Advances in the Estimation of Fractionally Integrated Models

Kanchana Nadarajah

MPhil, Monash University, Australia

BSc (Hons), University of Colombo, Sri Lanka

A thesis submitted for the degree of Doctor of Philosophy at

Monash University in May 2019

The Department of Econometrics and Business Statistics

Monash Business School

Copyright notice

Under the Copyright Act 1968, this thesis must be used only under the normal conditions of scholarly fair dealing for the purposes of research, criticism or review. In particular no results or conclusions should be extracted from it, nor should it be copied or closely paraphrased in whole or in part without the written consent of the author. Proper written acknowledgement should be made for any assistance obtained from this thesis.

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

Contents

List of Figures

List of Tables

Conventions and notation

The following notation and conventions are used throughout the thesis. All limits are considered as the sample size $n \to \infty$, unless specified otherwise.

Important conventions

$H(.)$	refers to a function
$H(x)$	refers to the value of the function $H(.)$ at x
$H^{(k)}$	k^{th} derivative of $H(.)$
$\{h_t\}$	refers to a series of values indexed by the integer t
Theorem 4.2	refers to Theorem 2 of Chapter 4

Abbreviations

AR	autoregressive
ARMA	autoregressive moving average
ARIMA	autoregressive integrated moving average
ARFIMA	fractionally integrated autoregressive moving average
BLUE	best linear unbiased estimator
CLT	central limit theorem
CSS	conditional sum of squares
DFT	discrete Fourier transform
DGP	data generating process
DWH	discrete Whittle
EWH	exact Whittle
FML	frequency domain maximum likelihood
i.i.d	independent and identically distributed
LPR	log-periodogram regression

LS	least squares
LSE	least squares estimator
MA	moving average
MisM	mis-specified model
MSE	mean squared error
OLS	ordinary least squares
OLSE	ordinary least squares estimator
RMSE	root mean squared error
TDGP	true data generating process
TML	time domain maximum likelihood

Mathematical notation

$\sim$	the ratio of the left- and right-hand sides tends to one in the limit
$\approx$	approximately equal to
$\in$	contained in
$\mathbf{1}$	column vector of ones of size n
$a.s$	almost surely
$\to a^+$	limit from above
$[a, b]$	set of values x such that $a \leq x \leq b$
(a, b)	set of values x such that $a < x < b$
$(a, b]$	set of values x such that $a < x \leq b$
$\lvert\mathbf{A}\rvert$	determinant of matrix $\mathbf{A}$
$\lVert\mathbf{A}\rVert$	norm of matrix $\mathbf{A}$
$\mathbf{A}^\top$	transpose of matrix $\mathbf{A}$
$\imath$	complex number such that $\imath^2 = -1$
$\log(.)$, $\ln(.)$	logarithmic function with base e
$\lfloor x \rfloor$	the greatest integer not greater than x
$\mathbb{R}$	$(-\infty, \infty)$
$\mathbb{Z}$	$\{\ldots, -2, -1, 0, 1, 2, \ldots\}$

Statistical notation

$a.s$	almost surely
$\rightarrow^{D}$	convergence in distribution
$\rightarrow^{P}$	convergence in probability
$\theta_1, \ldots, \theta_q$	MA coefficients of the ARMA/ARIMA/ARFIMA model
$\phi_1, \ldots, \phi_p$	AR coefficients of the ARMA/ARIMA/ARFIMA model
d	fractional differencing parameter/long memory parameter
$\boldsymbol{\beta}$	vector of AR and MA coefficients
$\boldsymbol{\eta}$	vector of dynamics parameters such that $\boldsymbol{\eta} = \left(d, \boldsymbol{\beta}^{\top}\right)^{\top}$
$\boldsymbol{\vartheta}_0$	true value of the vector parameter $\boldsymbol{\vartheta}$
$\boldsymbol{\vartheta}_1$	pseudo-true value of the vector parameter $\boldsymbol{\vartheta}$
$Var\,(.)$	variance operator
$Cov\,(.,.)$	covariance operator
$Corr\,(.,.)$	correlation operator
$E\,(.)$	expectation operator
$f\,(\lambda)$	spectral density of the stochastic process $\{X_t\}$
$f_{XY}\,(\lambda)$	cross spectral density of the stochastic processes $\{X_t\}$ and $\{Y_t\}$
$I\,(\lambda)$	periodogram
$M_{X_1,\ldots,X_k}\,(t_1, \ldots, t_k)$	moment generating function of the multivariate random variables $X_1, \ldots, X_k$
μ	process mean
σ^2	error variance
$N\left(\mu, \sigma^2\right)$	normal distribution with zero mean and variance σ^2
$o_p\,(1)$	a sequence of random variables that is converging to zero, in probability
$O_p\,(1)$	a sequence of random variables that is bounded, in probability
$\gamma\,(k)$	autocovariance at lag k
$\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$	autocovariance matrix
$\chi^2_{(k)}$	chi-square distribution with k degrees of freedom
Z	standard normal $[N\,(0,1)]$ random variable, unless stated otherwise

