (1996) or Sec. 9.2 of Lahiri (2003) for further details.

4.4 Simulation Studies

This section investigates performance of the FDB distributional approximation for Whittle estimators from linear LRD processes. In the following simulation study, we consider several types of FARIMA($1,d,1$) processes (Granger and Joyeux, 1980; and Hosking, 1981) defined by

$$\phi(B)(1 - B)^d X_t = \psi(B) \varepsilon_t \tag{4.11}$$

where $B$ is a backshift operator (i.e., $B X_t = X_{t-1}$), $\{\varepsilon_t\}$ are iid mean zero innovation variables, long-memory parameter is given by $d \in [0, 1/2)$, and $\phi(z) = 1 - \phi z$ and $\psi(z) = 1 + \psi z$ denote autoregressive and moving average polynomials, respectively, with parameters $|\phi|, |\psi| < 1$. In the framework of (4.11), we separately treat FARIMA($0,d,0$), FARIMA($1,d,0$) and FARIMA($0,d,1$) processes. For FARIMA($0,d,0$) models, the innovation distributions are either the standard normal distribution, the t-distribution with 3 degrees of freedom, or a (centered) chi-square distribution with 2 degrees of freedom. We consider a variety of long-memory parameters $d \in \{0.10, 0.25, 0.40\}$, which range from weaker (0.10) to stronger (0.4) forms of LRD, along with several sample sizes $n \in \{100, 250, 500, 1000\}$. Because the results are qualitatively similar for the different innovation processes (as seen in the FARIMA($0,d,0$) models), we shall restrict the presentation of results for other models to standard normal innovations, in either a FARIMA($1,d,0$) model with an AR parameter $\phi_1 = -0.3$ or a FARIMA($0,d,1$) model with an MA parameter $\psi_1 = 0.4$. A main reason for using low dimensional FARIMA models, with this choice of long-memory parameters and sample sizes, is to reliably automate the simulations. For long-memory parameters $d$ too close to the 0 or 1/2 boundaries and with small
sample sizes, there can be a significant possibility of obtaining Whittle estimates outside of the parameter space or having numerical convergence issues with re-parameterizations (e.g., small data sets may even exhibit properties of non-stationarity). For the processes and sample sizes considered, we did not encounter issues in numerical optimization (using the \texttt{nlminb} routine in R) though, in a small proportion of instances, Whittle estimates for $d$ fell slightly outside of $(0, 1/2]$, with the chances of this decreasing rapidly for larger sample sizes (see Figure 4.1). This aspect did not adversely affect coverage rates. To approximate the coverage rates of interval estimation procedures, we simulated 4,000 datasets for each model and sample size, and used 1,000 Monte Carlo bootstrap renditions to approximate the FDB distribution of Whittle estimators.

### 4.4.1 Properties of Whittle estimators

To illustrate the finite-sample distribution of Whittle estimators, Figure 4.1 displays the (numerically approximated) distributions of Whittle estimators of the long memory parameter $d \in \{0.10, 0.25, 0.40\}$ from FARIMA$(0, d, 0)$ models with different innovation types and a sample size $n = 500$. Each sampling distribution appears somewhat asymmetric and leptokurtic, with average values that tend to be downward biased, which supports findings of Palma (2007, p. 92). Figure 4.2 shows the distributions of Whittle estimators for the two parameters in the FARIMA$(0, d, 1)$ and FARIMA$(1, d, 0)$ models, which appear more asymmetric and leptokurtic than in the case of the long memory parameter $d$ with FARIMA$(0, d, 0)$ processes (Figure 4.1). For the long memory parameter, the dispersion of Whittle estimates from FARIMA$(0, d, 1)$ processes are wider than those of FARIMA$(1, d, 0)$ processes. Estimates of the long memory parameter for both processes are downward biased and estimates of $\phi_1$ and $\psi_1$ parameters are upward biased. In addition, as $d$ increases to $1/2$ and the strength of the LRD increases, bias problems generally becomes more severe.

In Table 4.1, we report the skewness and kurtosis for Whittle estimates for all FARIMA models considered along with p-values to test against a normal distribution using the D’Agostino test (D’Agostino, 1970) and the Anscombe-Glynn test (Anscombe and Glynn, 1983). As sample sizes increase in the table, the skewness and kurtosis of Whittle estimates for the long-range
dependent parameter $d$ tend to 0 and 3, respectively, indicating normal distributional properties. Table 4.1 also shows that as the number of parameters increase, the asymmetry in the distribution of Whittle estimators of $d$ tends to increase, indicating that convergence to a normal limit can be slow and depend on the underlying LRD process (Mantegna and Stanley, 1994).
Figure 4.1  Approximated distributions of Whittle estimators of $d$ are shown based on size $n = 500$ samples from FARIMA($0, d, 0$) models with $d = 0.4, 0.25, 0.1$ and innovations as standard normal (1st row), Chi-square (2nd row) or t-distributed (3rd row).
Figure 4.2  Approximated distributions of Whittle estimators of the LRD parameter $d$ (1st and 3rd rows), AR parameter $\phi_1 = -0.3$ (2nd row) and MA parameter $\psi_1 = 0.4$ (4th row) are shown based on size $n = 500$ samples from FARIMA$(1,d,0)$ (rows 1-2) or FARIMA$(0,d,1)$ (rows 3-4) models with $d = 0.4, 0.25, 0.1$ and standard normal innovations.
Table 4.1 Skewness (Skew) and kurtosis (Kurt) values are shown for Whittle estimates of \( d = 0.4, 0.25, 0.1 \) from FARIMA\((0, d, 0)\) processes with standard normal, chi-square or student t-3 innovations based on sample sizes \( n = 100, 250, 500, 1000 \). These values are also reported for Whittle estimates in FARIMA\((1, d, 0)\) or FARIMA\((0, d, 1)\) processes with standard normal innovations. Associated p-values (Pval) are provided where the D’Agostino test is for data symmetry and the Anscombe-Glynn test assess if the kurtosis equals 3.

### Long-memory parameter estimates from FARIMA\((0, d, 0)\)

<table>
<thead>
<tr>
<th>( d )</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>Skew</td>
<td>Pval</td>
<td>Kurt</td>
</tr>
<tr>
<td>100</td>
<td>-0.2503</td>
<td>&lt;0.001</td>
<td>4.0667</td>
</tr>
<tr>
<td>250</td>
<td>-0.2045</td>
<td>0.0006</td>
<td>0.0060</td>
</tr>
<tr>
<td>500</td>
<td>-0.0631</td>
<td>0.2827</td>
<td>3.1212</td>
</tr>
<tr>
<td>1000</td>
<td>-0.1023</td>
<td>0.0821</td>
<td>3.1961</td>
</tr>
</tbody>
</table>

### Parameter estimates from FARIMA\((1, d, 0)\)

<table>
<thead>
<tr>
<th>( d )</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>Skew</td>
<td>Pval</td>
<td>Kurt</td>
</tr>
<tr>
<td>100</td>
<td>-1.1085</td>
<td>&lt;0.001</td>
<td>5.6057</td>
</tr>
<tr>
<td>250</td>
<td>-0.3771</td>
<td>&lt;0.001</td>
<td>3.4581</td>
</tr>
<tr>
<td>500</td>
<td>-0.1374</td>
<td>0.0197</td>
<td>3.1785</td>
</tr>
<tr>
<td>1000</td>
<td>-0.1554</td>
<td>0.0084</td>
<td>3.2433</td>
</tr>
</tbody>
</table>

### Parameter estimates from FARIMA\((0, d, 1)\)

<table>
<thead>
<tr>
<th>( d )</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>Skew</td>
<td>Pval</td>
<td>Kurt</td>
</tr>
<tr>
<td>100</td>
<td>1.1155</td>
<td>&lt;0.001</td>
<td>5.2776</td>
</tr>
<tr>
<td>250</td>
<td>0.3408</td>
<td>&lt;0.001</td>
<td>3.4843</td>
</tr>
<tr>
<td>500</td>
<td>0.1724</td>
<td>0.0035</td>
<td>3.1609</td>
</tr>
<tr>
<td>1000</td>
<td>0.0787</td>
<td>0.1806</td>
<td>2.9417</td>
</tr>
</tbody>
</table>

### Parameter estimates from FARIMA\((0, 0, 1)\)

<table>
<thead>
<tr>
<th>( d )</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>Skew</td>
<td>Pval</td>
<td>Kurt</td>
</tr>
<tr>
<td>100</td>
<td>-0.1028</td>
<td>0.0005</td>
<td>3.1366</td>
</tr>
<tr>
<td>250</td>
<td>-0.1294</td>
<td>0.0281</td>
<td>3.0997</td>
</tr>
<tr>
<td>500</td>
<td>-0.1004</td>
<td>0.0878</td>
<td>3.1642</td>
</tr>
<tr>
<td>1000</td>
<td>-0.0423</td>
<td>0.4712</td>
<td>3.0477</td>
</tr>
</tbody>
</table>

### Parameter estimates from FARIMA\((0, 1, 1)\)

<table>
<thead>
<tr>
<th>( d )</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>Skew</td>
<td>Pval</td>
<td>Kurt</td>
</tr>
<tr>
<td>100</td>
<td>-0.2356</td>
<td>0.0001</td>
<td>3.4539</td>
</tr>
<tr>
<td>250</td>
<td>-0.1767</td>
<td>0.0282</td>
<td>3.3550</td>
</tr>
<tr>
<td>500</td>
<td>-0.0835</td>
<td>0.1551</td>
<td>3.1304</td>
</tr>
<tr>
<td>1000</td>
<td>-0.1652</td>
<td>0.0052</td>
<td>3.0942</td>
</tr>
</tbody>
</table>
4.4.2 Numerical studies of coverage accuracy in estimation

For each FARIMA process and sample size, we considered three different versions of the FDB for setting confidence intervals (CIs) for individual, or univariate, parameters (i.e., $d$, $\phi_1$, $\psi_1$) based on Whittle estimators, involving non-studentization and two forms of studentization. The studentization-types differed by using either the estimated spectral density model or the periodogram to formulate a variance estimate. These FDB methods were compared against two versions of normal approximation-based intervals, also using the two same forms of studentization. For each method, we computed the coverage rates of nominal 95% two-sided and one-sided (both lower and upper) CIs for Whittle parameters. For two-sided FDB intervals, we also considered forming symmetric and asymmetric intervals. Finally, in the FARIMA$(1,d,0)$ and FARIMA$(0,d,1)$ models, we also calculated coverage rates of joint 95% confidence regions (CRs) regions for the model parameters $(d, \phi_1)$ or $(d, \psi_1)$.

Table 4.2 provides a complete listing of the estimation methods considered, described in...
more detail below (Section 4.2.1). The simulation results on coverage accuracy are then presented and summarized in Section 4.2.2

4.4.2.1 Description of estimation methods

Non-studentized FDB Intervals: A 95% two-sided non-studentized, symmetric FDB confidence interval for a univariate spectral model parameter \( \theta \in \mathbb{R} \) is defined as \( \hat{\theta}_n \pm n^{-1/2}a_{0.95}^* \), using the corresponding Whittle estimator \( \hat{\theta}_n \) and a bootstrap quantile \( a_{0.95}^* \) defined as

\[
P_* (|T_n^*| \leq a_{0.95}^*) = 0.95.
\]

for \( T_n^* = n^{1/2}(\hat{\theta}_n^* - \hat{\theta}_n) \).

A 95% two-sided non-studentized, asymmetric FDB confidence interval for \( \theta \in \mathbb{R} \) is defined as \( (\hat{\theta}_n - n^{-1/2}q_{0.975}^*, \hat{\theta}_n - n^{-1/2}q_{0.025}^*) \) using lower 2.5% and 97.5% bootstrap quantiles of \( T_n^* \), \( q_{0.025}^* \) and \( q_{0.975}^* \), satisfying

\[
P_* (q_{0.025}^* \leq T_n^* \leq q_{0.975}^*) = 0.95.
\]

Upper and lower one-sided 95% non-studentized FDB confidence intervals for \( \theta \) are defined as \( (-\infty, \hat{\theta}_n - n^{-1/2}q_{0.05}^*) \) and \( (\hat{\theta}_n - n^{-1/2}q_{0.95}^*, \infty) \) using lower quantiles of \( T_n^* \).

Studentized FDB and Normal Approximation Intervals: For the Whittle estimators \( \hat{\theta}_n \) of a parameter vector \( \theta = (\theta_1, \ldots, \theta_p)' \) in (4.2), we consider estimating the corresponding limiting covariance matrix \( V \equiv \sigma_0^2 D_0^{-1} \) from Theorem 1(ii) in two ways. One approach to studentization involves using the estimated spectral density along with the estimating functions (4.7) as

\[
\hat{V}_n \equiv 2\pi \left[ \frac{2\pi}{n} \sum_{j=1}^N \left( \frac{\partial f^{-1}(\lambda_j; \hat{\theta}_n^*)}{\partial \theta} \right) \left( \frac{\partial f^{-1}(\lambda_j; \hat{\theta}_n^*)}{\partial \theta} \right)^T f^2(\lambda_j; \hat{\theta}_n^*) \right]^{-1}.
\]

The other approach uses the estimating functions with the periodogram in place of the estimated spectral density,

\[
\hat{V}_{n,P} \equiv \left[ \frac{(2\pi)^2}{n\sigma_0^2} \sum_{j=1}^N \left( \frac{\partial f^{-1}(\lambda_j; \hat{\theta}_n^*)}{\partial \theta} \right) \left( \frac{\partial f^{-1}(\lambda_j; \hat{\theta}_n^*)}{\partial \theta} \right)^T I_n(\lambda_j) \right]^{-1}.
\]
Let $\theta_i$ denote an element of $\theta$ with corresponding Whittle estimator $\hat{\theta}_{i,n}$. Then, a two-sided 95% CI for $\theta_i$ based on a normal approximation with density-studentization is given by $\hat{\theta}_{i,n} \pm 1.96n^{-1/2}/\sqrt{\hat{v}_{i,n}}$ where $\hat{v}_{i,n}$ denotes the $i$th diagonal component of $\hat{V}_n$. The corresponding upper and lower one-sided CIs are $(-\infty, \hat{\theta}_{i,n} + 1.96n^{-1/2}/\sqrt{\hat{v}_{i,n}})$ and $(\hat{\theta}_{i,n} - 1.96n^{-1/2}/\sqrt{\hat{v}_{i,n}}, \infty)$. Two-sided and one-sided CIs for $\theta_i$ based on a normal approximation with periodogram-studentization are analogously defined using $\tilde{V}_n$.

A FDB approximation of $\tilde{T}_{i,n} = \sqrt{n}(\hat{\theta}_{i,n} - \theta_i)/\sqrt{\hat{v}_{i,n}}$ is defined as $\tilde{T}_{i,n}^\ast = \sqrt{n}(\hat{\theta}_{i,n} - \hat{\theta}_{i,n})/\sqrt{\hat{v}_{i,n}^\ast}$ where $\hat{v}_{i,n}^\ast$ denotes the $i$th diagonal component of $\hat{V}_n^\ast$ defined by substituting $\hat{\theta}_{i,n}^\ast$ for $\hat{\theta}_{i,n}$ in $\hat{V}_n$. Likewise, the FDB approximation of $\tilde{T}_{i,n,P} = \sqrt{n}(\hat{\theta}_{i,n} - \theta_i)/\sqrt{\hat{v}_{i,n,P}}$ is defined as $\tilde{T}_{i,n,P}^\ast = \sqrt{n}(\hat{\theta}_{i,n}^\ast - \hat{\theta}_{i,n})/\sqrt{\hat{v}_{i,n,P}^\ast}$ where $\hat{v}_{i,n,P}^\ast$ denotes the $i$th diagonal component of $\hat{V}_{n,P}^\ast$ defined by substituting $\hat{\theta}_{i,n}^\ast, \hat{\sigma}_{n}^2, \hat{I}_{n}$ for $\hat{\theta}_{i,n}, \hat{\sigma}_{n}^2, \hat{I}_{n}$ in $\hat{V}_{n,P}^\ast$.

Then, a 95% two-sided, density-studentized, symmetric FDB confidence interval for $\theta_i$ is defined as $\hat{\theta}_{i,n} \pm n^{-1/2}/\sqrt{\hat{v}_{i,n}^\ast} a_{0.95}$ using a bootstrap quantile $a_{0.95}$ defined as

$$P^\ast \left( |\tilde{T}_{i,n}^\ast| \leq a_{0.95} \right) = 0.95.$$ 

A 95% two-sided, density-studentized, asymmetric FDB confidence interval is defined as $(\hat{\theta}_{i,n} - n^{-1/2}/\sqrt{\hat{v}_{i,n} q_{0.975}}, \hat{\theta}_{i,n} - n^{-1/2}/\sqrt{\hat{v}_{i,n} q_{0.975}})$ using lower 2.5% and 97.5% bootstrap quantiles of $\tilde{T}_{i,n}^\ast$ satisfying

$$P^\ast \left( q_{0.025} \leq \tilde{T}_{i,n}^\ast \leq q_{0.975} \right) = 0.95.$$ 

Upper and lower one-sided 95% density-studentized FDB confidence intervals for $\theta_i$ are defined as $(-\infty, \hat{\theta}_{i,n} - n^{-1/2}/\sqrt{\hat{v}_{i,n} q_{0.05}})$ and $(\hat{\theta}_{i,n} - n^{-1/2}/\sqrt{\hat{v}_{i,n} q_{0.05}}, \infty)$ using the same lower quantiles of $\tilde{T}_{i,n}^\ast$.

Periodogram-studentized one-sided CIs, or two-sided (symmetric/asymmetric) CIs, for $\theta_i$ are analogously defined using $\tilde{T}_{i,n,P}$ and the appropriate quantiles of $\tilde{T}_{i,n,P}^\ast$.

**Joint Confidence Regions (CRs):** Joint CRs are determined as $\{ \theta : n(\hat{\theta}_n - \theta)^T \hat{V}_n^{-1}(\hat{\theta}_n - \theta) \leq c_1 \}$ or $\{ \theta : n(\hat{\theta}_n - \theta)^T \hat{V}_n^{-1,P}(\hat{\theta}_n - \theta) \leq c_2 \}$, depending on the form of the studentization. In normal approximations, the cut-offs $c_1 = c_2$ are determined by chi-square quantiles. In FDB
approximations, \( c_1 \) and \( c_2 \) are estimated by the quantiles of the bootstrap versions 
\[ n(\hat{\theta}_n^* - \hat{\theta}_n)\mathbf{V}_n^{-1}(\hat{\theta}_n - \hat{\theta}_n) \text{ or } n(\hat{\theta}_n^* - \hat{\theta}_n)\mathbf{V}_{n,P}^{-1}(\hat{\theta}_n - \hat{\theta}_n), \]
respectively.

### 4.4.2.2 Simulation summary

Figures 4.3-4.10 to follow summarize the coverage results, using the abbreviated names of methods listed in Table 3.2. Section 4.8 provides tables reporting the exact numerical values used in these figures, along with average CI lengths or CR volumes.

In Figures 4.3-4.5 present the coverage accuracies of CIs for the long-memory parameter \( d \) from FARIMA\((0, d, 0)\) processes, where each figure corresponds to a different innovation type. In FARIMA\((0, d, 0)\) models, non-studentized FDB intervals for \( d \) exactly match the density-studentized FDB intervals (i.e., in one-sided or two-sided cases), because the estimated covariance matrices with spectral density estimates do not depend on the parameter value \( d \) and so \( \mathbf{V}_n = \mathbf{V}_n^* \) in this case. Figures 4.6-4.10 show coverage results for the FARIMA\((1, d, 0)\) and FARIMA\((0, d, 1)\) processes. In these cases, there exists no equivalences between non-studentized FDB intervals and density-studentized FDB intervals.

All figures show that periodogram-studentized FDB CIs always performed better than their normal approximation counterparts. Since Whittle estimators can have asymmetric sampling distributions, the two-sided asymmetric FDB CIs tended to have better accuracy than the symmetric versions, especially in smaller sample sizes. As sample sizes increased, the performances of FDB methods and normal approximations became similar, particularly for two-sided CIs with density-studentization or joint CRs. We may highlight some further observations about the results in a process-wise manner.

**FARIMA\((0, d, 0)\) models:**

- One-sided non-studentized FDB intervals (or equivalently here density-studentized intervals) performed better than the one-sided normal approximation versions.

**FARIMA\((1, d, 0)\) models:**
• For one-sided CIs with density-studentization, the FDB was better than the normal approximation version.

• Non-studentized FDB CIs outperformed all corresponding normal approximations.

**FARIMA**(0, d, 1) **models:**

• All one- or two-sided FDB CIs provided better performance than corresponding normal approximation CIs.

In summary, the FDB CIs with Whittle estimators generally improved upon normal approximation versions. This is especially true in the one-sided interval cases where the normal approximation can fail in reflecting the asymmetry in the sampling distributions of Whittle estimators. The non-studentized FDB intervals often performed better than their normal approximation counterparts as well as other FDB versions, especially in the larger spectral density models.
Figure 4.3  Differences of coverage rates (observed - nominal) for 95% CIs for $d = 0.4, 0.25, 0.1$ in FARIMA(0, $d$, 0) models with standard normal innovations; these are indicated by sample size $n = 100, 250, 500, 1000$ and one-sided (lower/upper) or two-sided cases. FBD (B) and normal approximation (Nor) methods are divided by studentization type (periodogram (P) or estimated density (D)); FDB intervals can be non-studentized (NS) and, in two-sided cases, either symmetric (S) or asymmetric (A).

$d = 0.4$

$d = 0.25$

$d = 0.1$
Figure 4.4 Differences of coverage rates (observed - nominal) for 95% CIs for $d = 0.4, 0.25, 0.1$ in FARIMA($0,d,0$) models with chi-square innovations; these are indicated by sample size $n = 100, 250, 500, 1000$ and one-sided (lower/upper) or two-sided cases. FBD (B) and normal approximation (Nor) methods are divided by studentization type (periodogram (P) or estimated density (D)); FDB intervals can be non-studentized (NS) and, in two-sided cases, either symmetric (S) or asymmetric (A).
Figure 4.5  Differences of coverage rates (observed - nominal) for 95% CIs for $d = 0.4, 0.25, 0.1$ in FARIMA($0,d,0$) models with student-t innovations; these are indicated by sample size $n = 100, 250, 500, 1000$ and one-sided (lower/upper) or two-sided cases. FBD (B) and normal approximation (Nor) methods are divided by studentization type (periodogram (P) or estimated density (D)); FDB intervals can be non-studentized (NS) and, in two-sided cases, either symmetric (S) or asymmetric (A).
Figure 4.6 Differences of coverage rates (observed - nominal) for 95% CIs for $d = 0.4, 0.25, 0.1$ in FARIMA($1, d, 0$) models with $\phi_1 = -0.3$ and normal innovations; these are indicated by sample size $n = 100, 250, 500, 1000$ and one-sided (lower/upper) or two-sided cases. FBD (B) and normal approximation (Nor) methods are divided by studentization type (periodogram (P) or estimated density (D)); FDB intervals can be non-studentized (NS) and, in two-sided cases, either symmetric (S) or asymmetric (A).
Figure 4.7 Differences of coverage rates (observed - nominal) for 95% CIs for $\phi_1 = -0.3$ in FARIMA$(1,d,0)$ models with $d = 0.4, 0.25, 0.1$ and normal innovations; these are indicated by sample size $n = 100, 250, 500, 1000$ and one-sided (lower/upper) or two-sided cases. FBD (B) and normal approximation (Nor) methods are divided by studentization type (periodogram (P) or estimated density (D)); FDB intervals can be non-studentized (NS) and, in two-sided cases, either symmetric (S) or asymmetric (A).
Figure 4.8  Differences of coverage rates (observed - nominal) for 95% CIs for $d = 0.4, 0.25, 0.1$ in FARIMA$(0, d, 1)$ models with $\psi_1 = 0.4$ and normal innovations; these are indicated by sample size $n = 100, 250, 500, 1000$ and one-sided (lower/upper) or two-sided cases. FBD (B) and normal approximation (Nor) methods are divided by studentization type (periodogram (P) or estimated density (D)); FDB intervals can be non-studentized (NS) and, in two-sided cases, either symmetric (S) or asymmetric (A).
Figure 4.9 Differences of coverage rates (observed - nominal) for 95% CIs for $\psi_1 = 0.4$ in FARIMA($0, d, 2$) models with $d = 0.4, 0.25, 0.1$ and normal innovations; these are indicated by sample size $n = 100, 250, 500, 1000$ and one-sided (lower/upper) or two-sided cases. FBD (B) and normal approximation (Nor) methods are divided by studentization type (periodogram (P) or estimated density (D)); FDB intervals can be non-studentized (NS) and, in two-sided cases, either symmetric (S) or asymmetric (A).
Figure 4.10 Differences of coverage rates (observed - nominal) for joint 95% CRs for $(d, \phi_1 = -0.3)$ in FARIMA$(1, d, 0)$ or for $(d, \psi_1 = 0.4)$ in FARIMA$(0, d, 1)$ models with $d = 0.4, 0.25, 0.1$ and normal innovations; these are indicated by sample size $n = 100, 250, 500, 1000$. FBD (B) and normal approximation (Nor) methods are divided by studentization type (periodogram (P) or estimated density (D)).
4.5 Concluding Remarks

Whittle estimation is a common technique for fitting parametric spectral density models to time series, which is useful for estimating the underlying time covariance structure (e.g., for purposes of prediction). Whittle estimators are known to have normal limit distributions for both weakly and strongly dependent linear time series (cf. Fox and Taqqu, 1986; Giraitis and Surgailis, 1990). However, for time processes which exhibit forms of strong or long-range dependence (LRD), the convergence of Whittle estimators to their normal limits can be slow and sampling distributions in moderate sample sizes may be more asymmetric than normal distributions. This implies room for improving normal approximations in calibrating confidence intervals with Whittle estimators.

We have shown that a frequency domain bootstrap (FDB) method provides valid estimation of the sampling distribution of Whittle estimators for a broad class of linear (but not necessarily causal) time processes, which could exhibit either weak or strong time dependence. In particular, this class includes many popular time series models for LRD (e.g., FARIMA models). The FDB method has the advantage that it presupposes no knowledge on the part of the practitioner with regard to the exact dependence structure of the data-generating mechanism or the full joint distribution of time series observations. Simulations have indicated that the FDB-based interval estimators generally outperform normal approximation-based ones in terms of coverage accuracy. Additionally, non-studentized forms of the FDB, which are computationally simple to obtain, appear to produce good results over a range of dependence strengths and parameter types in spectral density models.

As mentioned in the Introduction, the FDB method here is a type of LRD extension of a bootstrap method originally proposed by Dahlhaus and Janas (1996) for short-range dependent processes. However, this extension is not a full one and the FDB here is particular to Whittle estimation. The difference lies in the type of spectral estimators available for scaling, and thereby stabilizing the variances of, periodogram ordinates prior to resampling these. Under weak dependence, nonparametric kernel estimators of the spectral density are available, which are uniformly consistent on the entire spectrum, and this is a key component in the original FDB
as well as in other bootstrap versions in the frequency domain (cf. Franke and Härdle, 1992; Kreiss and Paparoditis, 2003; Jentsch and Kreiss, 2010). Currently, a comparable theory for spectral density estimators does not seemingly exist under LRD. However, if such estimators were to become available, the theoretical results here could be easily modified to justified a new version of the FDB under weak or strong dependence, based on similar nonparametric re-scaling of periodogram ordinates. Under weak time dependence, recent advances have been made in developing bootstrap re-creations of time series through frequency domain resampling (cf. Jentsch and Kreiss, 2010; Kirch and Politis, 2011; Kreiss, Paparoditis and Politis, 2011). The FDB method here could also potentially be extended in similar, useful directions under LRD with appropriate nonparametric spectral density estimators.

4.6 Preliminary results and lemmas

We require some additional notation and preliminary results to establish the two main theorems in this paper. For clarity, the preliminary results are divided into Sections 4.6.1 and 4.6.2 in following, which provide lemmas needed for proving Theorems 1 and 2, respectively. The proofs of Theorems 1 and 2 are provided in the next Section 4.7.

In the following, $C$ or $C(\cdot)$ denote generic constants that may depend on arguments (if any), but do not depend on $n$, $N = \lfloor (n - 1)/n \rfloor$ or any discrete Fourier frequencies $\{\lambda_m\}_{m=1}^N$. Unless indicated otherwise, all limits will denote convergence as $n \to \infty$.

Define the mean corrected discrete Fourier transforms as

$$d_{nc}(\lambda) = \sum_{t=1}^{n} (X_t - \mu) e^{-\imath t\lambda}, \quad \lambda \in \Pi, \quad \imath = \sqrt{-1}.$$ 

Note that

$$I_{nc}(\lambda) = |d_{nc}(\lambda)|^2 = \frac{1}{2\pi n} d_{nc}(\lambda)d_{nc}(-\lambda),$$

and $I_{nc}(\lambda_m) = I_n(\lambda_m)$ for $m = 1, 2, \ldots, N$. Let $H_n(\lambda) = \sum_{t=1}^{n} e^{-\imath t\lambda}, \lambda \in \mathbb{R}$, and write $K_n(\lambda) = (2\pi n)^{-1}|H_n(\lambda)|^2$ to denote the Fejer kernel (Brookwell and Davis, 1991). The function $K_n(\cdot)$ is nonnegative, even with period $2\pi$ on $\mathbb{R}$, and satisfies $\int_{-\pi}^{\pi} K_n(\lambda) d\lambda = 1$. Let $L_{ns} : \mathbb{R} \to \mathbb{R}$
be the periodic extension of
\[
L_{ns}(\lambda) \equiv \begin{cases} 
  e^{-s}n & \text{if } |\lambda| \leq \frac{e^s}{n}, \\
  \log^s(n|\lambda|) \left| \frac{1}{\lambda} \right| & \text{if } \frac{e^s}{n} < |\lambda| \leq \pi,
\end{cases} \quad \lambda \in \Pi, \quad s = 0, 1,
\]
which is decreasing on \([0, \pi]\) for each \(n \geq 1\) and \(s = 0, 1\), and note that
\[
|H_n(\lambda)| \leq C L_{n0}(\lambda), \quad \lambda \in \mathbb{R}; \quad (4.12)
\]
see Dahlhaus (1983) for more details on the function \(L_{ns}\). In the following, \(\text{Cum}(Y_1, \ldots, Y_n)\) will denote the joint cumulant of generic random variables \(Y_1, \ldots, Y_n\) (Brillinger, 1981).

### 4.6.1 Preliminary results for proving Theorem 1: Lemmas 1-7

**Lemma 1.** Let \(1 \leq j \leq k \leq N \) \((n \geq 3)\), and \(a_1, a_2, \ldots, a_l \in \{\pm \lambda_j, \pm \lambda_k\}\), \(|a_1| \leq \cdots \leq |a_l|\) with \(2 \leq l \leq 8\). Under assumption \(L\), \(A.2\) and \(A.5\), it holds that
\[
(i) \quad |\text{Cum}(d_{nc}(a_1), d_{nc}(a_2))| \leq C |a_1|^{-2d} (|a_2|^{-1} + L_{n1}(a_1 + a_2));
(ii) \quad |\text{Cum}(d_{nc}(a_1), \ldots, d_{nc}(a_l))| \leq C \left\{ |a_l|^{d-1} |a_{l-1}|^{-1/2} + n \log^{l-1}(n) \right\} \prod_{j=1}^{l} |a_j|^{-d},
\]
where \(C\) does not depend on \(n\) or \(1 \leq j, k \leq N\).

**Proof:** Following by modifying Lemma 3 of Nordman and Lahiri (2006). \(\square\)

**Lemma 2.** Under assumptions \(L\) and \(A.5\),
\[
(i) \quad \text{EI}_{nc}(0) \leq C n^{2d} \quad \text{and} \quad \text{EI}_{nc}(\pi) \leq C \quad \text{for some} \ C \quad \text{not depending on} \ n;
(ii) \quad n^{-1}I_{nc}(0) \xrightarrow{p} 0 \quad \text{and} \quad n^{-1}I_{nc}(\pi) \xrightarrow{p} 0 \quad \text{as} \ n \xrightarrow{} \infty.
\]

**Proof:** Note that \(n^{-1}\text{EI}_{nc}(0) \rightarrow 0\) and \(n^{-1}\text{EI}_{nc}(\pi) \rightarrow 0\) follow by part (i), implying \(n^{-1}I_{nc}(0) \xrightarrow{p} 0\) and \(n^{-1}I_{nc}(\pi) \xrightarrow{p} 0\) in part (ii). Hence, it suffices to establish part (i). For any \(\tau > 0\), note that
\[
L_{n0}(\lambda) \leq n^\tau |\lambda|^{-1+\tau}, \quad \lambda \in \Pi. \quad (4.13)
\]
From Lemma 2 of Nordman and Lahiri (2006) and \( g(\pi) < \infty \), we have
\[
\frac{1}{n} \text{EI}_{nc}(\pi) = \frac{1}{2\pi n^2} \text{Cum} (d_{nc}(\pi), d_{nc}(-\pi)) = \frac{1}{2\pi n^2} \{2\pi H_n(0)g(\pi) + o(n)\} = \frac{g(\pi)}{n} + o\left(\frac{1}{n}\right) = o(1).
\]

We next expand \( n^{-1}\text{EI}_{nc}(0) \) as
\[
\frac{1}{n} \text{EI}_{nc}(0) = \frac{1}{2\pi} \text{E}(\bar{X}_n - \mu)^2 = \frac{1}{2\pi n} \sum_{k=-n}^{n} \left(1 - \frac{|k|}{n}\right) r(k) = \frac{1}{2\pi n} \sum_{k=-n}^{n} \left(1 - \frac{|k|}{n}\right) 2 \int_{0}^{\pi} g(\lambda) e^{-ik\lambda} d\lambda.
\]

Since \( n^{-1}H_n(\lambda)H_n(-\lambda) = \sum_{k=-n}^{n} \left(1 - \frac{|k|}{n}\right) e^{-ik\lambda} \), we may rewrite
\[
\frac{1}{n} \text{EI}_{nc}(0) = \frac{1}{n\pi} \int_{0}^{\pi} \frac{1}{n} H_n(\lambda)H_n(-\lambda)g(\lambda)d\lambda.
\]

By A.5, \( g(\lambda) \leq C|\lambda|^{-2d} \) for \( \lambda \in \Pi \). Pick \( \tau = (1 + 2d)/2 < 1 \). Then, by (4.12) and (4.13)
\[
\frac{1}{n} \text{EI}_{nc}(0) \leq \frac{C}{n^{2\tau}} \int_{0}^{\pi} n^{2\tau} |\lambda|^{-2+2\tau} |\lambda|^{-2d} d\lambda = O\left(\frac{n^{2\tau}}{n^{2\tau}}\right) = o(1),
\]
since \( \int_{0}^{\pi} \lambda^{-2+2\tau-2d} d\lambda < \infty \). \( \square \)

**Lemma 3.** Under assumptions L, A.2 and A.5, for \( k \in \mathbb{Z} \),
\[
\hat{r}_n(k) = \frac{1}{n} \sum_{j=1}^{n-k} (X_j - \mu)(X_{j+k} - \mu) \xrightarrow{p} r(k) \quad \text{as} \quad n \to \infty.
\]

**Proof:** See Lemma 4 of Nordman and Lahiri (2006). \( \square \)

**Lemma 4.** If assumptions L, A.2 and A.5 hold and \( h: [0, \pi] \to \mathbb{R} \) is continuous, then
\[
\frac{2\pi}{n} \sum_{m=1}^{N} h(\lambda_m)I_n(\lambda_m) \xrightarrow{p} \int_{0}^{\pi} h(\lambda)g(\lambda)d\lambda \quad \text{as} \quad n \to \infty,
\]
where \( g(\lambda) \equiv \sigma_0^2 f(\lambda; \theta_0)/(2\pi) \).
Proof: We extend $h(\cdot)$ from $[0, \pi]$ to $\Pi \equiv (-\pi, \pi]$ so that $h(\lambda) = h(-\lambda)$, $\lambda \in \Pi$. Let $q_M(\lambda) > 0, \lambda \in \Pi$, be the Cesàro mean of the Fourier series of $h(\lambda)$ taken to $M$ terms for $M$ large, i.e.,

$$q_M(\lambda) \equiv \sum_{k=-M}^{M} C_n h(k) \left(1 - \frac{|k|}{M}\right) e^{ik\lambda},$$

where $C_n h(k) = (2\pi)^{-1} \int_{-\pi}^{\pi} h(k)e^{-ik\lambda}d\lambda$ is the $k$-th Fourier coefficient of $h(\lambda)$. Let $\varepsilon > 0$. Since the Cesàro mean converges uniformly in $\lambda \in \Pi$ by the continuity of $h(\cdot)$, we may choose large $M$ so that

$$\sup_{\lambda \in \Pi} |h(\lambda) - q_M(\lambda)| < \varepsilon. \quad (4.14)$$

Let $N_1 = \lfloor n/2 \rfloor$. For large $n > M$, we may expand

$$\frac{2\pi}{n} \sum_{m=-N}^{N_1} q_M(\lambda_m) I_{\text{nc}}(\lambda_m)$$

$$= \frac{2\pi}{n} \sum_{m=-N}^{N_1} \left( \sum_{t=-M}^{M} C_n h(t) \left(1 - \frac{|t|}{M}\right) e^{it\lambda_m} \right) \left( \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} \hat{r}_n(k) e^{-it\lambda_m} \right)$$

$$= \sum_{t=-M}^{M} C_n h(t) \left(1 - \frac{|t|}{M}\right) \left( \frac{1}{n} \sum_{k=-n+1}^{n-1} \hat{r}_n(k) \sum_{m=-N}^{N_1} e^{-i(k-t)\lambda_m} \right)$$

using sample covariances $\hat{r}_n(k) = \hat{r}_n(|k|) = n^{-1} \sum_{i=1}^{n-|k|} (X_i - \mu)(X_{i+|k|} - \mu)$, $|k| < n$. We have

$$\sum_{m=-N}^{N_1} e^{-i\lambda_m(k-t)} = \begin{cases} 0 & \text{if } k \neq t \\ n & \text{if } k = t. \end{cases}$$

Hence, for large $n > M$,

$$\frac{2\pi}{n} \sum_{m=-N}^{N_1} q_M(\lambda_m) I_{\text{nc}}(\lambda_m) = \sum_{t=-M}^{M} C_n h(t) \left(1 - \frac{|t|}{M}\right) \hat{r}_n(t).$$

By Lemma 3, $\hat{r}_n(t) \xrightarrow{p} r(t)$ as $n \to \infty$ for each fixed integer $t$, so that we have

$$\frac{2\pi}{n} \sum_{m=-N}^{N_1} q_M(\lambda_m) I_{\text{nc}}(\lambda_m) \xrightarrow{p} \sum_{t=-M}^{M} C_n h(t) \left(1 - \frac{|t|}{M}\right) r(t) = 2 \int_{0}^{\pi} q_M(\lambda) g(\lambda)d\lambda, \quad (4.15)$$

using $2 \int_{0}^{\pi} e^{-ik\lambda} g(\lambda)d\lambda = r(k) \equiv \text{Cov}(X_0, X_k)$.

By Lemma 2,

$$\Delta_{1n} \equiv \left| \frac{2\pi}{n} \sum_{m=1}^{N} h(\lambda_m) I_{\text{nc}}(\lambda_m) \right|$$

$$\leq |h(0)| I_{\text{nc}}(0) \frac{2\pi}{n} + |h(\pi)| I_{\text{nc}}(\pi) \frac{2\pi}{n} \xrightarrow{p} 0,$$
and likewise,

\[
\Delta_{2n} \equiv \left| \frac{2\pi}{n} \sum_{m=1}^{N} q_{M}(\lambda_{m})I_{n}(\lambda_{m}) - \frac{2\pi}{n} \sum_{m=-N}^{N_{1}} q_{M}(\lambda_{m})I_{nc}(\lambda_{m}) \right|
\]

\[
\leq |q_{M}(0)|I_{nc}(0)\frac{2\pi}{n} + |q_{M}(\pi)|I_{nc}(\pi)\frac{2\pi}{n} \to 0,
\]

because \(\sup_{\lambda \in \Pi} |q_{M}(\lambda)| \leq \sup_{\lambda \in \Pi} |h(\lambda)| < \infty\). Since \(2\pi n^{-1} \sum_{m=-N}^{N_{1}} I_{nc}(\lambda_{m}) = \hat{r}_{n}(0)\), we have from (4.14) that

\[
\Delta_{3n} \equiv \left| \frac{2\pi}{n} \sum_{m=-N}^{N_{1}} q_{M}(\lambda_{m})I_{nc}(\lambda_{m}) - \frac{2\pi}{n} \sum_{m=-N}^{N_{1}} h(\lambda_{m})I_{nc}(\lambda_{m}) \right|
\]

\[
\leq \frac{2\pi}{n} \sum_{m=-N}^{N_{1}} I_{nc}(\lambda_{m}) |q_{m}(\lambda_{m}) - h(\lambda_{m})| \leq \varepsilon \frac{2\pi}{n} \sum_{m=-N}^{N_{1}} I_{nc}(\lambda_{m}) = \varepsilon \hat{r}_{n}(0) \to \varepsilon r(0),
\]

by Lemma 3. From (4.15), as \(n \to \infty\),

\[
\Delta_{4n} \equiv \left| \frac{2\pi}{n} \sum_{m=-N}^{N_{1}} q_{M}(\lambda_{m})I_{nc}(\lambda_{m}) - 2 \int_{0}^{\pi} q_{M}(\lambda)g(\lambda)d\lambda \right| \to 0,
\]

and, from (4.14),

\[
\Delta_{5n} \equiv \left| 2 \int_{0}^{\pi} q_{M}(\lambda)g(\lambda)d\lambda - 2 \int_{0}^{\pi} h(\lambda)g(\lambda)d\lambda \right| \leq 2\varepsilon \int_{0}^{\pi} g(\lambda)d\lambda = \varepsilon r(0).
\]

We have now established that, for \(i = 1, 2, 3, 4, 5\),

\[
P(\Delta_{in} < 2\varepsilon r(0)) \to 1 \quad \text{as} \quad n \to \infty.
\]

Thus,

\[
P \left( \left| \frac{2\pi}{n} \sum_{m=1}^{N} h(\lambda_{m})I_{n}(\lambda_{m}) - \int_{0}^{\pi} h(\lambda)g(\lambda)d\lambda \right| \leq 10\varepsilon r(0) \right)
\]

\[
\geq P(\Delta_{1n} \leq 2\varepsilon r(0), \ldots, \Delta_{5n} \leq 2\varepsilon r(0)) \to 1.
\]

Since \(\varepsilon > 0\) is arbitrary, this establishes the result. \(\square\)

**Lemma 5.** Under assumptions \(L\) and \(A.5\), if \(h : [0, \pi] \to \mathbb{R}\) is a Riemann integrable function such that \(|h(\lambda)| \leq C|\lambda|^{\alpha}\) for \(0 < \alpha < 1\) and \(\alpha - 2d > -1/2\), then

\[
\frac{n}{\log n} \text{Var} \left( \frac{2\pi}{n} \sum_{m=1}^{N} h(\lambda_{m})I_{n}(\lambda_{m}) \right) = O(1) \quad \text{as} \quad n \to \infty.
\]
Proof: Recalling $I_n(\lambda_m) = I_{nc}(\lambda_m)$ for $m = 1, \ldots, N$, we bound

$$V_n \equiv \text{Var} \left( \frac{2\pi}{n} \sum_{m=1}^{N} h(\lambda_m) I_{nc}(\lambda_m) \right).$$

as $V_n \leq V_{1n} + V_{2n}$ with $V_{1n}$ and $V_{2n}$ defined as

$$V_{1n} = \frac{C}{n^4} \sum_{j=1}^{N} |\lambda_j|^{2\alpha} \left| \text{Cum} \left( |d_{nc}(\lambda_j)|^2, |d_{nc}(\lambda_j)|^2 \right) \right|,$$

$$V_{2n} = \frac{C}{n^4} \sum_{1 \leq j < k \leq N} |\lambda_j \lambda_k|^\alpha \left| \text{Cum} \left( |d_{nc}(\lambda_j)|^2, |d_{nc}(\lambda_k)|^2 \right) \right|,$$

respectively. By Theorem 2.3.2 (Brillinger, 1981) and since $E_{nc}(\lambda) = 0$, $\lambda \in \Pi$, for $j, k = 1, 2, \ldots, N$, we may expand

$$\text{Cum} \left( |d_{nc}(\lambda_j)|^2, |d_{nc}(\lambda_k)|^2 \right) = \text{Cum} (d_{nc}(\lambda_j), d_{nc}(\lambda_k)) \text{Cum} (d_{nc}(\lambda_j), d_{nc}(\lambda_k)) + \text{Cum} (d_{nc}(\lambda_j), d_{nc}(\lambda_k)) \text{Cum} (d_{nc}(-\lambda_j), d_{nc}(\lambda_k)) + \text{Cum} (d_{nc}(\lambda_j), d_{nc}(\lambda_k), d_{nc}(\lambda_k), d_{nc}(-\lambda_k)).$$