Abstract

Data in the economic and financial spheres often exhibit dynamic patterns characterized by a long lasting response to past shocks. The correct modelling of such long range dependence is of paramount importance, both in the production of accurate forecasts over long term horizons and in the isolation of long run equilibrium relationships. While the convention in the area has been to adopt complete parametric specifications for the dynamics in the time series, semi-parametric approaches have also featured. This thesis contributes to both of these lines of research. In Chapter 3 we develop the asymptotic theory for quantifying the impact of mis-specification of short memory dynamics in the context of parametric estimation of the parameter controlling the long range dependence. The methodology is developed within the framework of fractionally integrated processes, as introduced by Granger and Joyeux (1980) and Hosking (1981). We provide a comprehensive set of new results on the impact of mis-specifying the short run dynamics in such processes. We show that four alternative parametric estimators – frequency domain maximum likelihood, Whittle, time domain maximum likelihood and conditional sum of squares – converge to the same pseudo-true value under common mis-specification, and that they possess a common asymptotic distribution. The results are derived assuming a completely general parametric specification for the short run dynamics of the estimated (mis-specified) fractional model, and with long memory, short memory and antipersistence in both the model and the true data generating process accommodated.

As well as providing new theoretical insights, we undertake an extensive set of numerical explorations, beginning with the numerical evaluation, and implementation, of the (common) asymptotic distribution that holds under the most extreme form of mis-specification. Simulation experiments are then conducted to assess the relative finite sample performance of all four mis-specified estimators, initially under the assumption of a known mean, as accords with the theoretical derivations in this chapter. The importance of the known mean assumption is illustrated via the production of an alternative set of bias and mean squared error results, in which the estimators are applied to demeaned data. The chapter concludes with a discussion of open problems.

In Chapter 4, we then establish the limiting behaviour of parametric estimators (time domain maximum likelihood, conditional sum of squares and exact Whittle) under mis-specification of short memory dynamics, while allowing the process mean to be unknown. We also show that the limiting distributions of the three parametric estimators are identical to those of the frequency domain maximum likelihood and discrete Whittle estimators, regardless of whether the process mean is known or unknown. In order to estimate the mean, we consider two estimators, namely, the sample mean estimator and the best linear unbiased estimator [BLUE]. Our results show that the sample mean estimator is unaffected by model mis-specification. However, the limiting behaviour of BLUE is sensitive to mis-specification of the short memory dynamics. We establish the consistency of both estimators of the unknown mean under correct specification as well as under mis-specification. Monte Carlo simulations are used to quantify the finite sample behaviour of the estimators of the long memory parameter when the mean is also estimated.

In Chapter 5 we then focus on semi-parametric estimation. Specifically, we use the jackknife to bias correct the log-periodogram regression [LPR] estimator of the fractional parameter, d, in a stationary fractionally integrated model. The weights used to construct the jackknife estimator are chosen such that bias reduction occurs to an order of $n^{-\alpha}$ (where n is the sample size) for some $0 < \alpha < 1$, while the increase in variance is minimized - with the weights viewed as 'optimal' in this sense. We show that under regularity, the bias-corrected estimator is consistent and asymptotically normal with the same asymptotic variance and $n^{\alpha/2}$ rate of convergence as the original LPR estimator. In other words, the use of optimal weights enables bias reduction to be achieved without the usual increase in asymptotic variance being incurred. These theoretical results are valid under both the non-overlapping and moving-block sub-sampling schemes that can be used in the jackknife technique, and do not require the assumption of Gaussianity for the data generating process. A Monte Carlo study explores the finite sample performance of different versions of the optimal jackknife estimator under a variety of data generating processes, including alternative specifications for the short memory dynamics. The comparators in the simulation exercise are the raw (unadjusted) LPR estimator and two alternative bias-adjusted estimators, namely the weighted average estimator of Guggenberger and Sun (2006) and the pre-filtered sieve bootstrap-based estimator of Poskitt *et al.* (2016). The chapter concludes with some discussion of open issues and possible extensions to the work.

Chapter 1 and 2 provide background material for the research contributions contained in Chapter 3 to 5. Chapter 6 concludes the thesis with an overview of what has been achieved herein, plus details of the future research agenda.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma in any other university or equivalent institution, and that, to the best of my knowledge and belief, this thesis contains no material previously published by another person, except where due reference is made in the text of thesis.

Signature: ....