Under A.5, we have $g(\lambda) \equiv g(\lambda; \sigma_0^2, \theta_0) \leq C|\lambda|^{-2d}$. By this and Lemma 1, we may write

$$V_{1n} \leq \frac{C}{n^4} \sum_{m=1}^{N} |\lambda_m|^{2\alpha} \left( |\lambda_m|^{-2d} (|\lambda_m|^{-1} + n) \right)^2 \sum_{m=1}^{N} |\lambda_m|^{2\alpha} |\lambda_m|^{-4d} \left\{ |\lambda_m|^{2d-3/2} + n \log^3 n \right\} = O \left( \frac{1}{n} \right),$$

since $\frac{2\pi}{n} \sum_{m=1}^{N} |\lambda_m|^{2\alpha-4d} \rightarrow \int_0^\pi \lambda^{2\alpha-4d} d\lambda < \infty$ as $n \rightarrow \infty$ by $2\alpha - 4d > -1$. We next consider $V_{2n}$ and bound $V_{2n} \leq U_{1n} + U_{2n} + U_{3n}$ with quantities defined as

$$U_{1n} = \frac{C}{n^4} \sum_{1 \leq j < k \leq N} (\lambda_j \lambda_k)^\alpha \left| \text{Cum} (d_{nc}(\lambda_j), d_{nc}(\lambda_k)) \text{Cum} (d_{nc}(\lambda_j), d_{nc}(\lambda_k)) \right|,$$

$$U_{2n} = \frac{C}{n^4} \sum_{1 \leq j < k \leq N} (\lambda_j \lambda_k)^\alpha \left| \text{Cum} (d_{nc}(\lambda_j), d_{nc}(\lambda_k)) \text{Cum} (d_{nc}(\lambda_j), d_{nc}(\lambda_k)) \right|,$$

$$U_{3n} = \frac{C}{n^4} \sum_{1 \leq j < k \leq N} (\lambda_j \lambda_k)^\alpha \left| \text{Cum} (d_{nc}(\lambda_j), d_{nc}(\lambda_j), d_{nc}(\lambda_k), d_{nc}(\lambda_k)) \right|.$$

By Lemma 1,

$$\max \{U_{1n}, U_{2n}\} \leq \frac{C}{n^4} \sum_{1 \leq j < k \leq N} (\lambda_j \lambda_k)^\alpha \left\{ |\lambda_j|^{-2d} (|\lambda_k|-1 + L_{n1}(\lambda_k - \lambda_j)) \right\}^2 \leq T_{1n} + T_{2n} + T_{3n},$$
where

\[ T_{1n} = \frac{C}{n^4} \sum_{1 \leq j < k \leq N} |\lambda_j|^{\alpha-4d} |\lambda_k|^{\alpha-2}, \]

\[ T_{2n} = \frac{C}{n^4} \sum_{1 \leq j < k \leq N} |\lambda_j|^{\alpha-4d} |\lambda_k|^{\alpha-1} L_{n1}(\lambda_k - \lambda_j), \]

\[ T_{3n} = \frac{C}{n^4} \sum_{1 \leq j < k \leq N} |\lambda_j|^{\alpha-4d} |\lambda_k|^{\alpha} L_{n1}^2(\lambda_k - \lambda_j). \]

For \( T_{1n} \) and \( T_{2n} \), we have that as \( n \to \infty \),

\[ T_{1n} \leq \frac{C}{n^4} \sum_{1 \leq j < k \leq N} |\lambda_j|^{\alpha-4d} \frac{j^2}{\alpha n^2} \leq \frac{C}{n^4} \sum_{j=1}^{N} |\lambda_j|^{\alpha-4d} \frac{1}{k^{2-\alpha}} = O\left( \frac{1}{n} \right), \]

and

\[ T_{2n} \leq \frac{C}{n^4} \sum_{1 \leq j < k \leq N} \lambda_j^{2\alpha-4d} n^2 \log n \frac{1}{k-j} \leq \frac{C \log n}{n^2} \sum_{j=1}^{N} \lambda_j^{2\alpha-4d} \sum_{k=1}^{n} \frac{1}{k} = O\left( \frac{\log n}{n} \right), \]

using \( 2\alpha - 4d > -1 \). Since \( 0 < \alpha < 1 \) and \( \int_{1}^{\infty} y^{-2+\alpha} dy < \infty \), we have

\[ \sum_{k=j+1}^{N} \frac{k^\alpha}{(k-j)^2} \leq (j+1)^\alpha + \int_{j+1}^{n} \frac{x^\alpha}{(x-j)^2} dx \leq (2j)^\alpha + \int_{1}^{\infty} \frac{(y+j)^\alpha}{y^2} dy \leq C j^\alpha, \]

and, therefore for \( T_{3n} \), we have that as \( n \to \infty \)

\[ T_{3n} \leq \frac{C}{n^4} \sum_{1 \leq j < k \leq N} \lambda_j^{\alpha-4d} \frac{k^\alpha n^2 \log^2 n}{n^\alpha (k-j)^2} \leq \frac{C}{n^4} \sum_{j=1}^{N} \lambda_j^{\alpha-4d} n^2 \log^2 n = O\left( \frac{\log^2 n}{n} \right). \]

Likewise, by Lemma 1 and since \( \alpha - 2d > -1/2 \) and \( 0 \leq 2d < 1 \),

\[ U_{3n} \leq \frac{C}{n^4} \sum_{1 \leq j < k \leq N} (\lambda_j \lambda_k)^{\alpha-2d} \left\{ \lambda_k^{(2d-3)/2} + n \log^3 n \right\} \]

\[ \leq \frac{C \log^3 n}{n^3} \sum_{1 \leq j < k \leq N} (\lambda_j \lambda_k)^{\alpha-2d} + \frac{C}{n^2} \sum_{j=1}^{N} \lambda_j^{\alpha-2d} \cdot n^{(2d-1)/2} \sum_{k=1}^{\infty} k^{-(3+2d)/2} \]

\[ = O\left( \frac{\log^3 n}{n} \right). \]

This completes the proof. □

**Lemma 6.** Under assumptions \( L, A.1-A.5 \) and \( \int_{0}^{\pi} \frac{\partial f^{-1}(\lambda, \theta_0)}{\partial \theta} g(\lambda) d\lambda = 0_p \),

\[ \sqrt{n} E (\ell_n(\theta_0)) = o(1) \quad \text{as} \quad n \to \infty. \]
\textbf{Proof:} By the proof of Lemma 6 of Nordman and Lahiri (2006), it follows that

\[ L_{31} = \left\| E \int_0^\pi \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta} I_{nc}(\lambda)d\lambda - \int_0^\pi \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta} g(\lambda)d\lambda \right\| = o \left( \frac{1}{n} \right). \]

The proof of Lemma 10 of Nordman and Lahiri (2006) shows that

\[ L_{32} \equiv E \left\| C_{1n} - \int_0^\pi \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta} I_{nc}(\lambda)d\lambda \right\| = o \left( \frac{1}{n} \right). \]

where

\[ C_{1n} \equiv \int_0^\pi C_{1h}(\lambda)I_{nc}(\lambda)d\lambda \quad \text{for} \quad C_{1h}(\lambda) \equiv \int_{\Pi} \frac{\partial f^{-1}(y; \theta_0)}{\partial \theta} K_n(\lambda - y)dy, \]

and also

\[ L_{33} \equiv 2\sqrt{n}E \left\| \frac{2\pi}{n} \sum_{m=1}^N \frac{\partial f^{-1}(\lambda_m; \theta_0)}{\partial \theta} I_{nc}(\lambda_m) - C_{1n} \right\| \leq 4\pi\sqrt{n}(E t_{3n} + E t_{4n}) \]

where

\[ t_{3n} \equiv \frac{1}{n} \sum_{m=1}^N \left\| C_{1h}(\lambda_m) - \frac{\partial f^{-1}(\lambda_m; \theta_0)}{\partial \theta} \right\| I_{nc}(\lambda_m), \]

\[ t_{4n} \equiv \frac{1}{n} \left( \| C_{1h}(0) \| I_{nc}(0) + \| C_{1h}(\pi) \| I_{nc}(\pi) \right). \]

Since \( \int_0^\pi \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta} g(\lambda)d\lambda = 0_p \), to complete the proof, it is now enough to show \( E t_{3n} = o(n^{-1/2}) \) and \( E t_{4n} = o(n^{-1/2}) \).

We consider \( \sqrt{n}E t_{3n} \) first. Because \( \| \partial^2/\partial \theta_j \partial \lambda f^{-1}(\lambda; \theta_0) \| \leq C(\theta_0)|\lambda|^{2d-1} \) holds under \textbf{A.4}, it follows that, for a \( C > 0 \) independent of \( 1 \leq m \leq N \) \((n > 3)\),

\[ \left\| C_{1h}(\lambda_m) - \frac{\partial f^{-1}(\lambda_m; \theta_0)}{\partial \theta} \right\| \leq C|\lambda_m|^{2d} \left( \frac{\log n}{m} + \frac{\mathbb{I}(m > n/4)}{n - 2m} \right), \quad 1 \leq m \leq N, \]

where \( \mathbb{I}(\cdot) \) denotes the indicator function (cf. Lemma 10, Nordman and Lahiri, 2006). By Lemma 1, we then have \( E I_n(\lambda_m) \leq C|\lambda_m|^{-2d} \) for \( m = 1, 2, \ldots, N \) so that

\[ E(t_{3n}) \leq \frac{1}{n} \sum_{m=1}^N E I_{nc}(\lambda_m) \left\| C_{1h}(\lambda_m) - \frac{\partial f^{-1}(\lambda_m; \theta_0)}{\partial \theta} \right\| \]

\[ \leq C \sum_{m=1}^N |\lambda_m|^{-2d} |\lambda_m|^{2d} \left( \frac{\log n}{m} + \frac{\mathbb{I}(m > n/4)}{n - 2m} \right) \]

\[ \leq \frac{C}{n} \sum_{m=1}^N \frac{\log n}{m} = O \left( \frac{\log^2 n}{n} \right) = o(n^{-1/2}). \]
We next consider $E(t_{4n}) = n^{-1} |C_{1n}\lambda(0)| E_{\text{nc}}(0) + n^{-1} |C_{1n}\lambda(\pi)| E_{\text{nc}}(\pi)$. Since $\partial f^{-1}(\lambda, \theta_0)/\partial \theta$ is bounded and $\int_{-\pi}^\pi K_n(\lambda)d\lambda = 1$, it holds that $|C_{1n}\lambda(\pi)|$ is bounded, so that

$$\frac{1}{n} |C_{1n}\lambda(\pi)| E_{\text{nc}}(\lambda) = O(n^{-1}) = o(n^{-1/2})$$

holds by Lemma 2(i). By (4.12), (4.13), and A.4,

$$|C_{1n}\lambda(0)| = \frac{C}{n} \left| \int_0^\pi H_n(y)H_n(-y) \frac{\partial f^{-1}(y; \theta_0)}{\partial \theta} dy \right| \leq \frac{C}{n} \int_0^\pi n^2 |y|^{-2} |y| 2d dy = C n^{-2d+4^{-1}}$$

for $\tau = (1 - 2d + 4^{-1})/2 > 0$. Hence, by Lemma 2, $n^{-1} |C_{1n}\lambda(0)| E_{\text{nc}}(0) = O(n^{-3/4}) = o(n^{-1/2})$. □

**Lemma 7.** Under assumptions L, A.1-A.5 and $\int_0^\pi \theta^2 \frac{\partial f^{-1}(\lambda, \theta_0)}{\partial \sigma^2} g(\lambda) d\lambda = 0$, as $n \to \infty$,

$$\sqrt{n} \ell_n(\theta_0) = \sqrt{n} \left( \frac{2\pi}{n} \sum_{m=1}^N \frac{\partial f^{-1}(\lambda_m; \theta_0)}{\partial \theta} I_n(\lambda_m) - \int_0^\pi \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta} g(\lambda) d\lambda \right) \quad \overset{d}{\to} N(0, \sigma_n^2 D_0)$$

**Proof:** Under the assumptions, Lemma 6 of Nordman and Lahiri (2006) gives

$$\sqrt{n} \ell_n(\theta_0) = \sqrt{n} \left( \frac{2\pi}{n} \sum_{m=1}^N \frac{\partial f^{-1}(\lambda_m; \theta_0)}{\partial \theta} I_n(\lambda_m) - \int_0^\pi \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta} g(\lambda) d\lambda \right) \quad \overset{d}{\to} N(0, \sigma_n^2 D_0)$$

where

$$D_0 \equiv \int_0^\pi \frac{2\pi}{\sigma_n^2} \left( \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta} \right) \left( \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta} \right)^T g(\lambda) d\lambda = \int_0^\pi \frac{\partial^2 f^{-1}(\lambda; \theta_0)}{\partial \theta \partial \theta^T} g(\lambda) d\lambda$$

the last equality follows from A.1. □

### 4.6.2 Preliminary results for proving Theorem 2: Lemmas 8-10

**Lemma 8.** Let $r(0) = \text{Var}(X_i)$, $g(\lambda) \equiv g(\lambda; \sigma_n^2, \theta_0) \equiv \sigma_n^2 f(\lambda_m; \theta_0)/2\pi$ and $\hat{g}_n(\lambda) \equiv g(\lambda; \hat{\sigma}_n^2, \hat{\theta}_n) \equiv \hat{\sigma}_n^2 f(\hat{\lambda}_m; \hat{\theta}_n)/2\pi$ for $\lambda \in \Pi$. Under assumptions of Theorem 1 and assuming the results of Theorem 1 hold, as $n \to \infty$,

(i) $Z_n \equiv \max_{1 \leq m \leq N} I_n(\lambda_m)/g(\lambda_m) = o_p(n^{1/2})$,

(ii) $\hat{Z}_n \equiv \max_{1 \leq m \leq N} I_n(\lambda_m)/\hat{g}_n(\lambda_m) = o_p(n^{1/2})$,
(iii) \(2\pi n^{-1} \sum_{m=1}^{N} I_n(\lambda_m)/\tilde{g}_n(\lambda_m) \xrightarrow{p} \pi\),

(iv) \(2\pi n^{-1} \sum_{m=1}^{N} I_n^2(\lambda_m)/\tilde{g}_n^2(\lambda_m) \xrightarrow{p} 2\pi\),

(v) \(2\pi n^{-1} \sum_{m=1}^{N} \tilde{g}_n(\lambda_m) \xrightarrow{p} \int_0^\pi g(\lambda)d\lambda = r(0)/2\).

**Proof:** Pick \(\varepsilon > 0\). By Jensen’s inequality and A.4 and A.5, we have

\[
P\left(\frac{1}{n^{1/2}} Z_n > \varepsilon\right) \leq \frac{1}{\varepsilon n^{1/2}} E \left( \left[ \max_{1 \leq m \leq N} \frac{I_n^4(\lambda_m)}{g^4(\lambda_m)} \right]^{1/4} \right)
\]

\[
\leq \frac{1}{\varepsilon n^{1/2}} \left[ E \left( \sum_{m=1}^{N} \frac{I_n^4(\lambda_m)}{g^4(\lambda_m)} \right) \right]^{1/4}
\]

\[
\leq \frac{1}{\varepsilon n^{1/2}} \left[ \sum_{m=1}^{N} C|\lambda_m|^{-4(2d)+4(2d)} \right]^{1/4}
\]

\[
= \frac{C}{\varepsilon n^{1/2}} n^{1/4} = o(1),
\]

since \(g^{-1}(\lambda) \leq C|\lambda|^{2d}\) by A.5 and \(E I_n^4(\lambda_m) \leq C|\lambda_m|^{-4(2d)}\) by Lemma 1(ii). This establishes part (i).

Let \(Z_{1n} = \max_{1 \leq m \leq N} I_n(\lambda_m)/\tilde{g}(\lambda_m)\) where \(\tilde{g}_n(\lambda_m) = \sigma^2_0 f(\lambda_m; \hat{\theta}_n)/(2\pi)\). In order to show \(\tilde{Z}_n = o_p(n^{1/2})\), it suffices to show that \(Z_{1n} = o_p(n^{1/2})\) because

\[
\frac{\tilde{Z}_n}{n^{1/2}} = \frac{\sigma^2_0 Z_{1n}}{\sigma^2_0 n^{1/2}} \xrightarrow{p} 0
\]

would then follow by Theorem 1. Fix \(\delta > 0\) such that \(B \equiv \{\theta \in \Theta : ||\theta_0 - \theta|| \leq \delta\} \subset \Theta\) (cf. A.5). By A.3, it holds that

\[
f^{-1}(\lambda; \theta) = f^{-1}(\lambda; \theta_0) + \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta} (\theta - \theta_0) + R(\lambda; \theta) = (4.16)
\]

where \(||R(\lambda, \theta)|| \leq C||\theta - \theta_0||^2\) for \(C\) which does not depend on \(\lambda \in \Pi, \theta \in B\). By Theorem 1, \(\hat{\theta}_n - \theta_0 = O_p(n^{-1/2})\) and when \(\hat{\theta}_n \in B\), we write

\[
Z_{1n} = \max_{1 \leq m \leq N} I_n(\lambda_m)/\tilde{g}(\lambda_m) = \max_{1 \leq m \leq N} I_n(\lambda_m) \frac{2\pi}{\sigma^2_0} f^{-1}(\lambda_m; \hat{\theta}_n)
\]

\[
= \max_{1 \leq m \leq N} I_n(\lambda_m) \frac{2\pi}{\sigma^2_0} \left\{ f^{-1}(\lambda_m; \theta_0) + \frac{\partial f^{-1}(\lambda_m; \theta_0)}{\partial \theta} (\hat{\theta}_n - \theta_0) + R(\lambda_m; \hat{\theta}_n) \right\}
\]

\[
\leq Z_n + C \max_{1 \leq m \leq N} I_n(\lambda_m)|\lambda_m|^{2d}||\hat{\theta}_n - \theta_0|| + C \max_{1 \leq m \leq N} I_n(\lambda_m)||\hat{\theta}_n - \theta_0||^2
\]

\[
\equiv \Delta_{1n} + \Delta_{2n} + \Delta_{3n}
\]
by A.3-A.4. We have $\Delta_1 = Z_n = o_p(n^{1/2})$ by Lemma 8(i), implying

$$\Delta_{2n} \leq C \Delta_{1n} |\hat{\theta}_n - \theta_0| = o_p(n^{1/2})O_p(n^{-1/2}) = o_p(1).$$

In addition, by Lemma (1),

$$E \max_{1 \leq m \leq N} I_n(\lambda_m) \leq \sum_{m=1}^{N} E I_n(\lambda_m) \leq C \sum_{m=1}^{N} \lambda_m^{-2d} = O(n),$$

so that

$$\Delta_3 \leq C \max_{1 \leq m \leq N} I_n(\lambda_m) \|\theta - \theta_0\|^2 = O_p(n) \cdot O_p\left(\frac{1}{n}\right) = o_p\left(n^{1/2}\right).$$

Thus,

$$Z_{n1} = o_p\left(n^{1/2}\right),$$

establishing part (ii).

By (4.16), for $\hat{\theta}_n \in B$, we may write

$$\frac{2\pi}{n} \sum_{m=1}^{N} \frac{I_n(\lambda_m)}{g_n(\lambda_m)} = \frac{2\pi}{n} \sum_{m=1}^{N} \frac{I_n(\lambda_m)}{g(\lambda_m)} + R_{1n}$$

where

$$R_{1n} \leq \frac{C}{n} \sum_{m=1}^{N} I_n(\lambda_m) \left[\|\hat{\theta}_n - \theta_0\| |\lambda_m|^{2d} + \|\hat{\theta}_n - \theta_0\|^2\right]$$

using A.4. By Lemma 4,

$$\frac{2\pi}{n} \sum_{m=1}^{N} I_n(\lambda_m) \xrightarrow{p} \int_{0}^{\pi} g(\lambda) d\lambda < \infty, \quad \frac{2\pi}{n} \sum_{m=1}^{N} \frac{I_n(\lambda_m)}{g(\lambda_m)} \xrightarrow{p} \pi,$$

and

$$\frac{2\pi}{n} \sum_{m=1}^{N} I_n(\lambda_m) |\lambda_m|^{2d} \xrightarrow{p} \int_{0}^{\pi} g(\lambda) |\lambda|^{2d} d\lambda < \infty,$$

and by Theorem 1, $\|\hat{\theta}_n - \theta_0\| = O_p(n^{-1/2})$, so that $R_{1n} = o_p(1)$ and

$$\frac{2\pi}{n} \sum_{m=1}^{N} \frac{I_n(\lambda_m)}{g_n(\lambda_m)} \xrightarrow{p} \pi,$$

By Theorem 1, $\hat{\sigma}_n^2 \xrightarrow{p} \sigma_0^2 > 0$ so that

$$\frac{2\pi}{n} \sum_{m=1}^{N} \frac{I_n(\lambda_m)}{g_n(\lambda_m)} = \frac{\sigma_0^2}{\hat{\sigma}_n^2} \frac{2\pi}{n} \sum_{m=1}^{N} \frac{I_n(\lambda_m)}{g_n(\lambda_m)} \xrightarrow{p} \pi,$$

establishing part(iii).
To show part (iv), it suffices to show
\[
\frac{2\pi}{N} \sum_{m=1}^{N} \frac{I_n^2(\lambda_m)}{g_n^2(\lambda_m)} \overset{p}{\to} 2\pi
\]
(4.17)
so that \( \sigma_n^2/\sigma_n^2 \overset{p}{\to} 1 \) as \( n \to \infty \),
\[
\frac{2\pi}{n} \sum_{m=1}^{N} \frac{I_n^2(\lambda_m)}{g_n^2(\lambda_m)} = \left( \frac{\sigma_n^2}{\sigma_n^2} \right) \frac{2\pi}{n} \sum_{m=1}^{N} \frac{I_n^2(\lambda_m)}{g_n^2(\lambda_m)} \overset{p}{\to} 2\pi.
\]