Print name: Kanchana Nadarajah

Date: 27 May 2019

Thesis including published works declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

The core theme of the thesis is 'Advances in the Estimation of Fractionally Integrated Models'. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the Department of Econometrics and Business Statistics under the supervision of Professor Gael Margaret Martin and Professor Donald Stephen Poskitt. This thesis includes one paper submitted for publication to the Journal of Econometrics, and now under first-round revision for that journal.

In the case of Chapter 3 my contribution to the work is summarized as follows:

Thesis Chapter	Publication Title	Status (published, in press, accepted or returned for revision, submitted)	Nature and % of student contribution	Co-author name(s) Nature and % of Co-author's contribution	Co-author(s), Monash student Y/N
3	Issues in the estimation of mis-specified models of fractionally integrated processes	Under first-round revision for Journal of Econometrics	Overall writing and technical responsibility 60%	1) Gael M Martin Technical and editorial assistance [T & EDA] 20% 2) Don S Poskitt [T & EDA] 20%	No No

I have / have not renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Student signature: .... Date: ....

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor signature: .... Date: ....

Acknowledgments

First and foremost, I offer my sincerest gratitude to my supervisors, Professor Gael Martin and Professor Donald Poskitt, Department of Econometrics and Business Statistics, Monash University, who have supported me with patience and knowledge whilst allowing me to think independently. This thesis grew out of a series of dialogues with my supervisors. Their comments on chapter drafts are themselves a source from which I will always draw. Amid the chapter drafts, they were always there to offer quiet encouragement, and I was at times impressed by their extraordinary kindness. It is my great pleasure to express my deepest gratitude to them. I attribute the level of my Doctorate degree to their encouragement and effort and without them this thesis, too, would not have been completed or written. One simply could not wish for better or friendlier supervisors.

I wish to thank Professor Rob Hyndman, the Head of the Department of Econometrics and Business Statistics for his immense support in postgraduate affairs and for being a referee. I also take this opportunity to express my deepest gratitude to Rob and my other PhD panel members, Professor Mervyn Silvapulle and Dr Hsien Kew, for assisting the progress of my research via three milestones during the candidature, and for their insightful comments. I wish to thank Professor Heather Anderson, the former Head of the Department of Econometrics and Business Statistics for her invaluable advice and help in my first few years of my candidature. I would like to convey my thanks to Professor Farshid Vahid, the PhD coordinator, for the support during the thesis submission process. Also, my sincere thanks goes to all the other academic and administrative staff members of the Department who willingly helped me in numerous ways during this study. I gratefully acknowledge the Department for providing the financial support for international and domestic conferences during the candidature. My thanks also goes to Monash University for awarding me a Monash Postgraduate travel grant.

With reference to my Monte Carlo simulation work, I am particularly indebted to Mr. Philip Chan, Senior e-Research Software Specialist, Monash e-Research Centre who helped me access the high-performance computing facilities at Monash. His support is very much appreciated.

Besides the staff members at the Department, I would like to convey my special thanks to Professor Liudas Giraitis whom I met at the 4th International Society of Nonparametric Statistics conference held in Salerno, Italy. Thanks are due to him for providing valuable suggestions on my thesis chapters. My great respect and sincere thanks go to Professor Vance Martin for all the comments on my chapters and for giving professional guidance.

I sincerely acknowledge the Department of Education, Australia, for the Endeavour Postgraduate Scholarship Award, which made my PhD degree possible.

My time at Monash introduced me to many friends and groups that became a part of my life. I am grateful for the time spent with my colleagues and friends. They have always been supportive and encouraged me to do my best in all matters of life. My cheerful moments and our memorable trips are always in my memory.

Most importantly, I would like to thank my parents for their unconditional love, moral support and encouragement. Their good examples have taught me to work hard for the things that I aspire to achieve in my life. My initial learning in mathematics was guided by my father, Mr. Sockalingam Nadarajah. He raised me with a love for mathematics. My mother, Mrs. Vijayaluxmy Nadarajah's continuous support and guidance in every step in my life, has caused me to be steadfast and never bend at my difficult times. She always let me know that she is so proud of me, which motivates me to work hard and show my best. Most invaluable of all has been the love and support provided by my encouraging and patient sister, Miss. Sailajah Nadarajah with whom, up until now, I have shared most of my time. Without my family, it would have been impossible for me to reach the present position in my life. I have no words to express my heartfelt thanks to my family. My very special thanks go to my beloved husband, best friend and mentor Dr. Bethmage Sandun Indeewara Perera, who has been a constant source of support and encouragement during the challenges of post graduate and personal life. I am truly thankful for having him in my life. This work is not possible without his constant support. My dearest Indee, I do not know how to extend my gratitude towards you. To them, I dedicate this thesis.

Finally, I thank all those who have helped me directly or indirectly in the successful completion of my thesis. Anyone missed in this acknowledgement is also thanked.