To show (4.17), write for \( \theta \in B \),
\[
f^{-2}(\lambda; \theta) = f^{-2}(\lambda; \theta_0) + (-2)f^{-3}(\lambda; \theta) \left( \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta} \right)^T (\theta - \theta_0) + R(\lambda, \theta)
\]
where \( ||R(\lambda, \theta)|| \leq C||\theta - \theta_0||^2 \) for a constant \( C \) not depending on \( \lambda \in \Pi \), \( \theta \in B \) by A.3. From this, when \( \hat{\theta}_n \in B \) by Theorem 1, we then have
\[
\frac{2\pi}{n} \sum_{m=1}^{N} \frac{I_n^2(\lambda_m)}{g_n^2(\lambda_m)} = \frac{2\pi}{n} \sum_{m=1}^{N} \frac{I_n^2(\lambda_m)}{g_n^2(\lambda_m)} + R_{2n} + R_{3n},
\]
for remainder terms \( R_{2n} \) and \( R_{3n} \) defined below. Under assumptions A.2-A.4, Lemma 7 of Nordman and Lahiri (2006) gives
\[
\frac{2\pi}{n} \sum_{m=1}^{N} \frac{I_n^2(\lambda_m)}{g_n^2(\lambda_m)} \overset{p}{\to} 2\pi.
\]
Hence, (4.17) and part (iv) will follow by showing \( R_{2n}, R_{3n} \overset{p}{\to} 0 \). Note that
\[
E \left( \sum_{m=1}^{N} \frac{I_n^2(\lambda_m)}{n} \right) \leq \frac{C}{n} \sum_{m=1}^{N} |\lambda_m|^{-2(2d)} \leq C n^{2d-1} \frac{1}{n} \sum_{m=1}^{N} |\lambda_m|^{-2d} = O \left( n^{2d} \right),
\]
by Lemma (1), implying
\[
||R_{3n}|| \leq C \left( \frac{1}{n} \right) \left( \frac{1}{n} \right) = O_p \left( n^{2d} \right) = o_p(1)
\]
by Theorem ?? since \( 2d < 1 \). From A.3-A.4, we may bound
\[
||R_{2n}|| \equiv \left\| \frac{2\pi}{n} \sum_{m=1}^{N} \frac{I_n^2(\lambda_m)(-2)f^{-3}(\lambda_m; \theta_0) \left( \frac{\partial f^{-1}(\lambda_m; \theta_0)}{\partial \theta} \right)^T (\hat{\theta}_n - \theta_0) \right\|
\]
\[
\leq \frac{C}{n} \left( \frac{1}{n} \right) \frac{C}{n} \sum_{m=1}^{N} I_n^2(\lambda_m)|\lambda_m|^{4(2d)}.
\]
By Lemma 1 (ii), \( EI_n^2(\lambda_m) \leq C|\lambda_m|^{-4d} \) implying \( n^{-1} \sum_{m=1}^{N} I_n^2(\lambda_m)|\lambda_m|^{4(2d)} = O_p(1) \) and
\[
||R_{3n}|| = O_p(n^{-1/2})o_p(1) = o_p(1)
\]
by Theorem 1.
To show part (iv), suppose \( \hat{\theta}_n \in \mathcal{B} \) by Theorem 1 for \( \mathcal{B} \) defined in A.5. Fix \( 0 < \rho < \pi \). Because \( f(\lambda; \theta) \) is continuous on \([\rho, \pi] \times \mathcal{B}\), which is a compact set, we have \( R_{4n}(\rho) \equiv \sup_{\lambda \in [\rho, \pi]} |f(\lambda; \theta_0) - f(\lambda; \hat{\theta}_n)| \overset{p}{\to} 0 \). Hence, under A.5, when \( \hat{\theta}_n \in \mathcal{B} \), we may bound

\[
\frac{2\pi}{n} \sum_{m=1}^{N} |f(\lambda_m; \hat{\theta}_n) - f(\lambda_m; \theta_0)| \leq \frac{C}{n} \sum_{\lambda_m \leq \rho} |\lambda_m|^{-2d - \eta} + 4\pi R_{4n}(\rho)
\]

since \(-2d - \eta > -1\) where the constant \( C > 0 \) does not depend on \( \rho \). Because \( \rho^{1-2d-\eta} \) can be made arbitrarily small for small \( \rho \), and \( R_{4n}(\rho) \overset{p}{\to} 0 \) given \( \rho \), we have

\[
\frac{2\pi}{n} \sum_{m=1}^{N} |f(\lambda_m; \hat{\theta}_n) - f(\lambda_m; \theta_0)| \overset{p}{\to} 0.
\]

(4.18)

Because \( \hat{\sigma}_n^2 - \sigma_0^2 \overset{p}{\to} 1 \) by Theorem 1, the above implies

\[
\frac{2\pi}{n} \sum_{m=1}^{N} |\hat{g}_n(\lambda_m) - g(\lambda_m)| \overset{p}{\to} 0.
\]

(4.19)

for \( g(\lambda) = \sigma_0^2 f(\lambda; \theta_0)/(2\pi) \) and \( \hat{g}_n(\lambda) = \hat{\sigma}_n^2 f(\lambda; \hat{\theta}_n)/(2\pi) \). Finally, \( n^{-1}(2\pi) \sum_{m=1}^{N} g(\lambda_m) \to \int_{0}^{\pi} g(\lambda)d\lambda \) by the Dominated Convergence Theorem because the step function on \([0, \pi]\)

\[
g_{n, step}(\lambda) = \sum_{m=1}^{N} g(\lambda_m) I(\lambda_{m-1} < \lambda \leq \lambda_m) \to g(\lambda)
\]

almost everywhere under the Lebesgue measure on \([0, \pi]\) under A.2 (where above \( I(\cdot) \) is the indicator function and \( \lambda_0 = 0 \)) with \( \int_{0}^{\pi} g_{n, step}(\lambda)d\lambda = n^{-1}(2\pi) \sum_{m=1}^{N} g(\lambda_m) \) and because \( g(\lambda) \leq C|\lambda|^{-2d} \) by A.5 where \( \int_{0}^{\pi} |\lambda|^{-2d}d\lambda < \infty \). This with (4.19) yields

\[
\frac{2\pi}{n} \sum_{m=1}^{N} \hat{g}_n(\lambda_m) \overset{p}{\to} \int_{0}^{\pi} g(\lambda)d\lambda = r(0)/2,
\]

establishing part (v). \( \square \)

Lemma 9. Under assumptions of Theorem 1 and assuming the results of Theorem 1 hold, as \( n \to \infty \),

(i) \( \mathbb{E}_n \ell_n^*(\theta_0) \overset{p}{\to} \int_{0}^{\pi} \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta} g(\lambda)d\lambda(\equiv 0) \),

(ii) \( \mathbb{E}_n D_n^*(\theta_0) \overset{p}{\to} \int_{0}^{\pi} \frac{\partial^2 f^{-1}(\lambda; \theta_0)}{\partial \theta^2} g(\lambda)d\lambda(\equiv D_0) \),
\( (iii) \) \( \text{Var}_* (\varepsilon_1^*) \xrightarrow{p} 1, \)

\( (iv) \) \( n \text{Var}_* (a^T D_n^* (\theta_0) b) \xrightarrow{p} 2 \pi \int_0^\pi \left( a^T \frac{\partial^2 f^{-1}(\lambda; \theta_0)}{\partial \theta \partial \theta^T} b \right)^2 g^2(\lambda) d\lambda, \) for any \( a, b \in \mathbb{R}^p, \) and

\( (v) \) \( n \text{Var}_* (\ell_n^*(\theta_0)) \xrightarrow{p} 2 \pi \int_0^\pi \left[ \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta} \right]^T \left[ \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta} \right] g^2(\lambda) d\lambda = \sigma_0^2 D_0. \)

where

\[
D_n^*(\theta_0) = \frac{2\pi}{n} \sum_{m=1}^{N} \frac{\partial^2 f^{-1}(\lambda_m; \theta_0)}{\partial \theta \partial \theta^T} I_n^*(\lambda_m) \quad \text{and} \quad \ell_n^*(\theta_0) = \frac{2\pi}{n} \sum_{m=1}^{N} \frac{\partial f^{-1}(\lambda_m; \theta_0)}{\partial \theta} I_n^*(\lambda_m).
\]

**Proof:** To show part (i), we take the bootstrap expectation

\[
E_* \ell_n^*(\theta_0) = \frac{2\pi}{n} \sum_{m=1}^{N} \frac{\partial f^{-1}(\lambda_m; \theta_0)}{\partial \theta} g_n(\lambda_m).
\]

Since \( \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta} \) is continuous, and hence, bounded on \( \Pi, \) \( (4.19) \) implies that

\[
\left\| \frac{2\pi}{n} \sum_{m=1}^{N} \frac{\partial f^{-1}(\lambda_m; \theta_0)}{\partial \theta} \left[ g_n(\lambda_m) - g(\lambda_m) \right] \right\| \xrightarrow{p} 0, \quad (4.20)
\]

while the Dominated Convergence Theorem gives \( n^{-1}(2\pi) \sum_{m=1}^{N} g(\lambda_m) f^{-1}(\lambda_m; \theta_0) / \partial \theta \rightarrow f_0^\pi g(\lambda) f^{-1}(\lambda; \theta_0) / \partial \theta d\lambda = 0_p \) under \( A.2 \) and \( A.5; \) this step is analogous to the proof of Lemma 8(v). By this and \( (4.20), \) Lemma 9(i) follows. Additionally, because \( \frac{\partial^2 f^{-1}(\lambda; \theta_0)}{\partial \theta \partial \theta^T} \) is also bounded and continuous on \( \Pi, \) the same method of proof of Lemma 9(i) yields Lemma 9(ii),

\[
E_* D_n^*(\theta_0) \xrightarrow{p} \int_0^\pi \frac{\partial^2 f^{-1}(\lambda; \theta_0)}{\partial \theta \partial \theta^T} g(\lambda) d\lambda.
\]

Next using \( E_* \varepsilon_1^* = 1, \) we expand \( \text{Var}_* (\varepsilon_1^*) = E_* (\varepsilon_1^*)^2 - 1 = V_{2n} \cdot V_{1n} - 1 \) where

\[
V_{1n} \equiv \left( \frac{1}{N} \sum_{m=1}^{N} I_n(\lambda_m) \right)^{-2} \quad \text{and} \quad V_{2n} \equiv \frac{1}{N} \sum_{m=1}^{N} \frac{I_n^2(\lambda_m)}{\tilde{g}_n^2(\lambda_m)}.
\]

By Lemma 8(iii)-(iv) and \( n/2N \rightarrow 1, \) we have \( \pi \sqrt{V_{2n}} \xrightarrow{p} 2 \pi \) and \( \pi V_{1n}^{-1/2} \xrightarrow{p} \pi, \) which implies \( V_{2n} V_{1n} \xrightarrow{p} 2 \) by the continuous mapping theorem. Hence, \( \text{Var}_* (\varepsilon_1^*) \xrightarrow{p} 1 \) follows in part(iii).

To show part (iv), the bootstrap variance can be written as

\[
n \text{Var}_* (a^T D_n^*(\theta_0)) = \frac{2\pi}{n} \sum_{m=1}^{N} \left( a^T \frac{\partial^2 f^{-1}(\lambda_m; \theta_0)}{\partial \theta \partial \theta^T} b \right)^2 f^2(\lambda_m; \hat{\theta}_n).
\]

\[
V_{3n} \equiv \frac{2\pi}{n} \sum_{m=1}^{N} \left( a^T \frac{\partial^2 f^{-1}(\lambda_m; \theta_0)}{\partial \theta \partial \theta^T} b \right)^2 f^2(\lambda_m; \hat{\theta}_n).
\]
Suppose \( \hat{\theta}_n \in \mathcal{B} \) by Theorem 1 for \( \mathcal{B} \) defined in \( A.5 \). Since \( \hat{\theta}_n \xrightarrow{p} \theta_0 \) by Theorem 1, if \( \{n_j\} \) is a subsequence of \( \{n\} \), extract a further subsequence \( \{n_k\} \subset \{n_j\} \) along which \( \hat{\theta}_{n_k} \rightarrow \theta_0 \) almost surely. Then by \( A.2-A.3 \), the step function

\[
h_{n_k, \text{step}}(\lambda) = \sum_{m=1}^{N_k} \left( a^T \frac{\partial^2 f^{-1}(\lambda_m; \theta_0)}{\partial \theta \partial \theta^T} b \right)^2 f^2(\lambda; \hat{\theta}_{n_k}) \mathbb{I}(\lambda_{m-1} < \lambda \leq \lambda_m)
\]

almost everywhere under the Lebesgue measure on \( [0, \pi] \) with \( \int_0^\pi h_{n_k, \text{step}}(\lambda) d\lambda = V_{3n_k} \) and also

\[
\left\| \left( a^T \frac{\partial^2 f^{-1}(\lambda; \theta_0)}{\partial \theta \partial \theta^T} b \right)^2 f^2(\lambda; \hat{\theta}_{n_k}) \right\| \leq C|\lambda|^{-2\eta}, \lambda \in (0, \pi)
\]

holds by \( A.5 \) when \( \hat{\theta}_n \in \mathcal{B} \) where \( \int_0^\pi |\lambda|^{-2\eta} d\lambda < \infty \) by \(-2\eta > -1\). Hence, by the Dominated Convergence Theorem,

\[
V_{3n_k} \xrightarrow{\text{a.s.}} \int_0^\pi \left( a^T \frac{\partial^2 f^{-1}(\lambda; \theta_0)}{\partial \theta \partial \theta^T} b \right)^2 f^2(\lambda; \theta_0)
\]

almost surely, implying

\[
V_{3n} \xrightarrow{p} \int_0^\pi \left( a^T \frac{\partial^2 f^{-1}(\lambda; \theta_0)}{\partial \theta \partial \theta^T} b \right)^2 f^2(\lambda; \theta_0).
\]

By this, Lemma 9(iii) and \( \hat{\sigma}_n^2 \xrightarrow{p} \sigma_0^2 \) under Theorem 1, we have

\[
n \text{Var}_* (a^T D_n^*(\theta_0) b) \xrightarrow{p} 2\pi \int_0^\pi \left( a^T \frac{\partial^2 f^{-1}(\lambda; \theta_0)}{\partial \theta \partial \theta^T} b \right)^2 g^2(\lambda) d\lambda,
\]

where \( g(\lambda) = \sigma_0^2 f(\lambda; \theta_0)/(2\pi) \). This establishes part(iv) and the same argument as above establishes part(v), namely

\[
n \text{Var}_* (\ell_n^*(\theta_0)) = \frac{(2\pi)^2}{n} \sum_{m=1}^N \left[ \frac{\partial f^{-1}(\lambda_m; \theta_0)}{\partial \theta} \right] \left[ \frac{\partial f^{-1}(\lambda_m; \theta_0)}{\partial \theta} \right]^T \var_0^2(\lambda_m) \text{Var}_* (\varepsilon_n^*)
\]

\[
\xrightarrow{p} 2\pi \int_0^\pi \left[ \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta} \right] \left[ \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta} \right]^T g^2(\lambda) d\lambda.
\]

\( \square \)

**Lemma 10.** Under assumptions of Theorem 1 and assuming the results of Theorem 1 hold, as \( n \rightarrow \infty \),

\[
\sqrt{n} (\ell_n^*(\theta_0) - E_* \ell_n^*(\theta_0)) \xrightarrow{d} N \left( 0_p, \sigma_0^2 D_0 \right) \quad \text{in probability as} \quad n \rightarrow \infty.
\]
Proof: Let

\[ Z^*_n \equiv \sqrt{n}(\ell_n^*(\theta_0) - E_\varepsilon \ell_n^*(\theta_0)) = \sqrt{n} \left( \frac{2\pi}{n} \sum_{m=1}^{N} \frac{\partial f^{-1}(\lambda_m; \theta_0)}{\partial \theta} g_n(\lambda_m)(\varepsilon_m^* - 1) \right) \]

and

\[ Z^*_n_1 \equiv \sqrt{n} \left( \frac{2\pi}{n} \sum_{m=1}^{N} \frac{\partial f^{-1}(\lambda_m; \theta_0)}{\partial \theta} g(\lambda_m)(\varepsilon_m^* - 1) \right). \]

We will first show that

\[ Z^*_n_1 \xrightarrow{d} N(0, \sigma_0^2 D_0) \text{ in probability as } n \to \infty. \]

Note that by the iid property of the bootstrap innovations \( \varepsilon_m^* \) with \( E_\varepsilon \varepsilon_m^* = 1, m = 1, 2, \ldots, N, \) we can get

\[ \text{Var}_s(Z^*_n_1) = \frac{(2\pi)^2}{n} \sum_{m=1}^{N} \left( \frac{\partial f^{-1}(\lambda_m; \theta_0)}{\partial \theta_0} \right)^T g^2(\lambda_m) \text{Var}_s(\varepsilon_1^*). \]

By A.4, A.5 and the Dominated Convergence Theorem, we have

\[ \left( \frac{2\pi}{n} \sum_{m=1}^{N} \left( \frac{\partial f^{-1}(\lambda_m; \theta_0)}{\partial \theta_0} \right) \right)^T g^2(\lambda_m) \to 2\pi \int_0^{\pi} \left( \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta_0} \right)^T g^2(\lambda) d\lambda = \sigma_0^2 D_0, \]

and we also have

\[ W_n \equiv \text{Var}_s(\varepsilon_1^*) \xrightarrow{p} 1 \quad (4.21) \]

by Lemma 9, implying \( V_n \equiv \text{Var}_s(Z^*_n_1) \xrightarrow{p} \sigma_0^2 D_0 \equiv V, \) where \( V \) is positive definite by A.6.

Also, by Lemma 8, as \( n \to \infty, \) for \( m = 1, 2, \ldots, N, \)

\[ Y_n \equiv \max_{1 \leq m \leq N} U_m = o_p(n^{1/2}), \quad S_n \equiv \frac{1}{N} \sum_{m=1}^{N} U_m \xrightarrow{p} 1, \quad \text{for} \quad U_m \equiv \frac{I_n(\lambda_m)}{g_n(\lambda_m)}. \quad (4.22) \]

Let \( n_k \) be any subsequence of \( \{n\}. \) By (4.21) and (4.22), there exists a subsequence \( \{n_j\} \subset \{n_k\} \) such that

\[ V_{n_j} \to V, \quad S_{n_j} \to 1, \quad \frac{Y_{n_j}}{n_j^{1/2}} \to 0 \quad \text{and} \quad W_{n_j} \to 1 \quad \text{almost surely (a.s.)} \quad n \to \infty. \quad (4.23) \]
Pick \( \varepsilon > 0 \) and \( b \in \mathbb{R}^p \) such that \( b^T b = 1 \). Define

\[
X_{m,b}^* = \frac{2\pi b^T \partial f^{-1}(\lambda_m; \theta_0)}{\sqrt{n}} g(\lambda_m) (\varepsilon_m^* - 1), \quad m = 1, \ldots, N,
\]

\[
V_{n,b} = b^T V_n b = \text{Var}_* \left( \sum_{m=1}^N X_{m,b}^* \right),
\]

\[
G_{n,b,\varepsilon} = \frac{1}{V_{n,b}} \sum_{m=1}^{N_n} E_* \left| X_{m,b}^* \right|^2 \mathbb{I} \left( \left| X_{m,b}^* \right| > \varepsilon \sqrt{V_{n,b}} \right)
\]

where \( N_n \equiv N = \lfloor (n - 1)/2 \rfloor \) and we suppress the dependence of \( X_{m,b}^* \) on \( n \) in our notation here. By A.4 and A.5,

\[
\sup_{b \in \mathbb{R}^p, \lambda \in \Pi, ||b||=1} \left| b^T \partial f^{-1}(\lambda; \theta_0) g(\lambda) \right| \leq C
\]

for some \( C > 0 \). Considering \( G_{n,b,\varepsilon} \) along the subsequence \( n_j \), we may bound

\[
G_{n_j,b,\varepsilon} \leq \frac{C^2(2\pi)^2}{n_j V_{n_j,b}} \sum_{m=1}^{N_{n_j}} E_* \left| \varepsilon_m^* - 1 \right|^2 \mathbb{I} \left( 2\pi C |\varepsilon_m^* - 1| > \sqrt{n_j} \varepsilon \sqrt{V_{n_j,b}} \right)
\]

\[
\leq \frac{C^2(2\pi)^2}{n_j V_{n_j,b}} \sum_{m=1}^{N_{n_j}} E_* \left| \varepsilon_m^* - 1 \right|^2 \mathbb{I} \left( \max_{1 \leq k \leq N} \left| S_{n_j}^{-1} I_{n_j}(\lambda_k) / g_{n_j}(\lambda_k) \right| - 1 > \varepsilon \sqrt{n_j} \sqrt{V_{n_j,b}} / 2\pi C \right)
\]

\[
\leq \frac{C^2(2\pi)^2}{n_j V_{n_j,b}} N_{n_j} W_{n_j} \left[ \mathbb{I} \left( \frac{Y_{n_j}}{\sqrt{n_j}} > \frac{S_{n_j} \varepsilon \sqrt{V_{n_j,b}}}{4\pi C} \right) + \mathbb{I} \left( 1 > \frac{\sqrt{n_j} \varepsilon \sqrt{V_{n_j,b}}}{4\pi C} \right) \right].
\]

using that \( \varepsilon_m^* \) assumes values among \( S_{n_j}^{-1} I_{n_j}(\lambda_k) / g_{n_j}(\lambda_k) \), \( k = 1, \ldots, N_{n_j} \). As \( n_j \to \infty \), by (4.22)

\[
V_{n_j} \to b^T V b > 0, \quad S_{n_j} \to 1 \quad \text{and} \quad \frac{Y_{n_j}}{\sqrt{n_j}} \to 0 \quad \text{a.s.}
\]

which implies

\[
\mathbb{I} \left( \frac{Y_{n_j}}{\sqrt{n_j}} > \frac{S_{n_j} \varepsilon \sqrt{V_{n_j,b}}}{4\pi C} \right) \to 0 \quad \text{and} \quad \mathbb{I} \left( 1 > \frac{\sqrt{n_j} \varepsilon \sqrt{V_{n_j,b}}}{4\pi C} \right) \to 0 \quad \text{a.s.}
\]

as \( n \to \infty \). Since additionally, by (4.22), we have

\[
\frac{W_{n_j}}{V_{n_j,b}} \to \frac{1}{b^T V b} \quad \text{a.s.} \quad \text{and} \quad \frac{N_{n_j}}{n_j} \to \frac{1}{2} \quad \text{as} \quad n \to \infty,
\]

so that it follows that \( G_{n_j,b,\varepsilon} \to 0 \) a.s. In fact, since \( b \in \mathbb{R}^p \), and \( \varepsilon > 0 \) are arbitrary, it holds that

\[
P \left( \lim_{n_j \to \infty} G_{n_j,b,\varepsilon} = 0 \quad \text{for any} \quad b \in \mathbb{R}^p, \varepsilon > 0 \right) = 1.
\]
As a consequence of this almost surely convergence of $G_{n_j,b,\epsilon}$ (for any $b,\epsilon$), it holds that

\[ Z_{1n_j}^* \xrightarrow{d} N(0_p, V) \quad \text{a.s. as } n_j \to \infty, \]

by the Lindeberg-Feller CLT and Cramér-Wold device. In other words,

\[ P \left( \lim_{n_j \to \infty} P_* \left( Z_{1n_j}^* \leq x \right) = \Phi(x, V), \forall x \in \mathbb{R}^p \right) = 1 \]

where $\Phi(x, V)$ is the cdf of a $N(0_p, V)$ or by Polya’s theorem,

\[ P \left[ \lim_{n_j \to \infty} \left( \sup_{x \in \mathbb{R}^p} |P_* \left( Z_{1n_j}^* \leq x \right) - \Phi(x, V)| \right) = 0 \right] = 1. \]

Since the original subsequence $n_k$ was arbitrary, we have

\[ Z_{1n}^* \xrightarrow{d} N(0_p, V) \quad \text{in probability as } n \to \infty, \]

i.e., for any $\epsilon > 0$,

\[ P \left( \sup_{x \in \mathbb{R}^p} |P_* \left( Z_{1n}^* \leq x \right) - \Phi(x, V)| > \epsilon \right) \xrightarrow{n \to \infty} 0. \]

Now, let $\tilde{g}_n(\lambda_m) \equiv \frac{\sigma_0^2 f(\lambda_m; \hat{\theta}_n)}{(2\pi)}$ and define

\[ Z_{2n}^* \equiv \sqrt{n} \left( \frac{2\pi}{n} \sum_{m=1}^{N} \frac{\partial f^{-1}(\lambda_m; \hat{\theta}_n)}{\partial \theta} \tilde{g}_n(\lambda_m) \left( \varepsilon_m^* - 1 \right) \right). \]

In the following, we will show

\[ E_* \|Z_{1n}^* - Z_{2n}^*\|^2 \xrightarrow{p} 0 \quad \text{and} \quad E_* \|Z_n^* - Z_{2n}^*\|^2 \xrightarrow{p} 0, \quad (4.25) \]

which implies $E_* \|Z_{1n}^* - Z_n^*\|^2 \xrightarrow{p} 0$. Then given any subsequence $\{n_k\}$ of $\{n\}$, we may choose a further subsequence $\{n_j\}$ of $\{n_k\}$ such that

\[ Z_{1n_j}^* \xrightarrow{d} N(0_p, V) \quad \text{and} \quad E_* \|Z_{n_j}^* - Z_{1n_j}^*\|^2 \xrightarrow{} 0 \quad \text{a.s. as } n \to \infty. \]

This implies

\[ Z_{1n_j}^* \xrightarrow{d} N(0_p, V) \quad \text{and} \quad Z_{n_j}^* - Z_{1n_j}^* \xrightarrow{p} 0 \quad \text{a.s. as } n \to \infty \]

so that by the continuous mapping theorem,

\[ Z_{n_j}^* = Z_{1n_j}^* + Z_{n_j}^* - Z_{1n_j}^* \xrightarrow{d} N(0, V) \quad \text{a.s..} \]
Since \( \{n_k\} \) was an arbitrary subsequence, we have

\[
Z_{n_k}^* \overset{d}{\longrightarrow} \mathbf{N}(0, \mathbf{V}) \text{ in probability as } n \to \infty,
\]

establishing Lemma 10.