Chapter 1

Introduction

1.1 Background

Stochastic processes relating to the phenomenon of a long lasting response to past shocks have been recognized in many different fields, after the initial investigation of yearly Nile river flow minima by Hurst (1951). Notable empirical evidence of those processes has occurred in various fields including hydrology (Lawrance and Kottegoda, 1977 and Ooms and Franses, 2001), meteorology (Gil-Alana, 2012), computer science (Leland *et al.*, 1993, Karagiannis *et al.*, 2004 and Scherrer *et al.*, 2007), human science (Wagenmakers *et al.*, 2004), image texture recognition (Lundahl *et al.*, 1986), economics (Diebold and Rudebusch, 1989 and Baillie, 1996) and finance (Hassler and Wolters, 1995, Bollerslev and Mikkelsen, 1996, Andersen *et al.*, 2003, Cheung, 2016 and Varneskov and Perron, 2018). The important characterization of such 'long memory (or long range dependent) processes' is the hyperbolic rate of decay in autocorrelations or autocovariances, a rate that is too slow produce summability; in contrast to the usual exponential, and summable, decay associated with a short memory process. For more details on such processes and their applications refer to Beran (1994), Doukhan *et al.* (2003), Giraitis *et al.* (2011) or Pipiras and Taqqu (2017), and the review article by Baillie (1996).

In the literature, several definitions of long memory have been extensively discussed (see, Beran, 1994 and Palma, 2010), while the articles by Cox (1984) and Heyde and Yang (1997) give an overview of different definitions of long memory processes. Most of the definitions of long range dependence appearing in literature are based on the second-order properties of a time series, such as, the autocovariances, the spectral density, and the variances of partial sums. In this thesis, we define a long memory stationary time series as follows. The idea behind the definition is that *long range* dependence occurs when the autocovariances tend to zero in the form of a power function of lags as the lags tend to infinity, and decline slowly enough for the sum of autocovariances to diverge.

Definition 1.1 *A stationary series $\{y_t\}$, $t \in \mathbb{Z}$, with finite variance, has* (potential) *long memory if the autocovariance at lag k,*

$$\gamma(k) = Cov(y_t, y_{t+k}) \sim c_\gamma k^{2d-1}, \quad 0 < c_\gamma < \infty, \quad as \quad k \to \infty,$$

where $-0.5 < d < 0.5$. The symbol $\sim$ means that the ratio of the left- and right-hand sides tends to one in the limit.

Definition 1.1 allows the process $\{y_t\}$ to be classified as a short memory, long memory or intermediate memory process depending on the value of the parameter d. When $d = 0$, then the process $\{y_t\}$ is said to have short memory and the autocovariances are absolutely summable. The sum of the autocovariances converges to a positive real number and the autocovariances decay at an exponential rate. The process $\{y_t\}$ is identified as a long memory process if $d > 0$, and, the autocovariances are non-summable and decay hyperbolically. Note that the autocovariances are mostly positive. If $d < 0$, then the process is said to have intermediate memory and the autocovariances are mostly negative. In this case the autocovariances decay at a hyperbolic rate, but fast enough to be summable.

A class of statistical models that plays a central role in modelling such long memory processes and which was introduced by Granger and Joyeux (1980) and Hosking (1981), describes a time series $\{y_t\}$ in the form of

$$\phi(L)(1-L)^d\{y_t-\mu\}=\theta(L)\varepsilon_t, \tag{1.1}$$

where $\mu = E(y_t)$, L is the lag operator such that, $L^k x_t = x_{t-k}$ for $k \geqslant 0$ and $\phi(z) = 1 + \phi_1 z + \ldots + \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \ldots + \theta_q z^q$ are the autoregressive and moving average operators respectively, where it is assumed that $\phi(z)$ and $\theta(z)$ have no common roots and that the roots lie outside the unit circle. The errors $\{\varepsilon_t\}$ define a white noise sequence with finite variance $\sigma^2 > 0$. Further, d is the fractional differencing parameter assumed to lie within the range mentioned in Definition 1.1. If $d > -1$, the characteristic polynomial of the lag operator, $(1-z)^d$, can be represented by the binomial expansion with $|z| < 1$ as follows,

$$(1-z)^d = 1 - dz + d(d-1)\frac{z^2}{2!} - d(d-1)(d-2)\frac{z^3}{3!} + \ldots .$$

Therefore,

$$(1-z)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} z^j,$$

where $\Gamma(.)$ is the gamma function. The class of models in (1.1) is known as *fractionally integrated autoregressive moving average* [ARFIMA] class, and is closely related to the class of autoregressive moving average [ARMA] models. In summary notation, the above model is denoted by ARFIMA(p, d, q), where p and q are the numbers of AR and MA coefficients in the model. The parameters in (1.1) can be described as dynamic and static parameters. For example, the differencing parameter as well as the ARMA coefficients are referred to as dynamic parameters. Examples of static parameters are the location and scale parameters.

The literature deals with the methods and theory associated with parametric and semi-parametric estimation of the parameters of the model in (1.1). A short review of this literature is provided in Chapter 2, with further review material appearing in each of the introductory sections in Chapters 3 to 5.

The common practice adopted in the parametric literature is to assume that the model is correctly specified for the data generating process [DGP]. Conditional on this assumption, the convergence and distributional properties of various parametric estimators are developed – including their asymptotic equivalence in some cases; for example, see Fox and Taqqu (1986), Dahlhaus (1989), Giraitis and Surgailis (1990), Sowell (1992), Beran (1995), Robinson (2006) and Hualde and Robinson (2011), among others. In practice, the true values of p or q - the number of AR and MA components - are not known. Whilst there are several diagnostic tests available for the fit of a model; for example, goodness-of-fit tests (see Beran, 1994, Chapter 10) such as the Akaike information criterion [AIC], the Bayes information criterion [BIC] and the final prediction error [FPE], these methods have a low success rate in identifying the correct order of the short memory dynamics (see Crato and Ray, 1996). Incorrect specification of p or q can, in turn, lead to serious consequences for the characteristics of the parametric estimators (see Yajima, 1993). As is described in Chapter 2, some attempts have been made to address the issue of mis-specification of the short memory dynamics in certain fractionally integrated models (see Chen and Deo, 2006).

An alternative to undertaking (potentially mis-specified) parametric estimation is to use a semi-parametric method, which provides an estimate of the fractional differencing parameter without requiring the specification of the short memory component; for example, the log-periodogram regression estimator of Geweke and Porter-Hudak (1983), Gaussian semi-parametric estimator of Robinson (1995a) and the exact local Whittle estimator of Shimotsu and Phillips (2005), among others. The price paid for such an approach, however, is the occurrence of large finite sample bias, and several bias-correction procedures have been proposed in the literature as a consequence (see, Andrews and Guggenberger, 2003, Andrews and Sun, 2004, and Poskitt *et al.*, 2016).

The objective of this thesis is to develop new methodological and theoretical results in the fields of both parametric and semi-parametric estimation of long memory processes. First, we explore the consequences for parametric estimation of mis-specification of the short memory dynamics, allowing the process mean to be, respectively, known and unknown. Secondly, we develop an optimally bias-corrected semi-parametric estimator of the long memory parameter.

1.2 Outline of the thesis

In Chapter 2 we first briefly review the literature on parametric estimation methods for fractionally integrated models. A short review of the existing theory for parametric estimation in mis-specified long memory models is also provided. Secondly, semi-parametric estimation methods are described. This includes a brief review of existing bias-reduction methods for semi-parametric estimators.

In Chapter 3 we then proceed to provide a comprehensive new set of results on parametric estimation under mis-specification in the fractional setting. The starting point here has been the work of Chen and Deo (2006), who first investigated the consequences of mis-specification of short memory dynamics on the asymptotic properties of the approximate frequency domain maximum likelihood [FML] estimator. Under mis-specification, the standard asymptotic properties are shown not to hold for this estimator. In particular, when certain regularity conditions hold, the FML estimator of the dynamic parameters is shown to converge to a well-defined limit called the pseudo-true value that is different from the true value. The rate of convergence is sometimes slower than $\sqrt{n}$ and only under certain conditions is the $\sqrt{n}$ rate still achieved. Under certain forms of mis-specification, the limiting distribution is also non-Gaussian, with asymptotic normality shown to hold only in some instances. All of the asymptotic results of Chen and Deo are established for a Gaussian DGP.

In this chapter we extend the results of Chen and Deo (2006) to three different parametric estimators, namely, the discretized version of the Whittle [DWH] estimator, and the time domain maximum likelihood [TML] and conditional sum of squares [CSS] estimators, under the assumption that the process mean is known. We show that all three parametric estimators converge to the same pseudo-true value under common mis-specification. Closed-form expressions for the first-order conditions that define the pseudo-true values of the parameters are provided. Further, the limiting distributions of DWH, TML and CSS are identical to those of FML; that is, all four parametric estimators (FML, DWH, TML and CSS) are asymptotically equivalent. Furthermore, it is shown that our results are valid for any stationary ARFIMA process, including a non-Gaussian process. In a simulation study, we investigate the finite sample performance of the four parametric estimators of the long memory parameter (specifically), in terms of bias, RMSE and the form of the sampling distribution. This work is under first-round revision at *Journal of Econometrics* (Martin *et al.*, 2018) and is reproduced in Chapter 3 in its complete form.