To show (4.25), we use \( E_* Z_{n_k}^* = 0 = E_* Z_{2n_k}^* \) along with A.4 and A.5 to expand

\[
E_* \|Z_{2n_k}^* - Z_n^*\|^2 = \text{tr} \left( \frac{\sigma_n^2 - \sigma_0^2}{n} \sum_{m=1}^{N} \left( \frac{\partial f^{-1}(\lambda_m; \theta_0)}{\partial \theta_0} \right) \left( \frac{\partial f^{-1}(\lambda_m; \theta_0)}{\partial \theta_0} \right)^T f^2(\lambda_m; \theta_0)\text{Var}_*(\varepsilon_1^*) \right)
\]

\[
= C(\sigma_n^2 - \sigma_0^2)\text{Var}_*(\varepsilon_1^*) \overset{p}{\longrightarrow} 0 \text{ as } n \to \infty
\]

by Lemma 9(iii) and Theorem 1. Next, by A.4 and A.5 and \( E_* Z_{1n}^* = 0 \), we have

\[
E_* \|Z_{1n}^* - Z_{2n_k}^*\|^2 \leq \frac{C}{n}\text{Var}_*(\varepsilon_1^*)^2 \sum_{m=1}^{N} |\lambda_m|^{4d} \left( f(\lambda_m; \theta_0) - f(\lambda_m; \hat{\theta}_n) \right)^2 \equiv C\text{Var}_*(\varepsilon_1^*)T_n.
\]

To show \( E_* \|Z_{1n}^* - Z_{2n_k}^*\|^2 \overset{p}{\longrightarrow} 0 \), it suffices by Lemma 9 to show \( T_n \overset{p}{\longrightarrow} 0 \) as \( n \to \infty \). When \( \hat{\theta}_n \in \mathcal{B} \) by Theorem 1 for \( \mathcal{B} \) defined in A.5, then \( |\lambda|^{4d} \left( f(\lambda; \theta_0) - f(\lambda; \hat{\theta}_n) \right)^2 \leq C|\lambda|^{-2\eta} \), \( \lambda \in (0, \pi] \) where \( \int_0^\pi |\lambda|^{-2\eta}d\lambda < \infty \) by \( -2\eta > -1 \). Hence, using the Dominated Convergence Theorem with the same arguments for showing (4.18), it follows that \( T_n \overset{p}{\longrightarrow} 0 \), establishing (4.25) and completing the proof of Lemma 10. \( \square \)

### 4.7 Proofs of main results

#### 4.7.1 Proof of Theorem 1

We first establish the consistency results in Theorem 1(i). Let \( \psi_{j,k}(\lambda; \theta) \equiv \frac{\partial^2 f^{-1}(\lambda, \theta)}{\partial \theta_j \partial \theta_k} \) for \( j, k = 1, \ldots, p, \lambda \in \Pi, \theta \in \Theta \). Pick \( \delta > 0 \) so that for a closed ball around \( \theta_0, \mathcal{B} \equiv \{ \theta \in \mathbb{R}^p : ||\theta - \theta_0|| \leq \delta \} \subset \Theta \), it holds that

\[
\sup_{\lambda \in \Pi} \sup_{\theta \in \mathcal{B}} |\psi_{j,k}(\lambda; \theta) - \psi_{j,k}(\lambda; \theta_0)| \leq \varepsilon \equiv \frac{1}{4r(0)||D_{0}^{-1}||}
\]

(4.26)
for \( r(0) = \text{Var}(X_1) \) which is possible because \( \Pi \times \mathcal{B} \) is compact and \( \psi_{j,k}(\lambda; \theta) \) is continuous at all \((\lambda, \theta)\) from \( A.3 \) (i.e., \( \psi_{j,k}(\lambda; \theta) \) is uniformly continuous on \( \Pi \times \mathcal{B} \)). Note that by \( A.1 \),

\[
D_0 = \frac{2\pi}{\sigma^2_0} \int_0^\pi \left( \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta} \right) \left( \frac{\partial f^{-1}(\lambda; \theta_0)}{\partial \theta} \right)^T g^2(\lambda)d\lambda = \int_0^\pi \frac{\partial^2 f^{-1}(\lambda; \theta)}{\partial \theta \partial \theta^T} g(\lambda)d\lambda
\]

which is positive definite by \( A.6 \). By \( A.3 \), for \( \theta \in \mathcal{B} \), we apply Taylor’s expansion around \( \theta_0 \) to obtain

\[
\ell_n(\theta) = \ell_n(\theta_0) + D_n(\theta_0)(\theta - \theta_0) + R_n(\theta)(\theta - \theta_0),
\]

defining

\[
D_n(\theta) \equiv \frac{2\pi}{n} \sum_{m=1}^N \frac{\partial^2 f^{-1}(\lambda_m; \theta)}{\partial \theta \partial \theta^T} I_n(\lambda_m), \quad \theta \in \Theta,
\]

and a remainder \( R_n(\theta) = D_n(\theta^+) - D_n(\theta_0) \) for some \( \theta^+ \in \mathcal{B} \) depending on \( \theta \) with \( ||\theta^+ - \theta_0|| \leq ||\theta - \theta_0|| \). Lemma 4 implies that

\[
D_n(\theta_0) \xrightarrow{p} D_0 \quad \text{as} \quad n \to \infty.
\]

By (4.28) and \( A.5 \), we have that

\[
||D_n(\theta_0) - D_0|| \leq \frac{\delta^*}{||D_0||}
\]

holds for \( \delta^* \equiv 1/2 \) with probability arbitrarily close to 1 for large \( n \). Lemma 4.2 of Lahiri (2003) implies that \( D_n(\theta_0) \) is nonsingular whenever (4.29) holds and that

\[
||D_n^{-1}(\theta_0)|| \leq \frac{1}{1 - \delta^*} ||D_0^{-1}|| = 2 ||D_0^{-1}||.
\]

By Markov’s inequality, we get

\[
P \left( ||\ell_n(\theta_0)|| > n^{-1/2} \log^2 n \right) \leq \frac{4n}{\log^4 n} \left\{ \mathbb{E} ||\ell_n(\theta_0) - \mathbb{E}\ell_n(\theta_0)||^2 + ||\mathbb{E}\ell_n(\theta_0)||^2 \right\}
\]

\[
= \frac{4n}{\log^4 n} \sum_{j=1}^p \text{Var} \left( e_j^T \ell_n(\theta_0) \right) + \frac{4n}{\log^4 n} ||\mathbb{E}\ell_n(\theta_0)||^2
\]

where \( e_j = (0, \ldots, 0, 1, 0, \ldots, 0)^T \in \mathbb{R}^p \) is the standard basis vector with the \( j \)-th element being 1 for \( j = 1, \ldots, p \). Then, Lemmas 5 and 6 show that as \( n \to \infty \),

\[
\frac{4n}{\log^4 n} \sum_{j=1}^p \text{Var} \left( e_j^T \ell_n(\theta_0) \right) = o(1) \quad \text{and} \quad \frac{4n}{\log^4 n} ||\mathbb{E}\ell_n(\theta_0)||^2 = o(1),
\]
so that
\[ P \left( \| \ell_n(\theta_0) \| \leq n^{-1/2} \log^2 n \right) \longrightarrow 1 \quad \text{as} \quad n \to \infty. \] (4.31)

Define the event
\[ A_n \equiv \left\{ \| \ell(\theta_0) \| \leq n^{-1/2} \log^2 n, \left\| D_n^{-1}(\theta_0) \right\| \leq 2 \left\| D_0^{-1} \right\|, \frac{2\pi}{n} \sum_{m=1}^{N} I_n(\lambda_m) \leq r(0) \right\}. \]

By Lemma 4,
\[ \frac{2\pi}{n} \sum_{m=1}^{N} I_n(\lambda_m) \xrightarrow{p} \frac{1}{2} r(0) > 0 \quad \text{as} \quad n \to \infty \] (4.32)
so that this with (4.28)-(4.31) gives
\[ P(A_n) \longrightarrow 1 \quad \text{as} \quad n \to \infty. \]

Next, we define a function
\[ q(\theta_0 - \theta) \equiv D_n^{-1}(\theta_0) \left\{ \ell_n(\theta_0) + R_n(\theta)(\theta - \theta_0) \right\}, \quad \theta \in \mathcal{B}, \]
based on (4.26)-(4.28). With \( \varepsilon \) in (4.26), we may bound
\[ \sup_{\theta \in \mathcal{B}} \left\| R_n(\theta) \right\| \leq \sup_{\lambda_m \in \Pi, \theta \in \mathcal{B}} \left| \psi_{j,k}(\lambda_m; \theta^+) - \psi_{j,k}(\lambda_m; \theta_0) \right| \frac{2\pi}{n} \sum_{m=1}^{N} I_n(\lambda_m) \leq \varepsilon r(0) \]

on \( A_n \). Let \( C_1 \equiv 4 \left\| D_0^{-1} \right\| \). For large \( n \), on the set \( A_n \), if \( \| \theta - \theta_0 \| \leq C_1 n^{-1/2} \log^2 n \), we have
\[ \| q(\theta_0 - \theta) \| \leq \left\| D_n^{-1}(\theta_0) \right\| \left( \left\| \ell_n(\theta_0) \right\| + \left\| R_n(\theta) \right\| \| \theta - \theta_0 \| \right) \]
\[ \leq 2 \left\| D_0^{-1} \right\| \left\{ n^{-1/2} \log^2 n + \varepsilon r(0)C_1 n^{-1/2} \log^2 n \right\} \]
\[ = 2 \left\| D_0^{-1} \right\| n^{-1/2} \log^2 n \{ 1 + 1 \} \]
\[ = C_1 n^{-1/2} \log^2 n. \] (4.33)

Hence, by (4.33), on the set \( A_n \), Brouwer’s Fixed Point Theorem (Lahiri, 2003) implies that there exists \( \hat{\theta}_n \in \mathcal{B} \) with \( \| \hat{\theta}_n - \theta_0 \| \leq C_1 n^{-1/2} \log^2 n \) such that \( \theta_0 - \hat{\theta}_n = q(\theta_0 - \hat{\theta}_n) \) or equivalently \( D_n(\theta_0)(\hat{\theta}_n - \theta_0) = \ell_n(\theta_0) + R_n(\hat{\theta}_n)(\hat{\theta}_n - \theta_0) \) implying that
\[ \ell_n(\hat{\theta}_n) = \ell_n(\theta_0) + D_n(\theta_0) \left( \hat{\theta}_n - \theta_0 \right) + R_n(\hat{\theta}_n) \left( \hat{\theta}_n - \theta_0 \right) = 0_p. \]
That is, on the set $A_n$, there exists a root $\hat{\theta}_n \in \mathcal{B}$ where $\ell_n(\hat{\theta}_n) = 0_p$ and $||\hat{\theta}_n - \theta_0|| \leq C_1 n^{-1/2} \log^2 n$. Since $P(A_n) \to 1$ as $n \to \infty$, we have

$$\hat{\theta}_n - \theta_0 = O_p\left(n^{-1/2} \log^2 n\right),$$

and $\hat{\theta}_n \xrightarrow{p} \theta_0$. From (4.9), we have

$$\hat{\sigma}_n^2 = 2 \cdot 2\pi \sum_{m=1}^{N} f^{-1}(\lambda_m; \theta_0)I_n(\lambda_m) + R_{1n}$$

where

$$|R_{1n}| \leq \frac{2\pi}{n} \sum_{m=1}^{N} I_n(\lambda_m) \cdot \sup_{\lambda \in \Pi} \left| f^{-1}(\lambda; \hat{\theta}_n) - f^{-1}(\lambda; \theta_0) \right|.$$

By (4.32) and (4.34), $|R_{1n}| \xrightarrow{p} 0$ holds and $\frac{2\pi}{n} \sum_{m=1}^{N} f^{-1}(\lambda_m; \theta_0)I_n(\lambda_m) \xrightarrow{p} \int_{\Pi} f^{-1}(\lambda; \theta_0)g(\lambda)d\lambda = \sigma_0^2/2$ holds by Lemma 4, so that $\hat{\sigma}_n^2 \xrightarrow{p} \sigma_0^2$ follows.

We next prove the asymptotic normality in Theorem 1(ii). Recall, using (4.27), that $\hat{\theta}_n - \theta_0 = -D_n^{-1}(\theta_0) \left\{ \ell_n(\theta_0) + R_n(\hat{\theta}_n) \left( \hat{\theta}_n - \theta_0 \right) \right\}$ on the set $A_n$, so that

$$\sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) \left\{ 1 + D_n^{-1}(\theta_0)R_n(\hat{\theta}_n) \right\} = \sqrt{n}\ell_n(\theta_0) \left\{ -D_n^{-1}(\theta_0) \right\}.$$

Note that, as $n \to \infty$,

$$\left\| R_n(\hat{\theta}_n) \right\| \leq \sup_{1 \leq j,k \leq p} \left| \psi_{j,k}(\lambda; \hat{\theta}_n) - \psi_{j,k}(\lambda; \theta_0) \right| \frac{2\pi}{n} \sum_{m=1}^{N} I_n(\lambda_m) = o_p(1)$$

by (3.32) and (4.34) and the fact that $\psi_{j,k}(\lambda; \theta)$ is uniformly continuous on $\mathcal{B}$ by A.3. From this, $D_n^{-1}(\theta_0) \xrightarrow{p} D_0^{-1}$ by (4.28) and $\sqrt{n}\ell_n(\theta_0) \xrightarrow{d} N(0_p, \sigma_0^2 D_0)$ by Lemma 7, we have

$$\sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) \left\{ 1 + O_p(1) \cdot o_p(1) \right\} = \sqrt{n}\ell_n(\theta_0) \left\{ -D_0^{-1} + o_p(1) \right\},$$

and

$$\sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) \xrightarrow{d} N\left(0_p, \sigma_0^2 D_0^{-1}\right) \quad \text{as} \quad n \to \infty$$

by Slutsky’s theorem. □

4.7.2 Proof of Theorem 2

We first consider showing Theorem 2(i). Pick $\delta > 0$ so that the closed ball $\mathcal{B} \equiv \{ \theta \in \mathbb{R}^p : ||\theta - \theta_0|| \leq \delta \} \subset \Theta$ satisfies (4.26) with $\varepsilon = \left(10r(0) \max\{1, ||D_0^{-1}||\} \right)^{-1}$ for $r(0) = \text{Var}(X_t)$, and
where

$$\sup_{\lambda \neq 0, \lambda \in \Pi} \sup_{\theta \in B} f(\lambda; \theta)|\lambda|^{2d+\eta} \leq C(\eta) \quad (4.35)$$

holds for some $C(\eta) > 0$ and $\eta \in (0, 1/2)$ with $2d + \eta < 1$ by $\text{A.5}$. For $\theta \in \mathcal{B}$, let

$$\ell_n^*(\theta) \equiv \frac{2\pi}{n} \sum_{m=1}^{N} \frac{\partial f^{-1}(\lambda_m; \theta_0)}{\partial \theta} I_n^*(\lambda_m) \quad \text{and} \quad D_n^*(\theta) \equiv \frac{2\pi}{n} \sum_{m=1}^{N} \frac{\partial^2 f^{-1}(\lambda_m; \theta_0)}{\partial \theta \partial \theta^T} I_n^*(\lambda_m)$$

where $I_n^*(\lambda_m) = g_n(\lambda_m) \varepsilon_m^* = \hat{\sigma}_n^2 f(\lambda_m; \hat{\theta}_n) \varepsilon_m^*/2\pi$. In addition, $E_n I_n^*(\lambda_m) = \hat{g}_n(\lambda_m)$. By Lemma 9(ii), $E_n D_n^*(\theta_0) \xrightarrow{p} D_0$, implying

$$\|E_n D_n^*(\theta_0)\|^{-1} \leq 2 \|D_0^{-1}\| \quad (4.36)$$

holds with arbitrarily large probability as $n \to \infty$ (similarly to (4.30)). Define a set $A_n$ as the intersection of the following events

$$\|\hat{\theta}_n - \theta_0\| \leq n^{-1/2} \log^2(n), \quad \text{Var}_n \left(\varepsilon_i^*\right) \leq 2$$

$$\frac{2\pi}{n} \sum_{m=1}^{N} E_n I_n^*(\lambda_m) \leq r(0), \quad \left\|E_n D_n^*(\hat{\theta}_n)\right\| \leq 2 \|D_0^{-1}\|,$$

$$\sup_{\lambda \neq 0, \lambda \in \Pi} \hat{g}_n(\lambda)|\lambda|^{2d+\tau} \leq C_0, \quad \max_{1 \leq i, j \leq p} \text{Var}_n \left(e_i^T D_n^*(\theta_0) e_j\right) + \|\text{Var}_n (\ell_n^*(\theta_0))\| \leq \frac{C_0}{n},$$

using standard coordinate vectors $e_i \in \mathbb{R}^p$, $i = 1, \ldots, p$ and an appropriate constant $C_0 > 0$. We can choose $C_0$ so that $P(A_n) \to 1$ as $n \to \infty$, using (4.35), (4.36), Theorem 1(ii), Lemma 8(v) and Lemmas 9(iv),(v). By $\text{A.3}$, we use Taylor's theorem to expand $\ell_n^*(\theta) - E_n \ell_n^*(\hat{\theta}_n)$, $\theta \in \mathcal{B}$, around $\theta_0$ as

$$\ell_n^*(\theta) - E_n \ell_n^*(\hat{\theta}_n) = \ell_n^*(\theta_0) - E_n \ell_n^*(\theta_0) + E_n D_n^*(\theta_0) \left(\theta - \hat{\theta}_n\right) + R_n^*(\theta), \quad (4.37)$$

where

$$R_n^*(\theta) = (D_n^*(\theta_0) - E_n D_n^*(\theta_0)) (\theta - \theta_0) + (D_n^*(\theta_{1n}) - D_n^*(\theta_0)) (\theta - \theta_0) \quad (4.38)$$

$$+ E_n \left(D_n^*(\theta_{2n}) - D_n^*(\theta_0)\right) \left(\theta_0 - \hat{\theta}_n\right)$$

where $\theta_{1n}, \theta_{2n} \in \mathcal{B}$ depend on $\theta$ and $\hat{\theta}_n$ so that $\|\theta_{1n} - \theta_0\| \leq \|\theta - \theta_0\|$ and $\|\theta_{2n} - \theta_0\| \leq \|\hat{\theta}_n - \theta_0\|$, respectively. Let $A_n^*$ be the bootstrap event defined as

$$A_n^* = \left\{ \max_{1 \leq i, j \leq p} \left| e_i^T \left[ D_n^*(\theta_0) - E_n D_n^*(\theta_0)\right] e_j \right| \leq n^{-1/2} \log n, \right.$$\begin{align*}
\left. \|\ell_n^*(\theta_0) - E_n \ell_n^*(\theta_0)\| \leq n^{-1/2} \log n, \sup_{\theta \in \mathcal{B}} \|D_n^*(\theta) - D_n^*(\theta_0)\| \leq 2\varepsilon r(0) \right\},
\end{align*}
and define a function

\[ q_n^*(\theta - \hat{\theta}_n) \equiv [E_* D_n^*(\theta_0)]^{-1} \{ \ell_n^*(\theta_0) - E_* \ell_n^*(\theta_0) + R_n^*(\theta) \}, \quad \theta \in B. \]

On the set \( A_n \), if the bootstrap event \( A_n^* \) holds and if \( \| \hat{\theta}_n - \theta \| \leq C_1 n^{-1/2} \log n \) for \( C_1 \equiv 8 \| D_0^{-1} \| \), then

\[
\| q_n^*(\theta - \hat{\theta}_n) \| \leq 2 \| D_0^{-1} \| \left\{ n^{-1/2} \log n(C_1 n^{-1/2} \log n + n^{-1/2} \log n) + n^{-1/2} \log n + 2\varepsilon r(0)(C_1 + 1)n^{-1/2} \log n + 2\varepsilon r(0)n^{-1/2} \log n \right\} \\
\leq C_1 n^{-1/2} \log n,
\]

provided that \( n \) is large so that \( (C_1 + 1)n^{-1/2} \log n \leq 1 \) and using \( \| \theta - \theta_0 \| \leq \| \theta - \hat{\theta}_n \| + \| \hat{\theta}_n - \theta_0 \| \).