From a theoretical perspective, it is restrictive to impose the requirement of the known process mean. However, relaxing this assumption may have consequences for the estimation of the dynamic parameters, as the mean estimators that are commonly used in the context of fractionally integrated models, such as the sample mean or the best linear unbiased estimator [BLUE], have a slower rate of convergence than the usual $\sqrt{n}$ (see Adenstedt, 1974 and Hosking, 1996). Therefore, the questions that arise are; (i) How does the limiting behaviour of the parametric estimators of the dynamic parameters under mis-specification differ in the two cases of known and unknown mean? and, (ii) Does the nature of the mean estimation influence the finite sample ranking of the parametric estimators of the dynamic parameters under mis-specification?

The FML and DWH estimation methods are invariant to the mean. Therefore, in Chapter 4 we consider only the TML and CSS estimation methods, in addition to the exact version of the Whittle estimator [EWH], when examining the impact of estimating a mis-specified fractionally integrated model with an unknown mean. We prove that all three of these parametric estimators converge to the same pseudo-true value as do the invariant FML and DWH estimators, and that all five estimators share the same limiting distribution, regardless of the form of mean estimation. A simulation study is provided to illustrate the finite sample performance of the EWH, TML and CSS estimators of the long memory parameter (specifically), when the mean is estimated by either the sample mean or the BLUE. The study reveals that the choice of mean estimator does not influence the finite sample performance. However, the ranking of the estimators does alter due to mean estimation *per se*. Whilst in Chapter 3 a small Monte-Carlo study for the unknown mean case is included, with the sample mean used, in this chapter we not only provide the theoretical results relevant to mean estimation, but also a more extensive numerical study based on both the sample mean and the BLUE. Furthermore, a detailed technical discussion is provided on BLUE under both the correct and incorrect specification of the model. This chapter has been written as a draft for a self-contained article for journal submission.

Chapter 5 switches focus to semi-parametric estimation of the long memory parameter in a fractional model. Semi-parametric methods have the advantage that they do not require specification the short memory dynamics and, hence, are not subject to the problem of misspecification thereof (see Robinson, 2014). Moreover, these methods are generally easy to implement and computationally attractive compared to certain parametric techniques. However, the most commonly used semi-parametric methods, for example, the log-periodogram regression [LPR] estimator of Geweke and Porter-Hudak (1983) and the local Whittle [LW] estimator of Robinson (1995a), exhibit large finite sample bias in the presence of true, and unmodelled, short memory dynamics. As a consequence, several bias-reduced semi-parametric methods have been proposed, as referenced earlier.

This bias reduction, however, comes at a cost of increased sampling variance. In the context of a unit root process, Chen and Yu (2015) propose a methodology using a non-parametric technique called the jackknife to optimally correct the bias while minimizing the associated increase in variance. By adopting their approach in the fractionally integrated context, we develop an optimally bias-corrected LPR estimator of the long memory parameter. We show that the proposed estimator is consistent for the true value of the fractional differencing parameter, and that a limiting normal distribution is achieved along with no loss in asymptotic efficiency. One requirement of the proposed method is that the true values of the short memory parameters of the underlying model are known. We thus also propose a feasible version of the estimator, based on an iterative procedure. A simulation study is used to illustrate the finite sample performance of both proposed estimators, in comparison with relevant alternatives, including the method of Guggenberger and Sun (2006), which also produces bias reduction with reduced variance inflation. This chapter has also been written as a draft for a self-contained article for journal submission.

Chapter 6 provides an outline of the thesis, restating the overall objectives and the main issues investigated. We also discuss possible avenues for future research that have emerged from the research conducted in the thesis.

Chapter 2

Parametric and semi-parametric estimation of fractionally integrated models

In this chapter, we briefly review existing parametric and semi-parametric estimation methods for the class of ARFIMA models introduced in Chapter 1. We first provide a brief review of the literature on the parametric modelling of long memory processes using ARFIMA models in Section 2.1. A short discussion of the literature on the estimation of the location parameter is provided in Section 2.2. The properties of parametric estimators of the dynamic parameters, under both correct and incorrect specification of the model, are outlined in Section 2.3. Section 2.4 describes the semi-parametric estimation methods, including available bias-correction procedures. Section 2.5 concludes the chapter.