Then, as in the proof of Theorem 1, by Brouwer’s Fixed Point theorem, there exists \( \hat{\theta}_n^* \in B \) with \( \| \hat{\theta}_n^* - \hat{\theta}_n \| \leq C_1 n^{-1/2} \log n \) such that \( q_n^*(\hat{\theta}_n - \hat{\theta}_n^*) = \hat{\theta}_n - \hat{\theta}_n \) or

\[ \ell_n^*(\hat{\theta}_n^*) - E_* \ell_n^*(\hat{\theta}_n) = 0_p. \]

Hence, on the set \( A_n \) with large \( n \) so that \( C_1 n^{-1/2} \log n \leq 1 \), we use the definition of \( A_n \) to bound bootstrap moments in the following

\[
P_\star \left( \text{There exists a bootstrap solution } \hat{\theta}_n^* \text{ with } \| \hat{\theta}_n^* - \hat{\theta}_n \| \leq C_1 n^{-1/2} \log n \right) \\
\geq P_\star (A_n^*) \\
\geq 1 - \sum_{j=1}^{\rho} \sum_{k=1}^{\rho} P_\star \left( \| e_j^T [D_n^*(\theta_0) - E_* D_n^*(\theta_0)] e_k \| > n^{-1/2} \log n \right) \\
-P_\star \left( \| \ell_n^*(\theta_0) - E_* \ell_n^*(\theta_0) \| > n^{-1/2} \log n \right) \\
-P_\star \left( \sup_{\theta \in B} \| D_n^*(\theta) - D_n^*(\theta_0) \| > \varepsilon r(0) \right) \\
\geq 1 - \sum_{j=1}^{\rho} \sum_{k=1}^{\rho} \frac{\text{Var}_\star \left( e_j^T [D_n^*(\theta_0) - E_* D_n(\theta_0)] e_k \right)}{n \log^2 n} - \frac{C \| \text{Var}_\star (\ell_n^*(\theta_0)) \|}{n \log^2 n} \\
-P_\star \left( \frac{2\pi}{n} \sum_{m=1}^{N} I_n^*(\lambda_m) > 2r(0) \right) \\
\geq 1 - \frac{C_0 \rho^2}{n^2 \log^2 n} - \frac{CC_0}{n^2 \log^2 n} - \frac{\text{Var}_\star \left( \frac{2\pi}{n} \sum_{m=1}^{N} I_n^*(\lambda_m) \right)}{\left( 2r(0) - \frac{2\pi}{n} \sum_{m=1}^{N} E_* I_n^*(\lambda_m) \right)^2}.
\]
\[
\geq 1 - \frac{C}{n^2 \log^2 n} - \left(\frac{2\pi}{n}\right)^2 \sum_{m=1}^{N} \hat{g}_n^2(\lambda_m) \text{Var}_*(e_j^* e_j^*') \frac{1}{r^2(0)}
\]
\[
\geq 1 - \frac{C}{n^2 \log^2 n} - \frac{C}{n} \max_{1 \leq m \leq N} \hat{g}_n(\lambda_m) \geq 1 - \frac{C}{n^2 \log^2 n} - \frac{CC_0 n^{-2d-\eta}}{n} \to 1
\]

since \(2d + \eta < 1\). Hence, on the set \(A_n\), as \(n \to \infty\),

\[
P_*(\text{a bootstrap estimator } \hat{\theta}_n^* \text{ exists and } \|\hat{\theta}_n^* - \hat{\theta}_n\| \leq C_1 n^{-1/2} \log n) \to 1
\]

and, since \(P(A_n) \to 1\) as \(n \to \infty\), we have established Theorem 2(i).

To show Theorem 2(ii), note that, for any \(\epsilon > 0\),

\[
P_*(\sup_{\|\theta - \theta_0\| \leq 2C_1 n^{-1/2} \log n} \|D_n^*(\theta) - D_n^*(\theta_0)\| > \epsilon)
\]
\[
\leq \frac{1}{\epsilon} \sup_{\|\theta - \theta_0\| \leq 2C_1 n^{-1/2} \log n} \max_{\lambda \in \Pi, 1 \leq i, k \leq p} |\psi_{j,k}(\theta, \lambda) - \psi_{j,k}(\theta_0, \lambda)| E_* \left(\frac{2\pi}{n} \sum_{m=1}^{N} I_n^*(\lambda_m)\right)
\]
\[
= o(1) \cdot \frac{2\pi}{n} \sum_{m=1}^{N} \hat{g}_n(\lambda_m) = o(1) O_p(1) = o_p(1)
\]

by A.4 and Lemma 8(v), while likewise

\[
P_*(\sqrt{n} \|\hat{\theta}_n - \theta_0\| \sup_{\|\theta - \theta_0\| \leq 2C_1 n^{-1/2} \log n} \|D_n^*(\theta) - D_n^*(\theta_0)\| > \epsilon) = o_p(1)
\]

and

\[
\max\{1, \sqrt{n} \|\hat{\theta}_n - \theta_0\|\} \sup_{\|\theta - \theta_0\| \leq 2C_1 n^{-1/2} \log n} E_* \|D_n^*(\theta) - D_n^*(\theta_0)\| = O_p(1) o_p(1) = o_p(1)
\]

by Theorem 1(ii). Also,

\[
P_*(\|D_n^*(\theta_0) - E_* D_n^*(\theta_0)\|^2 > \epsilon) \leq \frac{1}{n \epsilon} \sum_{j=1}^{p} n \text{Var}_*(e_j^* D_n^*(\theta_0)e_j) = O\left(\frac{1}{n}\right) O_p(1) = o_p(1)
\]

by Lemma 9(iv). Using \(P(A_n) \to 1\) with the Borel-Cantelli lemma along with Lemma 9(ii) and Lemma 10 with the above results, let \(\{n_k\}\) be any subsequence of \(\{n\}\) and take a further
subsequence \( \{n_j\} \) of \( \{n_k\} \) so that the events

\[
\begin{align*}
B_1 & \equiv \left\{ \sqrt{n_j} \left( \ell_{n_j}^* (\theta_0) - E_* \ell_{n_j}^* (\theta_0) \right) \overset{d}{\rightarrow} N \left( 0, \sigma^2_0 D_0 \right) \right\}, \\
B_2 & \equiv \left\{ E_* D_{n_j}^* (\theta_0) \rightarrow D_0 \right\}, \\
B_3 & \equiv \left\{ P_* \left( \left\| D_{n_j}^* (\theta_0) - E_* D_{n_j}^* (\theta_0) \right\| \geq \frac{1}{m} \right) \rightarrow 0, \forall m \geq 1 \right\}, \\
B_4 & \equiv \left\{ P_* \left( \sup_{\| \theta-\theta_0 \| \leq 2c_{1} n^{-1/2} \log n} \left\| D_{n_j}^* (\theta) - D_{n_j}^* (\theta_0) \right\| > \frac{1}{m} \right) \rightarrow 0, \forall m \geq 1 \right\}, \\
B_5 & \equiv \left\{ P_* \left( \sqrt{n_j} \left\| \hat{\theta}_{n_j} - \theta_0 \right\| \sup_{\| \theta-\theta_0 \| \leq 2c_{1} n^{-1/2} \log n} \left\| D_{n_j}^* (\theta) - D_{n_j}^* (\theta_0) \right\| > \frac{1}{m} \right) \rightarrow 0, \forall m \geq 1 \right\}, \\
B_6 & \equiv \left\{ P_* \left( \left\| \hat{\theta}_{n_j} - \theta_0 \right\| \leq C_1 n^{-1/2} \log n_j \right) \rightarrow 1 \right\}, \text{ and} \\
B_7 & \equiv \left\{ \max\{1, \sqrt{n_j} \left\| \hat{\theta}_{n_j} - \theta_0 \right\| \sup_{\| \theta-\theta_0 \| \leq 2c_{1} n^{-1/2} \log n} E_* \left\| D_n^* (\theta) - D_n^* (\theta_0) \right\| \rightarrow 0 \right\}
\end{align*}
\]

have probability 1, i.e., \( P(B) = 1 \) for \( B = \bigcap_{i=1}^7 B_i \). For large \( n \), pointwise on \( B \), we use \((4.37)\) and \((4.38)\) to write

\[
\hat{\theta}_{n_j}^* - \hat{\theta}_{n_j} = \left[ E_* D_{n_j}^* (\theta_0) \right]^{-1} \left\{ \ell_{n_j}^* (\theta_0) - E_* \ell_{n_j}^* (\theta_0) + R_{n_j}^* (\hat{\theta}_{n_j}) \right\}
\]

where

\[
\left\| R_{n_j}^* (\hat{\theta}_{n_j}) \right\| \leq \left( \sup_{\| \theta-\theta_0 \| \leq 2c_{1} n^{-1/2} \log n} \left\| D_{n_j}^* (\theta) - D_{n_j}^* (\theta_0) \right\| + E_* \left\| D_{n_j}^* (\theta) - D_{n_j}^* (\theta_0) \right\| \right.
\]

\[
+ \left. \left\| D_{n_j}^* (\theta_0) - E_* D_{n_j}^* (\theta_0) \right\| \left( \left\| \hat{\theta}_{n_j}^* - \hat{\theta}_{n_j} \right\| + \left\| \hat{\theta}_{n_j} - \theta_0 \right\| \right) \right) .
\]

As \( n_j \rightarrow \infty \) pointwise on \( B \), we have

\[
\sqrt{n_j} \left\| \hat{\theta}_{n_j}^* - \hat{\theta}_{n_j} \right\| (1 + o_p (1)) = O_p (1)
\]

implying \( \sqrt{n_j} \left\| \hat{\theta}_{n_j}^* - \hat{\theta}_{n_j} \right\| = O_p (1) \) and consequently \( \sqrt{n_j} \left\| R_{n_j}^* (\hat{\theta}_{n_j}) \right\| = o_p (1) \), where here \( p^* \) denotes the bootstrap probability along the sequence of bootstrap distributions for a point in \( B \) (i.e., the induced sequence of bootstrap distributions conditional on the data \( X_i(w), i \geq 1 \) for a point \( w \in B \)). Hence, pointwise on \( B \) as \( n_j \rightarrow \infty \),

\[
\sqrt{n_j} \left( \hat{\theta}_{n_j}^* - \hat{\theta}_{n_j} \right) \overset{d}{\rightarrow} N \left( 0_p, \sigma^2 D_0^{-1} \right).
\]
Since the subsequence \( \{n_k\} \) was arbitrary, we now have that 
\[
T^*_n = \sqrt{n}(\hat{\theta}^*_n - \hat{\theta}_n) \xrightarrow{d} N\left(0_p, \sigma^2D_0^{-1}\right)
\]
in probability as \( n \to \infty \); in other words, using Poyla’s Theorem,
\[
\sup_{x \in \mathbb{R}^p} \left| P^*\left(T^*_n \leq x\right) - \Phi(x, \sigma^2D_0^{-1}) \right| \xrightarrow{p} 0
\]
as \( n \to \infty \), where \( \Phi(x, \sigma^2D_0^{-1}) \) denotes the distribution function of a \( N(0_p, \sigma^2D_0^{-1}) \) distribution.

For 
\[
T_n = \sqrt{n}(\hat{\theta}_n - \theta_0),
\]
\[
\sup_{x \in \mathbb{R}^p} \left| P(T_n \leq x) - \Phi(x, \sigma^2D_0^{-1}) \right| \to 0
\]
follows by Theorem 1(ii), implying the consistency result 
\[
\sup_{x \in \mathbb{R}^p} \left| P^*(T^*_n \leq x) - P(T_n \leq x) \right| \xrightarrow{p} 0
\]
of Theorem 2(ii). \( \square \).
References


4.8 Tabulated simulation results

This section provides the coverage rates from all methods (e.g., FDB or normal approximation), processes and sample sizes described in the simulation study of Section 4.4. The coverage rates of 95% confidence intervals (CIs), or joint confidence regions (CRs), for Whittle parameters are listed in tabular form along with the average/median lengths of CIs (or the average/median volume of CRs). The reported average lengths of one-sided intervals correspond to the average value of the upper or lower bound creating the interval. For a given data set, the volume of a joint CR (i.e., an ellipsoid in $\mathbb{R}^2$) was computed as

$$\frac{2\pi^2}{n\sqrt{|\det(A/c)|}}$$

where $A$ denotes an estimator of the limiting covariance matrix $V \equiv \sigma_0^2D_0^{-1}$ of Whittle estimators from Theorem 1(ii) (e.g., $A = \hat{V}_n$ or $\hat{V}_{n,p}$ from Section ??) and $c$ denotes a quantile (obtained by a chi-square/normal approximation or estimated by a FDB method) needed to calibrate a 95% confidence level. All coverage rates and average lengths (or volumes) were approximated by 4000 simulations from each process and sample size. All FDB methods were implemented with 1000 bootstrap resamples per data set.

Figures 4.3-4.10 reported in Section 4.4 were obtained from coverage rates listed in the following Tables 4.3-4.20. The correspondence is as follows

- Figure 4.3 is based on Tables 4.3-4.4 (FARIMA(0, $d$, 0), normal innovations).
- Figure 4.4 is based on Tables 4.5-4.6 (FARIMA(0, $d$, 0), chi-square innovations).
- Figure 4.5 is based on Tables 4.7-4.8 (FARIMA(0, $d$, 0), student-t innovations).
- Figure 4.6 is based on Tables 4.9, 4.10 and 4.13 (FARIMA(1, $d$, 0), normal innovations).
- Figure 4.7 is based on Tables 4.11, 4.12 and 4.13 (FARIMA(1, $d$, 0), normal innovations).
- Figure 4.8 is based on Tables 4.15, 4.16 and 4.19 (FARIMA(0, $d$, 1), normal innovations).
- Figure 4.9 is based on Tables 4.17, 4.18 and 4.19 (FARIMA(0, $d$, 1), normal innovations).
- Figure 4.10 is based on Tables 4.14 and 4.20 (joint CRs for FARIMA(0, $d$, 1), FARIMA(0, $d$, 1)).
Table 4.3 Empirical coverage rates (%) and median lengths for two-sided studentized 95% CIs for $d = 0.4, 0.25, 0.1$ from FARIMA($0, d, 0$) models with standard normal innovations using FDB and normal approximation methods. Note that density-studentized FDB intervals for $d$ here match those of non-studentized FDB intervals. (CP: coverage probability, Me(L): median length)

<table>
<thead>
<tr>
<th>$d$</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>CP</td>
<td>Me(L)</td>
<td>CP</td>
<td>Me(L)</td>
<td>CP</td>
<td>Me(L)</td>
</tr>
<tr>
<td>100</td>
<td>86.9</td>
<td>0.3063</td>
<td>85.6</td>
<td>0.3077</td>
<td>86.3</td>
<td>0.3059</td>
</tr>
<tr>
<td>B.PS 250</td>
<td>90.7</td>
<td>0.1964</td>
<td>91.6</td>
<td>0.1968</td>
<td>91.6</td>
<td>0.1978</td>
</tr>
<tr>
<td>500</td>
<td>92.2</td>
<td>0.1380</td>
<td>92.4</td>
<td>0.1394</td>
<td>93.1</td>
<td>0.1394</td>
</tr>
<tr>
<td>1000</td>
<td>93.7</td>
<td>0.0975</td>
<td>93.5</td>
<td>0.0978</td>
<td>93.5</td>
<td>0.0980</td>
</tr>
<tr>
<td>B.PA 500</td>
<td>88.1</td>
<td>0.3064</td>
<td>87.2</td>
<td>0.3068</td>
<td>88.4</td>
<td>0.3057</td>
</tr>
<tr>
<td>1000</td>
<td>91.3</td>
<td>0.1960</td>
<td>92.5</td>
<td>0.1971</td>
<td>92.5</td>
<td>0.1980</td>
</tr>
<tr>
<td>Nor.P 250</td>
<td>92.4</td>
<td>0.1383</td>
<td>92.9</td>
<td>0.1393</td>
<td>93.6</td>
<td>0.1397</td>
</tr>
<tr>
<td>1000</td>
<td>93.7</td>
<td>0.0975</td>
<td>94.0</td>
<td>0.0980</td>
<td>93.8</td>
<td>0.0981</td>
</tr>
<tr>
<td>Nor.P 500</td>
<td>91.6</td>
<td>0.2551</td>
<td>79.0</td>
<td>0.2565</td>
<td>78.1</td>
<td>0.2575</td>
</tr>
<tr>
<td>1000</td>
<td>81.3</td>
<td>0.1499</td>
<td>81.8</td>
<td>0.1512</td>
<td>81.6</td>
<td>0.1513</td>
</tr>
<tr>
<td>1000</td>
<td>81.9</td>
<td>0.1025</td>
<td>81.4</td>
<td>0.1033</td>
<td>82.9</td>
<td>0.1031</td>
</tr>
<tr>
<td>1000</td>
<td>83.2</td>
<td>0.0709</td>
<td>82.6</td>
<td>0.0712</td>
<td>82.2</td>
<td>0.0712</td>
</tr>
</tbody>
</table>
Table 4.4 Empirical coverage rates (%) and median lengths for one-sided (lower/upper) studentized 95% CIs for $d = 0.4, 0.25, 0.1$ from FARIMA$(0, d, 0)$ models with standard normal innovations using FDB and normal approximation methods. Note that density-studentized FDB intervals for $d$ here match those of non-studentized FDB intervals. (CP: coverage probability, Me(L): median length)

<table>
<thead>
<tr>
<th>$d$</th>
<th>0.40</th>
<th>Lower</th>
<th>0.25</th>
<th>0.10</th>
<th></th>
<th>0.40</th>
<th>Lower</th>
<th>0.25</th>
<th>0.10</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
</tr>
<tr>
<td>B.P</td>
<td>100</td>
<td>86.4 0.5112</td>
<td>85.3 0.3575</td>
<td>85.0 0.2044</td>
<td>94.9 0.2527</td>
<td>95.4 0.0942</td>
<td>96.3 -0.0593</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>90.7 0.4770</td>
<td>90.3 0.3237</td>
<td>90.6 0.1725</td>
<td>94.8 0.3119</td>
<td>96.4 0.1853</td>
<td>96.1 0.0058</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>92.6 0.4558</td>
<td>91.8 0.3042</td>
<td>92.5 0.1533</td>
<td>94.0 0.3400</td>
<td>95.3 0.1873</td>
<td>96.3 0.0367</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>93.5 0.3838</td>
<td>93.1 0.2387</td>
<td>92.5 0.1386</td>
<td>94.7 0.3583</td>
<td>95.8 0.2061</td>
<td>95.6 0.0566</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nor.P</td>
<td>100</td>
<td>87.3 0.4738</td>
<td>76.6 0.3204</td>
<td>75.3 0.1670</td>
<td>94.1 0.2633</td>
<td>94.2 0.1058</td>
<td>94.9 -0.0473</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>81.6 0.4500</td>
<td>80.1 0.2963</td>
<td>79.6 0.1456</td>
<td>92.0 0.3246</td>
<td>93.3 0.1720</td>
<td>93.5 0.0194</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>84.2 0.4366</td>
<td>82.1 0.2848</td>
<td>82.3 0.1335</td>
<td>90.2 0.3514</td>
<td>91.3 0.1990</td>
<td>92.8 0.0485</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>85.8 0.4276</td>
<td>83.3 0.2753</td>
<td>83.0 0.1253</td>
<td>89.7 0.3687</td>
<td>91.4 0.2154</td>
<td>91.0 0.0661</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B.D</td>
<td>100</td>
<td>90.2 0.5337</td>
<td>88.9 0.3780</td>
<td>89.2 0.2232</td>
<td>96.8 0.2345</td>
<td>97.1 0.0772</td>
<td>97.5 -0.0748</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>92.1 0.4828</td>
<td>92.0 0.3290</td>
<td>91.8 0.1771</td>
<td>96.1 0.3068</td>
<td>97.3 0.1523</td>
<td>96.8 0.0018</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>93.7 0.4586</td>
<td>92.7 0.3065</td>
<td>92.7 0.1554</td>
<td>94.9 0.3374</td>
<td>95.7 0.1854</td>
<td>96.7 0.0346</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>94.4 0.4421</td>
<td>93.3 0.2894</td>
<td>92.9 0.1395</td>
<td>95.2 0.3581</td>
<td>96.1 0.2058</td>
<td>96.1 0.0558</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nor.D</td>
<td>100</td>
<td>88.2 0.3202</td>
<td>87.4 0.3621</td>
<td>86.9 0.2087</td>
<td>98.0 0.2197</td>
<td>98.5 0.0616</td>
<td>98.4 -0.0198</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>90.4 0.4759</td>
<td>90.2 0.3221</td>
<td>90.2 0.1708</td>
<td>97.3 0.2995</td>
<td>98.2 0.1458</td>
<td>97.9 -0.0056</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>92.5 0.4543</td>
<td>91.1 0.3023</td>
<td>91.7 0.1516</td>
<td>96.0 0.3336</td>
<td>96.6 0.1817</td>
<td>97.6 0.0309</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>93.7 0.4399</td>
<td>92.6 0.2872</td>
<td>91.9 0.1375</td>
<td>96.0 0.3563</td>
<td>96.7 0.2035</td>
<td>96.8 0.0539</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.5 Empirical coverage rates (%) and median lengths for two-sided studentized 95% CIs for $d = 0.4, 0.25, 0.1$ from FARIMA$(0, d, 0)$ models with chi-square innovations using FDB and normal approximation methods. Note that density-studentized FDB intervals for $d$ here match those of non-studentized FDB intervals. (CP: coverage probability, Me(L): median length)

<table>
<thead>
<tr>
<th>$d$</th>
<th>0.40</th>
<th>Lower</th>
<th>0.25</th>
<th>0.10</th>
<th></th>
<th>0.40</th>
<th>Lower</th>
<th>0.25</th>
<th>0.10</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
</tr>
<tr>
<td>B.PS</td>
<td>100</td>
<td>87.5 0.3027</td>
<td>85.9 0.3021</td>
<td>86.9 0.3017</td>
<td>91.7 0.3286</td>
<td>91.4 0.3284</td>
<td>91.4 0.3277</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>92.0 0.1948</td>
<td>91.6 0.1948</td>
<td>90.9 0.1945</td>
<td>93.6 0.2017</td>
<td>93.3 0.2006</td>
<td>92.8 0.2014</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>92.5 0.1382</td>
<td>92.8 0.1374</td>
<td>93.4 0.1384</td>
<td>93.2 0.1406</td>
<td>93.7 0.1401</td>
<td>96.2 0.5357</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>93.8 0.0974</td>
<td>94.0 0.0974</td>
<td>94.6 0.0979</td>
<td>94.2 0.0982</td>
<td>94.1 0.0984</td>
<td>95.1 0.0985</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B.PA</td>
<td>100</td>
<td>88.6 0.3017</td>
<td>87.6 0.3016</td>
<td>88.2 0.2999</td>
<td>92.1 0.3513</td>
<td>92.4 0.3497</td>
<td>92.5 0.3495</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>92.2 0.1947</td>
<td>92.1 0.1949</td>
<td>92.0 0.1944</td>
<td>94.5 0.2080</td>
<td>93.9 0.2072</td>
<td>93.9 0.2078</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>92.4 0.1383</td>
<td>93.3 0.1376</td>
<td>94.0 0.1384</td>
<td>93.6 0.1428</td>
<td>94.1 0.1428</td>
<td>94.8 0.1430</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>93.8 0.0975</td>
<td>94.1 0.0977</td>
<td>94.7 0.0981</td>
<td>94.3 0.0994</td>
<td>94.6 0.0993</td>
<td>95.3 0.0995</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nor.P</td>
<td>100</td>
<td>80.5 0.2587</td>
<td>79.1 0.2586</td>
<td>80.2 0.2573</td>
<td>92.0 0.3581</td>
<td>92.0 0.3581</td>
<td>92.6 0.3581</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>82.0 0.1510</td>
<td>82.0 0.1520</td>
<td>81.7 0.1514</td>
<td>94.1 0.2101</td>
<td>93.8 0.2101</td>
<td>93.7 0.2101</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>82.1 0.1030</td>
<td>82.7 0.1029</td>
<td>83.6 0.1033</td>
<td>93.6 0.1438</td>
<td>94.2 0.1438</td>
<td>94.6 0.1438</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>83.2 0.0712</td>
<td>81.8 0.0713</td>
<td>83.5 0.0714</td>
<td>94.6 0.0997</td>
<td>94.5 0.0997</td>
<td>94.8 0.0997</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.6  Empirical coverage rate(%) and median lengths for one-sided (lower/upper) studentized 95% CIs for $d = 0.4, 0.25, 0.1$ from FARIMA$(0,d,0)$ models with chi-square innovations using FDB and normal approximation methods. Note that density-studentized FDB intervals for $d$ here match those of non-studentized FDB intervals. (CP: coverage probability, Me(L): median length)

<table>
<thead>
<tr>
<th>$d$</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>CP</td>
<td>Me(L)</td>
<td>CP</td>
<td>Me(L)</td>
<td>CP</td>
<td>Me(L)</td>
</tr>
<tr>
<td>B.P</td>
<td>100</td>
<td>86.7</td>
<td>0.5067</td>
<td>83.6</td>
<td>0.3497</td>
<td>85.2</td>
</tr>
<tr>
<td>250</td>
<td>91.2</td>
<td>0.4745</td>
<td>91.3</td>
<td>0.3224</td>
<td>89.9</td>
<td>0.1712</td>
</tr>
<tr>
<td>500</td>
<td>93.1</td>
<td>0.4559</td>
<td>92.6</td>
<td>0.3023</td>
<td>93.0</td>
<td>0.1535</td>
</tr>
<tr>
<td>1000</td>
<td>94.2</td>
<td>0.4345</td>
<td>93.7</td>
<td>0.2982</td>
<td>95.7</td>
<td>0.1585</td>
</tr>
<tr>
<td>Nor.P</td>
<td>100</td>
<td>77.9</td>
<td>0.4747</td>
<td>75.9</td>
<td>0.3171</td>
<td>76.5</td>
</tr>
<tr>
<td>250</td>
<td>81.3</td>
<td>0.4849</td>
<td>81.1</td>
<td>0.2975</td>
<td>79.6</td>
<td>0.1454</td>
</tr>
<tr>
<td>500</td>
<td>84.2</td>
<td>0.4373</td>
<td>82.6</td>
<td>0.2838</td>
<td>83.5</td>
<td>0.1351</td>
</tr>
<tr>
<td>1000</td>
<td>86.0</td>
<td>0.4273</td>
<td>84.1</td>
<td>0.2749</td>
<td>84.7</td>
<td>0.1253</td>
</tr>
<tr>
<td>B.D</td>
<td>100</td>
<td>90.2</td>
<td>0.5271</td>
<td>88.3</td>
<td>0.3703</td>
<td>89.2</td>
</tr>
<tr>
<td>250</td>
<td>92.7</td>
<td>0.4806</td>
<td>92.5</td>
<td>0.3281</td>
<td>91.2</td>
<td>0.1771</td>
</tr>
<tr>
<td>500</td>
<td>93.8</td>
<td>0.4582</td>
<td>92.7</td>
<td>0.3044</td>
<td>93.9</td>
<td>0.1551</td>
</tr>
<tr>
<td>1000</td>
<td>94.8</td>
<td>0.4417</td>
<td>94.0</td>
<td>0.2892</td>
<td>94.0</td>
<td>0.1391</td>
</tr>
<tr>
<td>Nor.D</td>
<td>100</td>
<td>89.0</td>
<td>0.5180</td>
<td>87.8</td>
<td>0.3624</td>
<td>88.0</td>
</tr>
<tr>
<td>250</td>
<td>91.5</td>
<td>0.4748</td>
<td>91.6</td>
<td>0.3227</td>
<td>90.2</td>
<td>0.1699</td>
</tr>
<tr>
<td>500</td>
<td>93.0</td>
<td>0.4547</td>
<td>91.9</td>
<td>0.3010</td>
<td>92.7</td>
<td>0.1523</td>
</tr>
<tr>
<td>1000</td>
<td>93.9</td>
<td>0.4399</td>
<td>92.9</td>
<td>0.2874</td>
<td>93.2</td>
<td>0.1371</td>
</tr>
</tbody>
</table>

Table 4.7  Empirical coverage rates (%) and median lengths for two-sided studentized 95% CIs for $d = 0.4, 0.25, 0.1$ from FARIMA$(0,d,0)$ models with student-t innovations using FDB and normal approximation methods. Note that density-studentized FDB intervals for $d$ here match those of non-studentized FDB intervals. (CP: coverage probability, Me(L): median length)

<table>
<thead>
<tr>
<th>$d$</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>CP</td>
<td>Me(L)</td>
<td>CP</td>
<td>Me(L)</td>
<td>CP</td>
<td>Me(L)</td>
</tr>
<tr>
<td>B.PS</td>
<td>100</td>
<td>87.2</td>
<td>0.2918</td>
<td>86.1</td>
<td>0.2957</td>
<td>86.7</td>
</tr>
<tr>
<td>250</td>
<td>91.3</td>
<td>0.1909</td>
<td>90.5</td>
<td>0.1909</td>
<td>91.5</td>
<td>0.1907</td>
</tr>
<tr>
<td>500</td>
<td>93.8</td>
<td>0.0960</td>
<td>94.1</td>
<td>0.0966</td>
<td>93.5</td>
<td>0.0967</td>
</tr>
<tr>
<td>1000</td>
<td>87.7</td>
<td>0.2981</td>
<td>86.9</td>
<td>0.2933</td>
<td>87.0</td>
<td>0.2928</td>
</tr>
<tr>
<td>B.PA</td>
<td>100</td>
<td>91.9</td>
<td>0.1913</td>
<td>90.9</td>
<td>0.1908</td>
<td>92.5</td>
</tr>
<tr>
<td>500</td>
<td>92.6</td>
<td>0.1354</td>
<td>93.2</td>
<td>0.1363</td>
<td>92.8</td>
<td>0.1356</td>
</tr>
<tr>
<td>1000</td>
<td>93.2</td>
<td>0.0960</td>
<td>94.6</td>
<td>0.0966</td>
<td>93.8</td>
<td>0.0969</td>
</tr>
<tr>
<td>Nor.P</td>
<td>100</td>
<td>81.6</td>
<td>0.2578</td>
<td>80.5</td>
<td>0.2611</td>
<td>81.8</td>
</tr>
<tr>
<td>250</td>
<td>83.0</td>
<td>0.1521</td>
<td>82.2</td>
<td>0.1523</td>
<td>83.8</td>
<td>0.1523</td>
</tr>
<tr>
<td>500</td>
<td>83.6</td>
<td>0.1033</td>
<td>83.5</td>
<td>0.1039</td>
<td>82.2</td>
<td>0.1038</td>
</tr>
<tr>
<td>1000</td>
<td>82.4</td>
<td>0.0714</td>
<td>84.3</td>
<td>0.0716</td>
<td>82.7</td>
<td>0.0719</td>
</tr>
</tbody>
</table>
Table 4.8 Empirical coverage rates (%) and median lengths for one-sided (lower/upper) studentized 95% CIs for \( d = 0.4, 0.25, 0.1 \) from FARIMA(0, \( d, 0 \)) models with student-t innovations using FDB and normal approximation methods. Note that density-studentized FDB intervals for \( d \) here match those of non-studentized FDB intervals. (CP: coverage probability, Me(L): median of length)

<table>
<thead>
<tr>
<th>( d )</th>
<th>Lower</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>Upper</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
</tr>
<tr>
<td>100</td>
<td>84.2 0.3955</td>
<td>83.5 0.3434</td>
<td>84.0 0.1926</td>
<td>95.4 0.2488</td>
<td>96.1 0.0956</td>
<td>96.3 -0.0557</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B.P</td>
<td>250</td>
<td>90.5 0.4751</td>
<td>90.0 0.3205</td>
<td>89.9 0.1681</td>
<td>95.3 0.3137</td>
<td>95.3 0.1604</td>
<td>96.8 0.0097</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>93.2 0.4557</td>
<td>91.5 0.3028</td>
<td>90.7 0.1502</td>
<td>94.4 0.3426</td>
<td>96.5 0.1886</td>
<td>96.3 0.0364</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>93.1 0.4396</td>
<td>93.2 0.2878</td>
<td>92.4 0.1369</td>
<td>94.2 0.3593</td>
<td>96.3 0.2074</td>
<td>95.9 0.0562</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nor.P</td>
<td>100</td>
<td>79.1 0.4752</td>
<td>77.7 0.3223</td>
<td>78.3 0.1719</td>
<td>94.3 0.2591</td>
<td>94.9 0.1053</td>
<td>94.9 -0.0451</td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>82.9 0.4526</td>
<td>82.0 0.2988</td>
<td>82.1 0.1474</td>
<td>92.6 0.3251</td>
<td>93.2 0.1726</td>
<td>95.0 0.0206</td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>85.1 0.4390</td>
<td>83.1 0.2858</td>
<td>81.6 0.1337</td>
<td>90.6 0.3529</td>
<td>92.7 0.1989</td>
<td>93.1 0.0470</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>85.9 0.4276</td>
<td>84.2 0.2756</td>
<td>83.1 0.1254</td>
<td>89.5 0.3676</td>
<td>92.2 0.2157</td>
<td>92.1 0.0651</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B.D</td>
<td>100</td>
<td>88.9 0.5155</td>
<td>86.9 0.3621</td>
<td>88.3 0.2133</td>
<td>97.5 0.2274</td>
<td>97.6 0.0766</td>
<td>97.6 -0.0749</td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>91.7 0.4798</td>
<td>91.2 0.3269</td>
<td>91.7 0.1742</td>
<td>96.6 0.3079</td>
<td>96.6 0.1547</td>
<td>97.8 0.0034</td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>93.6 0.4678</td>
<td>92.2 0.3048</td>
<td>91.7 0.1527</td>
<td>95.1 0.3402</td>
<td>97.1 0.1865</td>
<td>96.7 0.0344</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>93.6 0.4405</td>
<td>93.6 0.2866</td>
<td>92.8 0.1378</td>
<td>94.9 0.3585</td>
<td>96.6 0.2064</td>
<td>95.9 0.0557</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nor.D</td>
<td>100</td>
<td>89.6 0.5174</td>
<td>87.4 0.3644</td>
<td>88.9 0.2145</td>
<td>98.3 0.2169</td>
<td>98.6 0.0639</td>
<td>98.5 -0.0859</td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>91.2 0.4770</td>
<td>90.8 0.3237</td>
<td>91.2 0.1718</td>
<td>97.5 0.3006</td>
<td>97.7 0.1474</td>
<td>98.4 -0.0046</td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>93.3 0.4561</td>
<td>91.4 0.3026</td>
<td>91.2 0.1506</td>
<td>96.3 0.3354</td>
<td>97.9 0.1819</td>
<td>97.5 0.0299</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>93.3 0.4396</td>
<td>93.3 0.2874</td>
<td>92.4 0.1370</td>
<td>95.8 0.3559</td>
<td>97.2 0.2038</td>
<td>96.6 0.0534</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.9 Empirical coverage rates (%) and median lengths for two-sided studentized 95% CIs for \( d = 0.4, 0.25, 0.1 \) from FARIMA(1, \( d, 0 \)) models with \( \phi_1 = -0.3 \) and standard normal innovations using FDB and normal approximation methods. (CP: coverage probability, Me(L): median length)

<table>
<thead>
<tr>
<th>( d )</th>
<th>Lower</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>Upper</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
</tr>
<tr>
<td>100</td>
<td>73.8 0.3866</td>
<td>73.1 0.3837</td>
<td>72.4 0.3815</td>
<td>84.0 0.4367</td>
<td>82.1 0.4375</td>
<td>82.6 0.4339</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B.PS</td>
<td>250</td>
<td>85.6 0.2434</td>
<td>84.2 0.2432</td>
<td>84.9 0.2488</td>
<td>89.3 0.2569</td>
<td>87.9 0.2560</td>
<td>88.5 0.2571</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>88.8 0.1702</td>
<td>88.1 0.1716</td>
<td>88.2 0.1716</td>
<td>90.9 0.1752</td>
<td>90.0 0.1761</td>
<td>90.2 0.1757</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>90.7 0.1199</td>
<td>90.0 0.1206</td>
<td>91.1 0.1205</td>
<td>91.8 0.1216</td>
<td>90.9 0.1219</td>
<td>91.8 0.1219</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B.PA</td>
<td>100</td>
<td>76.4 0.3860</td>
<td>75.7 0.3849</td>
<td>75.2 0.3817</td>
<td>86.2 0.4831</td>
<td>84.3 0.4844</td>
<td>84.7 0.4813</td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>87.6 0.2436</td>
<td>86.6 0.2432</td>
<td>87.2 0.2487</td>
<td>91.0 0.2697</td>
<td>89.7 0.2687</td>
<td>90.4 0.2693</td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>89.4 0.1705</td>
<td>89.3 0.1718</td>
<td>89.4 0.1714</td>
<td>91.5 0.1799</td>
<td>90.9 0.1807</td>
<td>91.8 0.1804</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>91.0 0.1198</td>
<td>90.8 0.1207</td>
<td>92.0 0.1205</td>
<td>92.1 0.1235</td>
<td>91.7 0.1236</td>
<td>92.3 0.1234</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nor.P</td>
<td>100</td>
<td>68.4 0.3405</td>
<td>67.7 0.3498</td>
<td>66.9 0.3391</td>
<td>85.0 0.4838</td>
<td>83.5 0.4825</td>
<td>84.4 0.4816</td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>75.2 0.1906</td>
<td>73.5 0.1905</td>
<td>73.9 0.1900</td>
<td>89.2 0.2677</td>
<td>87.9 0.2677</td>
<td>88.4 0.2675</td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>76.9 0.1280</td>
<td>76.2 0.1281</td>
<td>75.8 0.1284</td>
<td>90.8 0.1796</td>
<td>89.9 0.1798</td>
<td>90.6 0.1798</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>78.6 0.0875</td>
<td>78.0 0.0881</td>
<td>78.6 0.0882</td>
<td>91.7 0.1232</td>
<td>90.8 0.1232</td>
<td>91.9 0.1233</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.10 Empirical coverage rates (%) and median lengths for one-sided (lower/upper) studentized 95% CIs for $d = 0.4, 0.25, 0.1$ from FARIMA($1, d, 0$) models with $\phi_1 = -0.3$ and standard normal innovations using FDB and normal approximation methods. (CP: coverage probability, Me(L): median length)

<table>
<thead>
<tr>
<th>$d$</th>
<th>Lower CP</th>
<th>Lower Me(L)</th>
<th>Upper CP</th>
<th>Upper Me(L)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.40</td>
<td>0.25</td>
<td>0.10</td>
<td>0.40</td>
</tr>
<tr>
<td>100</td>
<td>72.1</td>
<td>0.4910</td>
<td>70.8</td>
<td>0.3351</td>
</tr>
<tr>
<td>250</td>
<td>85.1</td>
<td>0.4818</td>
<td>82.8</td>
<td>0.3277</td>
</tr>
<tr>
<td>500</td>
<td>89.3</td>
<td>0.4634</td>
<td>86.8</td>
<td>0.3084</td>
</tr>
<tr>
<td>1000</td>
<td>91.3</td>
<td>0.4466</td>
<td>89.3</td>
<td>0.2949</td>
</tr>
</tbody>
</table>

Table 4.11 Empirical coverage rates (%) and median lengths for two-sided studentized 95% CIs for $\phi_1 = -0.3$ from FARIMA($1, d, 0$) models with $d = 0.4, 0.25, 0.1$ and standard normal innovations using FDB and normal approximation methods. (CP: coverage probability, Me(L): median length)

<table>
<thead>
<tr>
<th>$d$</th>
<th>Lower CP</th>
<th>Lower Me(L)</th>
<th>Upper CP</th>
<th>Upper Me(L)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.40</td>
<td>0.25</td>
<td>0.10</td>
<td>0.40</td>
</tr>
<tr>
<td>100</td>
<td>80.9</td>
<td>0.450</td>
<td>81.5</td>
<td>0.4512</td>
</tr>
<tr>
<td>250</td>
<td>88.6</td>
<td>0.2905</td>
<td>88.5</td>
<td>0.2905</td>
</tr>
<tr>
<td>500</td>
<td>91.4</td>
<td>0.2075</td>
<td>90.7</td>
<td>0.2088</td>
</tr>
<tr>
<td>1000</td>
<td>92.7</td>
<td>0.1478</td>
<td>92.9</td>
<td>0.1476</td>
</tr>
<tr>
<td>100</td>
<td>76.9</td>
<td>0.4447</td>
<td>76.7</td>
<td>0.4498</td>
</tr>
<tr>
<td>250</td>
<td>87.3</td>
<td>0.2913</td>
<td>87.7</td>
<td>0.2916</td>
</tr>
<tr>
<td>500</td>
<td>90.9</td>
<td>0.2082</td>
<td>90.2</td>
<td>0.2093</td>
</tr>
<tr>
<td>1000</td>
<td>92.5</td>
<td>0.1481</td>
<td>92.8</td>
<td>0.1478</td>
</tr>
<tr>
<td>100</td>
<td>75.0</td>
<td>0.3961</td>
<td>75.2</td>
<td>0.3972</td>
</tr>
<tr>
<td>250</td>
<td>78.7</td>
<td>0.2272</td>
<td>77.4</td>
<td>0.2266</td>
</tr>
<tr>
<td>500</td>
<td>80.0</td>
<td>0.1547</td>
<td>79.2</td>
<td>0.1551</td>
</tr>
<tr>
<td>1000</td>
<td>80.2</td>
<td>0.1072</td>
<td>80.0</td>
<td>0.1071</td>
</tr>
</tbody>
</table>
Table 4.12 Empirical coverage rates (%) and median lengths for one-sided (lower/upper) studentized 95% CIs for $\phi_1 = -0.3$ from FARIMA(1, $d$, 0) models with $d = 0.4, 0.25, 0.1$ and standard normal innovations using FDB and normal approximation methods. (CP: coverage probability, Me(L): median length)

<table>
<thead>
<tr>
<th>$d$</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td></td>
<td></td>
<td>Upper</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>CP</td>
<td>Me(L)</td>
<td>CP</td>
<td>Me(L)</td>
<td>CP</td>
<td>Me(L)</td>
<td>CP</td>
<td>Me(L)</td>
<td>CP</td>
<td>Me(L)</td>
<td>CP</td>
</tr>
<tr>
<td>100</td>
<td>95.6</td>
<td>-0.0194</td>
<td>96.3</td>
<td>-0.0123</td>
<td>95.6</td>
<td>-0.0199</td>
<td>73.8</td>
<td>-0.3996</td>
<td>72.8</td>
<td>-0.3948</td>
<td>72.9</td>
<td>-0.3970</td>
</tr>
<tr>
<td></td>
<td>B.P</td>
<td>250</td>
<td>95.6</td>
<td>-0.1437</td>
<td>95.3</td>
<td>-0.1425</td>
<td>96.4</td>
<td>-0.1465</td>
<td>85.2</td>
<td>-0.3907</td>
<td>84.9</td>
<td>-0.3912</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>95.7</td>
<td>-0.1969</td>
<td>95.7</td>
<td>-0.1948</td>
<td>96.0</td>
<td>-0.1936</td>
<td>89.2</td>
<td>-0.3725</td>
<td>88.9</td>
<td>-0.3705</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>95.0</td>
<td>-0.2303</td>
<td>95.6</td>
<td>-0.2301</td>
<td>95.3</td>
<td>-0.2280</td>
<td>91.3</td>
<td>-0.3545</td>
<td>90.9</td>
<td>-0.3542</td>
</tr>
<tr>
<td>100</td>
<td>94.5</td>
<td>-0.0533</td>
<td>95.2</td>
<td>-0.0516</td>
<td>94.3</td>
<td>-0.0529</td>
<td>72.2</td>
<td>-0.3864</td>
<td>71.4</td>
<td>-0.3814</td>
<td>71.5</td>
<td>-0.3856</td>
</tr>
<tr>
<td></td>
<td>Nor.D</td>
<td>250</td>
<td>92.3</td>
<td>-0.1733</td>
<td>91.9</td>
<td>-0.1744</td>
<td>92.8</td>
<td>-0.1759</td>
<td>77.9</td>
<td>-0.3648</td>
<td>77.5</td>
<td>-0.3647</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>90.4</td>
<td>-0.2214</td>
<td>91.0</td>
<td>-0.2185</td>
<td>91.3</td>
<td>-0.2184</td>
<td>81.3</td>
<td>-0.3512</td>
<td>80.3</td>
<td>-0.3488</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>89.5</td>
<td>-0.2477</td>
<td>89.7</td>
<td>-0.2479</td>
<td>90.3</td>
<td>-0.2453</td>
<td>82.7</td>
<td>-0.3376</td>
<td>82.2</td>
<td>-0.3377</td>
</tr>
<tr>
<td>100</td>
<td>96.7</td>
<td>-0.0063</td>
<td>97.3</td>
<td>-0.0008</td>
<td>96.4</td>
<td>-0.0061</td>
<td>75.3</td>
<td>-0.4088</td>
<td>74.3</td>
<td>-0.4053</td>
<td>74.7</td>
<td>-0.4073</td>
</tr>
<tr>
<td></td>
<td>B.D</td>
<td>250</td>
<td>95.9</td>
<td>-0.1404</td>
<td>95.8</td>
<td>-0.1406</td>
<td>96.7</td>
<td>-0.1418</td>
<td>85.8</td>
<td>-0.3930</td>
<td>85.7</td>
<td>-0.3944</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>95.8</td>
<td>-0.1957</td>
<td>95.9</td>
<td>-0.1936</td>
<td>96.0</td>
<td>-0.1931</td>
<td>89.5</td>
<td>-0.3739</td>
<td>89.0</td>
<td>-0.3718</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>95.0</td>
<td>-0.2296</td>
<td>95.7</td>
<td>-0.2298</td>
<td>95.5</td>
<td>-0.2272</td>
<td>91.3</td>
<td>-0.3547</td>
<td>91.2</td>
<td>-0.3542</td>
</tr>
<tr>
<td>100</td>
<td>97.9</td>
<td>0.0049</td>
<td>98.1</td>
<td>0.0079</td>
<td>97.2</td>
<td>0.0040</td>
<td>83.2</td>
<td>-0.4445</td>
<td>83.7</td>
<td>-0.4424</td>
<td>83.3</td>
<td>-0.4451</td>
</tr>
<tr>
<td></td>
<td>Nor.D</td>
<td>250</td>
<td>96.6</td>
<td>-0.1377</td>
<td>96.1</td>
<td>-0.1373</td>
<td>96.7</td>
<td>-0.1391</td>
<td>88.2</td>
<td>-0.4012</td>
<td>88.4</td>
<td>-0.4008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>96.1</td>
<td>-0.1953</td>
<td>96.1</td>
<td>-0.1931</td>
<td>96.3</td>
<td>-0.1930</td>
<td>90.7</td>
<td>-0.3766</td>
<td>90.2</td>
<td>-0.3747</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>95.3</td>
<td>-0.2293</td>
<td>95.8</td>
<td>-0.2294</td>
<td>95.7</td>
<td>-0.2273</td>
<td>92.0</td>
<td>-0.3557</td>
<td>91.6</td>
<td>-0.3558</td>
</tr>
</tbody>
</table>