2.1 Underpinnings of ARFIMA models

We consider the ARFIMA(p, d, q) model for long memory processes as specified in (1.1), with $-0.5 < d < 0.5$. An alternative form of this model (see, for example, Poskitt, 2008) is characterized with an infinite order moving average [MA] process as

$$y_t = \mu + \sum_{j=-\infty}^{\infty} b_j (\boldsymbol{\eta}) \varepsilon_{t-j}, \tag{2.1}$$

where $\sum_{j=-\infty}^{\infty} c_j^2 < \infty$ with

$$b_j(\boldsymbol{\eta}) = \sum_{s=0}^{j} \frac{\varsigma_{j-s}(\boldsymbol{\beta})\,\Gamma(s+d)}{\Gamma(s+1)\,\Gamma(d)}, \quad j = 1, 2, \ldots, \tag{2.2}$$

and $\varsigma(.)$ is the transfer function of a stable and invertible autoregressive moving average process. Here, we denote by $\boldsymbol{\eta}$ the vector of dynamic parameters as $\boldsymbol{\eta} = \left(d, \boldsymbol{\beta}^{\top}\right)^{\top} \in \mathbb{E}$ where $\boldsymbol{\beta}^{\top} = \left(\phi_1, \phi_2, ..., \phi_p, \theta_1, \theta_2, ..., \theta_q\right) \in \mathbb{B}$ is the vector of coefficients of the short memory components of the model such that $\mathbb{E} = \mathbb{D} \times \mathbb{B}$, with $\mathbb{D} = (-0.5, 0.5)$ and $\mathbb{B}$ is an l-dimensional compact convex set in $\mathbb{R}^l$ with $l = p + q$. The only requirement for the above representation is that the innovations have a finite mean and finite variance. Thus, Gaussian innovations are not necessary.

An alternate representation is as an infinite order autoregressive [AR] process,

$$y_t - \sum_{j=-\infty}^{\infty} \tau_j(\boldsymbol{\beta})\, y_{t-j} = \varepsilon_t, \tag{2.3}$$

where

$$\tau_j(\boldsymbol{\eta}) = \sum_{s=0}^{j} \frac{\alpha_{j-s}(\boldsymbol{\beta})\,\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}, \quad j = 1, 2, \ldots. \tag{2.4}$$

The spectral density of the ARFIMA(p, d, q) process in (1.1) is given by (refer, Theorem 13.2.2 of Brockwell and Davis, 1991, and Nielsen and Frederiksen, 2005)

$$f(\boldsymbol{\eta}, \lambda) = \frac{\sigma^2}{2\pi} \left[\sin(\lambda/2)\right]^{-2d} g(\boldsymbol{\beta}, \lambda), \quad -\pi \leq \lambda \leq \pi, \tag{2.5}$$

where $g(\boldsymbol{\beta}, \lambda)$ is the spectral density of the autoregressive moving average [ARMA] component taking the form,

$$g(\boldsymbol{\beta}, \lambda) = \frac{|\theta(\exp(\imath\lambda))|^2}{|\phi(\exp(\imath\lambda))|^2}. \tag{2.6}$$

Here $g(\boldsymbol{\beta}, \lambda)$ is bounded above and bounded away from zero with continuous second derivatives. This implies that as $\lambda \to 0$, the limit of $\lambda^{2d} f(\boldsymbol{\eta}, \lambda)$ exists and is finite. Refer .

The autocovariance at lag k, $\gamma(k)$ is defined as follows (see Sowell, 1992, pages. $171-174$),

$$\gamma(k) = \sigma^2 \sum_{l=-q}^{q} \sum_{j=1}^{p} \nu(l) \varsigma_j C(d, p+l-k, \rho_j), \tag{2.7}$$

where,

$$\begin{aligned} \nu(l) &= \sum_{k=\max(0,l)}^{\min(q,q-l)} \theta_k \theta_{k-l}, \\ \varsigma_j &= \left[\varrho_j \prod_{i=1}^{p} (1-\varrho_i \varrho_j) \prod_{m \neq j} (\varrho_j - \varrho_m) \right]^{-1}, \end{aligned}$$

with the ϱ_j's defined by, $\phi(z) = \prod_{j=1}^{k}(1-\rho_j z)$, and,

$$\begin{aligned} C(d, p+l-k, \rho_j) &= \frac{\Gamma(1-2d)\Gamma(d+p+l-k)}{\Gamma(1-d+p+l-k)\Gamma(1-d)\Gamma(d)} \\ &\quad \times [\rho_{j2} F_1(d+p+l-k, 1; 1-d+p+l-k; \rho_j) \\ &\quad +_2F_1(d-p-l+k, 1; 1-d-p-l+k; \rho_j) - 1]. \end{aligned}$$

Here, ${}_2F_1(a, b; c; x)$ is the hypergeometric function, which is a special representation of the hypergeometric series given below,

$${}_lF_m(a_1, a_2, ...a_l; b_1, b_2, ...b_m; x) = \sum_{j=0}^{\infty} \frac{(a_1)_j (a_2)_j ... (a_l)_j}{(b_1)_j (b_2)_j ... (b_m)_j} x^j,$$

where, for some parameter c, $(c)_j$ is defined as $(c)_0 = 1$, and,

$$(c)_j = c(c+1)(c+2)...(c+j-1), j = 1, 2, ...$$

The autocorrelation is defined by $\rho(k) = \gamma(k)/\gamma(0)$, for $k = 0, \pm 1, \pm 2, \ldots$.