Table 4.13 Empirical coverage rates (%) and median lengths of one-sided (lower/upper) and two-sided 95% CIs for $d = 0.4, 0.25, 0.1$ and for $\phi_1 = -0.3$ from FARIMA(1, $d$, 0) models with standard normal innovations using non-studentized FDB methods. (CP: coverage probability, Me(L): median length)

<table>
<thead>
<tr>
<th>$d$</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate of AR</td>
<td></td>
<td></td>
<td></td>
<td>Estimate of AR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.14  Empirical coverage rates (%) and median volumes of studentized 95% joint CRs for $(d, \phi_1 = -0.3)$ for FARIMA$(1,d,0)$ processes with $d = 0.4, 0.25, 0.1$ and standard normal innovations using FDB and normal approximation methods. (CP: coverage probability, Me(V): median of volume)

<table>
<thead>
<tr>
<th>d</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>CP</td>
<td>Me(V)</td>
<td>CP</td>
<td>Me(V)</td>
<td>CP</td>
<td>Me(V)</td>
</tr>
<tr>
<td>B.P</td>
<td>100</td>
<td>79.8</td>
<td>0.3287</td>
<td>79.3</td>
<td>0.3317</td>
<td>78.3</td>
</tr>
<tr>
<td>250</td>
<td>89.1</td>
<td>0.1437</td>
<td>88.2</td>
<td>0.1425</td>
<td>88.1</td>
<td>0.1432</td>
</tr>
<tr>
<td>500</td>
<td>91.8</td>
<td>0.0732</td>
<td>90.9</td>
<td>0.0739</td>
<td>91.2</td>
<td>0.0739</td>
</tr>
<tr>
<td>1000</td>
<td>93.6</td>
<td>0.0369</td>
<td>93.1</td>
<td>0.0372</td>
<td>93.5</td>
<td>0.0371</td>
</tr>
<tr>
<td>Nor.P</td>
<td>250</td>
<td>74.8</td>
<td>0.0868</td>
<td>73.5</td>
<td>0.0866</td>
<td>73.2</td>
</tr>
<tr>
<td>500</td>
<td>75.3</td>
<td>0.0406</td>
<td>74.2</td>
<td>0.0408</td>
<td>74.5</td>
<td>0.0410</td>
</tr>
<tr>
<td>1000</td>
<td>76.3</td>
<td>0.0195</td>
<td>76.0</td>
<td>0.0196</td>
<td>77.0</td>
<td>0.0197</td>
</tr>
</tbody>
</table>

Table 4.15  Empirical coverage rates (%) and median lengths for two-sided studentized 95% CIs for $d = 0.4, 0.25, 0.1$ from FARIMA$(0,d,1)$ models with $\psi_1 = 0.4$ and standard normal innovations using FDB and normal approximation methods. (CP: coverage probability, Me(L): median length)

<table>
<thead>
<tr>
<th>d</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>CP</td>
<td>Me(L)</td>
<td>CP</td>
<td>Me(L)</td>
<td>CP</td>
<td>Me(L)</td>
</tr>
<tr>
<td>B.P</td>
<td>100</td>
<td>78.0</td>
<td>0.3526</td>
<td>77.7</td>
<td>0.3523</td>
<td>77.5</td>
</tr>
<tr>
<td>250</td>
<td>86.2</td>
<td>0.2291</td>
<td>87.5</td>
<td>0.2300</td>
<td>86.3</td>
<td>0.2310</td>
</tr>
<tr>
<td>500</td>
<td>90.3</td>
<td>0.1634</td>
<td>88.7</td>
<td>0.1641</td>
<td>88.8</td>
<td>0.1636</td>
</tr>
<tr>
<td>1000</td>
<td>90.5</td>
<td>0.1149</td>
<td>91.4</td>
<td>0.1155</td>
<td>92.3</td>
<td>0.1156</td>
</tr>
<tr>
<td>B.PA</td>
<td>250</td>
<td>80.8</td>
<td>0.3503</td>
<td>80.6</td>
<td>0.3502</td>
<td>80.2</td>
</tr>
<tr>
<td>500</td>
<td>91.1</td>
<td>0.1634</td>
<td>90.0</td>
<td>0.1638</td>
<td>90.3</td>
<td>0.1635</td>
</tr>
<tr>
<td>1000</td>
<td>90.9</td>
<td>0.1151</td>
<td>91.4</td>
<td>0.1155</td>
<td>92.3</td>
<td>0.1156</td>
</tr>
<tr>
<td>Nor.P</td>
<td>250</td>
<td>71.7</td>
<td>0.3045</td>
<td>71.0</td>
<td>0.3031</td>
<td>70.9</td>
</tr>
<tr>
<td>500</td>
<td>78.4</td>
<td>0.1222</td>
<td>77.8</td>
<td>0.1227</td>
<td>77.6</td>
<td>0.1222</td>
</tr>
<tr>
<td>1000</td>
<td>78.0</td>
<td>0.0839</td>
<td>79.8</td>
<td>0.0844</td>
<td>79.8</td>
<td>0.0846</td>
</tr>
</tbody>
</table>
Table 4.16 Empirical coverage rates (%) and median lengths for one-sided (lower/upper) studentized 95% CIs for $d = 0.4, 0.25, 0.1$ from FARIMA($0, d, 1$) models with $\psi_1 = 0.4$ and standard normal innovations using FDB and normal approximation methods. (CP: coverage probability, median of length)

<table>
<thead>
<tr>
<th>$d$</th>
<th>CP</th>
<th>Me(L)</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.40</td>
<td>78.2</td>
<td>0.5061</td>
<td>76.9</td>
<td>0.4363</td>
<td>76.7</td>
<td>0.1943</td>
<td>94.0</td>
<td>0.2058</td>
<td>95.7</td>
<td>0.0471</td>
<td>95.6</td>
<td>-0.1024</td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>78.2</td>
<td>0.5061</td>
<td>76.9</td>
<td>0.4363</td>
<td>76.7</td>
<td>0.1943</td>
<td>94.0</td>
<td>0.2058</td>
<td>95.7</td>
<td>0.0471</td>
<td>95.6</td>
<td>-0.1024</td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>78.2</td>
<td>0.5061</td>
<td>76.9</td>
<td>0.4363</td>
<td>76.7</td>
<td>0.1943</td>
<td>94.0</td>
<td>0.2058</td>
<td>95.7</td>
<td>0.0471</td>
<td>95.6</td>
<td>-0.1024</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.17 Empirical coverage rates (%) and median lengths for two-sided studentized 95% CIs for $\psi_1 = 0.4$ from FARIMA($0, d, 1$) models with $d = 0.4, 0.25, 0.1$ and standard normal innovations using FDB and normal approximation methods. (CP: coverage probability, Me(L): median of length)

<table>
<thead>
<tr>
<th>$d$</th>
<th>CP</th>
<th>Me(L)</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.40</td>
<td>89.0</td>
<td>0.4470</td>
<td>87.6</td>
<td>0.4452</td>
<td>87.5</td>
<td>0.4412</td>
<td>92.4</td>
<td>0.4772</td>
<td>91.3</td>
<td>0.4764</td>
<td>91.2</td>
<td>0.4751</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>89.0</td>
<td>0.4470</td>
<td>87.6</td>
<td>0.4452</td>
<td>87.5</td>
<td>0.4412</td>
<td>92.4</td>
<td>0.4772</td>
<td>91.3</td>
<td>0.4764</td>
<td>91.2</td>
<td>0.4751</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>89.0</td>
<td>0.4470</td>
<td>87.6</td>
<td>0.4452</td>
<td>87.5</td>
<td>0.4412</td>
<td>92.4</td>
<td>0.4772</td>
<td>91.3</td>
<td>0.4764</td>
<td>91.2</td>
<td>0.4751</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$d$</th>
<th>CP</th>
<th>Me(L)</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.40</td>
<td>90.9</td>
<td>0.2718</td>
<td>92.0</td>
<td>0.2713</td>
<td>90.6</td>
<td>0.2715</td>
<td>91.9</td>
<td>0.2794</td>
<td>92.5</td>
<td>0.2795</td>
<td>92.1</td>
<td>0.2779</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>90.9</td>
<td>0.2718</td>
<td>92.0</td>
<td>0.2713</td>
<td>90.6</td>
<td>0.2715</td>
<td>91.9</td>
<td>0.2794</td>
<td>92.5</td>
<td>0.2795</td>
<td>92.1</td>
<td>0.2779</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>90.9</td>
<td>0.2718</td>
<td>92.0</td>
<td>0.2713</td>
<td>90.6</td>
<td>0.2715</td>
<td>91.9</td>
<td>0.2794</td>
<td>92.5</td>
<td>0.2795</td>
<td>92.1</td>
<td>0.2779</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$d$</th>
<th>CP</th>
<th>Me(L)</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.40</td>
<td>89.0</td>
<td>0.4394</td>
<td>92.3</td>
<td>0.4388</td>
<td>92.6</td>
<td>0.4372</td>
<td>95.7</td>
<td>0.5036</td>
<td>95.0</td>
<td>0.5018</td>
<td>95.3</td>
<td>0.4999</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>89.0</td>
<td>0.4394</td>
<td>92.3</td>
<td>0.4388</td>
<td>92.6</td>
<td>0.4372</td>
<td>95.7</td>
<td>0.5036</td>
<td>95.0</td>
<td>0.5018</td>
<td>95.3</td>
<td>0.4999</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>89.0</td>
<td>0.4394</td>
<td>92.3</td>
<td>0.4388</td>
<td>92.6</td>
<td>0.4372</td>
<td>95.7</td>
<td>0.5036</td>
<td>95.0</td>
<td>0.5018</td>
<td>95.3</td>
<td>0.4999</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$d$</th>
<th>CP</th>
<th>Me(L)</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.40</td>
<td>79.1</td>
<td>0.1384</td>
<td>79.7</td>
<td>0.1380</td>
<td>79.3</td>
<td>0.1378</td>
<td>92.5</td>
<td>0.1944</td>
<td>92.0</td>
<td>0.1935</td>
<td>92.1</td>
<td>0.1931</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>79.1</td>
<td>0.1384</td>
<td>79.7</td>
<td>0.1380</td>
<td>79.3</td>
<td>0.1378</td>
<td>92.5</td>
<td>0.1944</td>
<td>92.0</td>
<td>0.1935</td>
<td>92.1</td>
<td>0.1931</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>79.1</td>
<td>0.1384</td>
<td>79.7</td>
<td>0.1380</td>
<td>79.3</td>
<td>0.1378</td>
<td>92.5</td>
<td>0.1944</td>
<td>92.0</td>
<td>0.1935</td>
<td>92.1</td>
<td>0.1931</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 4.18 Empirical coverage rates (%) and median lengths for one-sided (lower/upper) studentized 95% CIs for $\psi_1 = 0.4$ from FARIMA(0, $d$, 1) models with $d = 0.4, 0.25, 0.1$ and standard normal innovations using FDB and normal approximation methods. (CP: coverage probability, Me(L): median length)

<table>
<thead>
<tr>
<th>$d$</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
</tr>
<tr>
<td>B.P</td>
<td>100</td>
<td>96.3</td>
<td>0.5991</td>
<td>97.0</td>
<td>0.6087</td>
<td>97.8</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>95.8</td>
<td>0.5210</td>
<td>96.5</td>
<td>0.5237</td>
<td>96.1</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>95.7</td>
<td>0.4825</td>
<td>95.8</td>
<td>0.4868</td>
<td>96.5</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>93.9</td>
<td>0.4592</td>
<td>95.6</td>
<td>0.4611</td>
<td>95.6</td>
</tr>
<tr>
<td>Nor.P</td>
<td>100</td>
<td>94.1</td>
<td>0.5795</td>
<td>95.3</td>
<td>0.5868</td>
<td>95.4</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>92.1</td>
<td>0.5015</td>
<td>93.3</td>
<td>0.5042</td>
<td>93.4</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>91.4</td>
<td>0.4635</td>
<td>91.7</td>
<td>0.4697</td>
<td>92.2</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>88.9</td>
<td>0.4456</td>
<td>90.9</td>
<td>0.4475</td>
<td>90.9</td>
</tr>
<tr>
<td>B.D</td>
<td>100</td>
<td>97.2</td>
<td>0.6161</td>
<td>97.8</td>
<td>0.6262</td>
<td>98.2</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>96.3</td>
<td>0.5251</td>
<td>96.9</td>
<td>0.5289</td>
<td>96.8</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>96.0</td>
<td>0.4846</td>
<td>96.0</td>
<td>0.4886</td>
<td>96.7</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>94.2</td>
<td>0.4599</td>
<td>96.0</td>
<td>0.4646</td>
<td>95.9</td>
</tr>
<tr>
<td>Nor.D</td>
<td>100</td>
<td>97.9</td>
<td>0.6313</td>
<td>98.5</td>
<td>0.6395</td>
<td>98.8</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>97.3</td>
<td>0.5340</td>
<td>97.9</td>
<td>0.5378</td>
<td>97.6</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>96.7</td>
<td>0.4890</td>
<td>96.9</td>
<td>0.4932</td>
<td>97.2</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>94.9</td>
<td>0.4624</td>
<td>96.5</td>
<td>0.4643</td>
<td>96.4</td>
</tr>
</tbody>
</table>

### Table 4.19 Empirical coverage rates (%) and median lengths for one-sided (lower/upper) and two-sided 95% CIs for $d = 0.4, 0.25, 0.1$ and for $\psi_1 = 0.4$ from FARIMA(0, $d$, 1) models with standard normal innovations using non-studentized FDB methods. (CP: coverage probability, Me(L): median length)

<table>
<thead>
<tr>
<th>$d$</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
<th>0.40</th>
<th>0.25</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
<td>CP Me(L)</td>
</tr>
<tr>
<td>B.NS</td>
<td>100</td>
<td>89.3</td>
<td>0.4816</td>
<td>88.8</td>
<td>0.4797</td>
<td>88.3</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>91.9</td>
<td>0.2714</td>
<td>91.8</td>
<td>0.2715</td>
<td>90.8</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>93.4</td>
<td>0.1835</td>
<td>92.2</td>
<td>0.1832</td>
<td>92.1</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>93.4</td>
<td>0.1262</td>
<td>93.3</td>
<td>0.1258</td>
<td>94.2</td>
</tr>
<tr>
<td>B.NA</td>
<td>100</td>
<td>89.8</td>
<td>0.4790</td>
<td>89.6</td>
<td>0.4776</td>
<td>88.5</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>92.8</td>
<td>0.2709</td>
<td>92.6</td>
<td>0.2711</td>
<td>92.1</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>94.4</td>
<td>0.1835</td>
<td>93.2</td>
<td>0.1833</td>
<td>92.8</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>93.8</td>
<td>0.1263</td>
<td>93.8</td>
<td>0.1269</td>
<td>94.5</td>
</tr>
<tr>
<td>Lower</td>
<td>100</td>
<td>84.7</td>
<td>0.5333</td>
<td>83.5</td>
<td>0.3813</td>
<td>83.0</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>89.4</td>
<td>0.4967</td>
<td>89.6</td>
<td>0.3420</td>
<td>88.1</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>91.8</td>
<td>0.4889</td>
<td>90.8</td>
<td>0.3163</td>
<td>90.0</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>92.7</td>
<td>0.4501</td>
<td>91.9</td>
<td>0.2984</td>
<td>91.8</td>
</tr>
<tr>
<td>Upper</td>
<td>100</td>
<td>99.2</td>
<td>0.1126</td>
<td>99.4</td>
<td>0.1147</td>
<td>99.5</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>97.4</td>
<td>0.2687</td>
<td>98.4</td>
<td>0.2692</td>
<td>98.4</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>97.1</td>
<td>0.3147</td>
<td>97.5</td>
<td>0.1629</td>
<td>97.8</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>95.5</td>
<td>0.3440</td>
<td>97.0</td>
<td>0.1926</td>
<td>97.4</td>
</tr>
</tbody>
</table>
Table 4.20  Empirical coverage rates (%) and median volumes of studentized 95\% joint CRs for \((d, \psi_1 = 0.4)\) for FARIMA\((0, d, 1)\) processes with \(d = 0.4, 0.25, 0.1\) and standard normal innovations using FDB and normal approximation methods. (CP: coverage probability, Me(V): median of volume)

<table>
<thead>
<tr>
<th>(d)</th>
<th>(0.40)</th>
<th>(0.25)</th>
<th>(0.10)</th>
<th>(0.40)</th>
<th>(0.25)</th>
<th>(0.10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n)</td>
<td>CP</td>
<td>Me(V)</td>
<td>CP</td>
<td>Me(V)</td>
<td>CP</td>
<td>Me(V)</td>
</tr>
<tr>
<td>100</td>
<td>B.P</td>
<td>84.6</td>
<td>0.3574</td>
<td>83.7</td>
<td>0.3572</td>
<td>82.8</td>
</tr>
<tr>
<td>250</td>
<td></td>
<td>89.4</td>
<td>0.1360</td>
<td>89.4</td>
<td>0.1367</td>
<td>89.1</td>
</tr>
<tr>
<td>500</td>
<td></td>
<td>92.7</td>
<td>0.0685</td>
<td>91.5</td>
<td>0.0682</td>
<td>90.9</td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td>92.8</td>
<td>0.0341</td>
<td>92.5</td>
<td>0.0340</td>
<td>93.5</td>
</tr>
<tr>
<td>100</td>
<td>Nor.P</td>
<td>67.4</td>
<td>0.1989</td>
<td>66.1</td>
<td>0.1978</td>
<td>65.6</td>
</tr>
<tr>
<td>250</td>
<td></td>
<td>72.2</td>
<td>0.0746</td>
<td>73.8</td>
<td>0.0747</td>
<td>72.1</td>
</tr>
<tr>
<td>500</td>
<td></td>
<td>75.6</td>
<td>0.0362</td>
<td>74.4</td>
<td>0.0362</td>
<td>74.4</td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td>74.7</td>
<td>0.0176</td>
<td>76.5</td>
<td>0.0176</td>
<td>76.7</td>
</tr>
</tbody>
</table>
CHAPTER 5. GENERAL CONCLUSIONS

5.1 General Discussion

We have introduced alternative version of blockwise empirical likelihood for stationary, weakly
dependent time processes in chapter 2, and have developed block bootstrap estimation of the
sampling distribution of the sample mean and a frequency domain bootstrap (FDB) estimation
of the sampling distribution of Whittle estimators for a large class of linear time series exhibiting
long-memory in chapters 3 and 4.

Since the proposed blockwise empirical likelihood method (PBEL) does not require block
length selection in comparision to the standard and tapered blockwise empirical likelihood
methods, it relatively provides stable, robust results. There is some continuity for distribu-
tional estimation in the validity of block bootstrap under short- and long-range dependence in
chapter 3. For variance estimation, the relative efficiencies of moving and non-overlapping block
bootstrap estimators under long-range dependence are drastically different than the weakly de-
pendent time processes, even though both bootstraps may have the same large-sample variance
when the underlying dependence is strong enough. Under strong dependence, we conjecture
that these blocks might also increase as the underlying dependence decreases, which is consistent
with our simulation evidence. For time processes which exhibit forms of strong or long-range
dependence (LRD), the convergence of Whittle estimators to their normal limits can be slow
and sampling distributions in moderate sample sizes may be more asymmetric than normal
distributions in chapter 4. A frequency domain bootstrap method for Whittle estimators have
been proposed to calibrate confidence intervals to improve normal approximation.
5.2 Recommendations for Future Research

We studied the progressive block empirical likelihood (PBEL) for time series, which involves data blocking scheme where each block size increases in length by an arithmetic progression $2^k$, $k = 1, 2, \ldots$. Theoretically, $1k$, $2k$, $3k$ and so forth are valid for this same setting. In addition, we can study more with a variety of data-blocking scheme such as $4, 2, 8, 6, \ldots$. In chapter 2, we studied the overlapping progression blocking scheme is invalid for a progressive block empirical likelihood. Since we already know the overlapping blocking methods can improve the performance of the PBEL in contrast to the nonoverlapping blocking mechanism, this can be a future study for PBEL.

Dalhaus and Janas (1996) provided the frequency domain bootstrap under short-range dependence with nonparametric spectral density estimates (e.g. kernel method). Nonparametric spectral density estimators under long-range dependence can not easily be used because the spectral density function has a pole at the origin. In chapter 4, the frequency domain bootstrap for Whittle estimation was considered. This method is not widely applicable in contrast to the FDB for weakly dependent time series. Nonparametric spectral density estimation under long-range dependence are recommended for future research. With the nonparametric spectral density estimators, we can apply an standard FDB and an autoregressive-aided frequency domain bootstrap under long-range dependence. In addition, we can extend this idea to estimate long-memory parameter using bootstrap methods.

References