Figure 2.1 displays the spectral density and the correlogram of an ARFIMA$(1, d, 0)$ process with an AR coefficient of $\phi = -0.9$. The spectral density diverges at zero frequency for $0 < d < 0.5$ and declines as the frequency increases. For $d \leq 0$, the spectral density is continuous on $[-\pi, \pi]$ and bounded above. The correlogram shows the slow decay of the autocorrelations

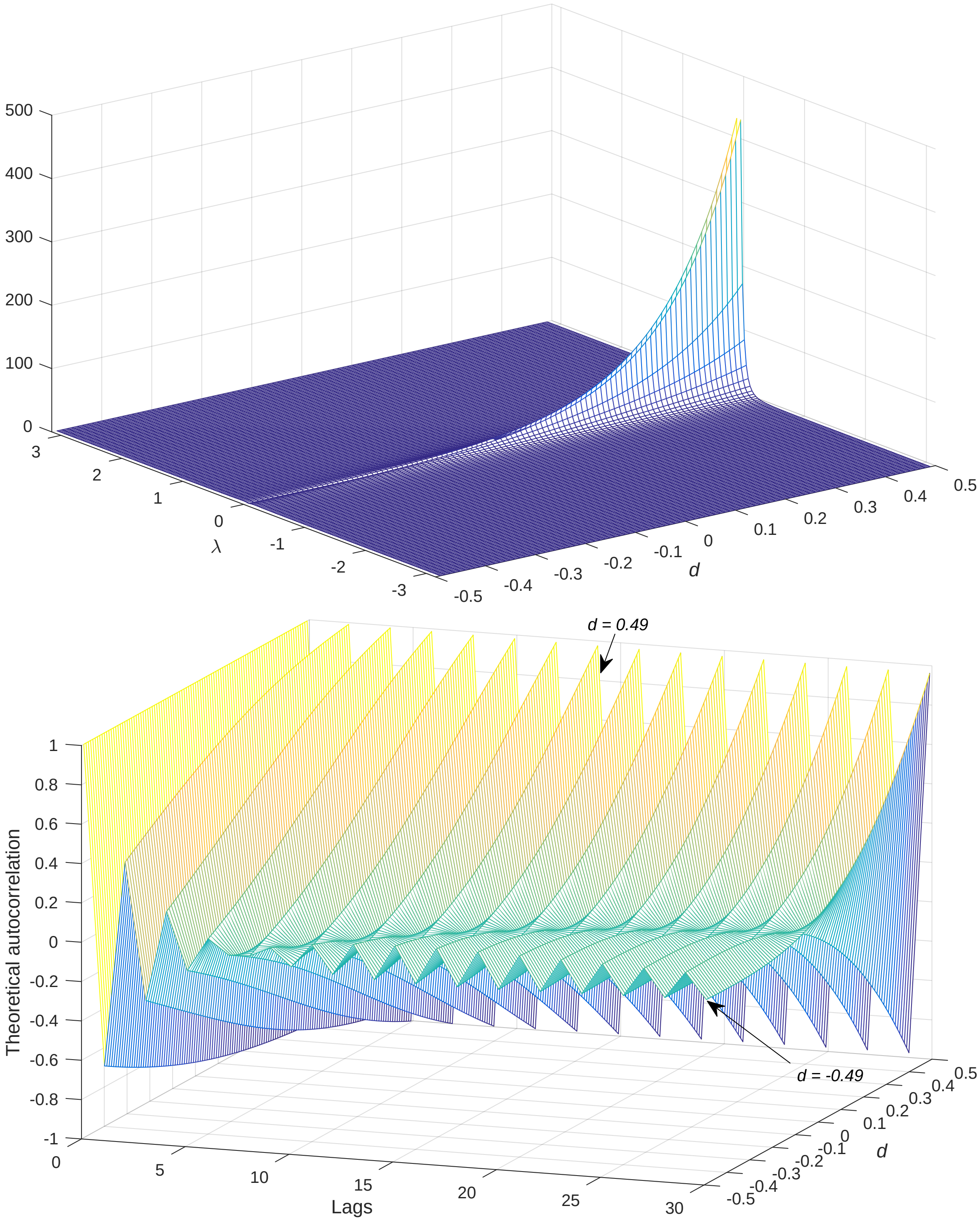

Figure 2.1: Spectral density (displayed in the top panel) and autocorrelation function (displayed in the bottom panel) of ARFIMA$(1, d, 0)$ with the AR coefficient $\phi = -0.9$.

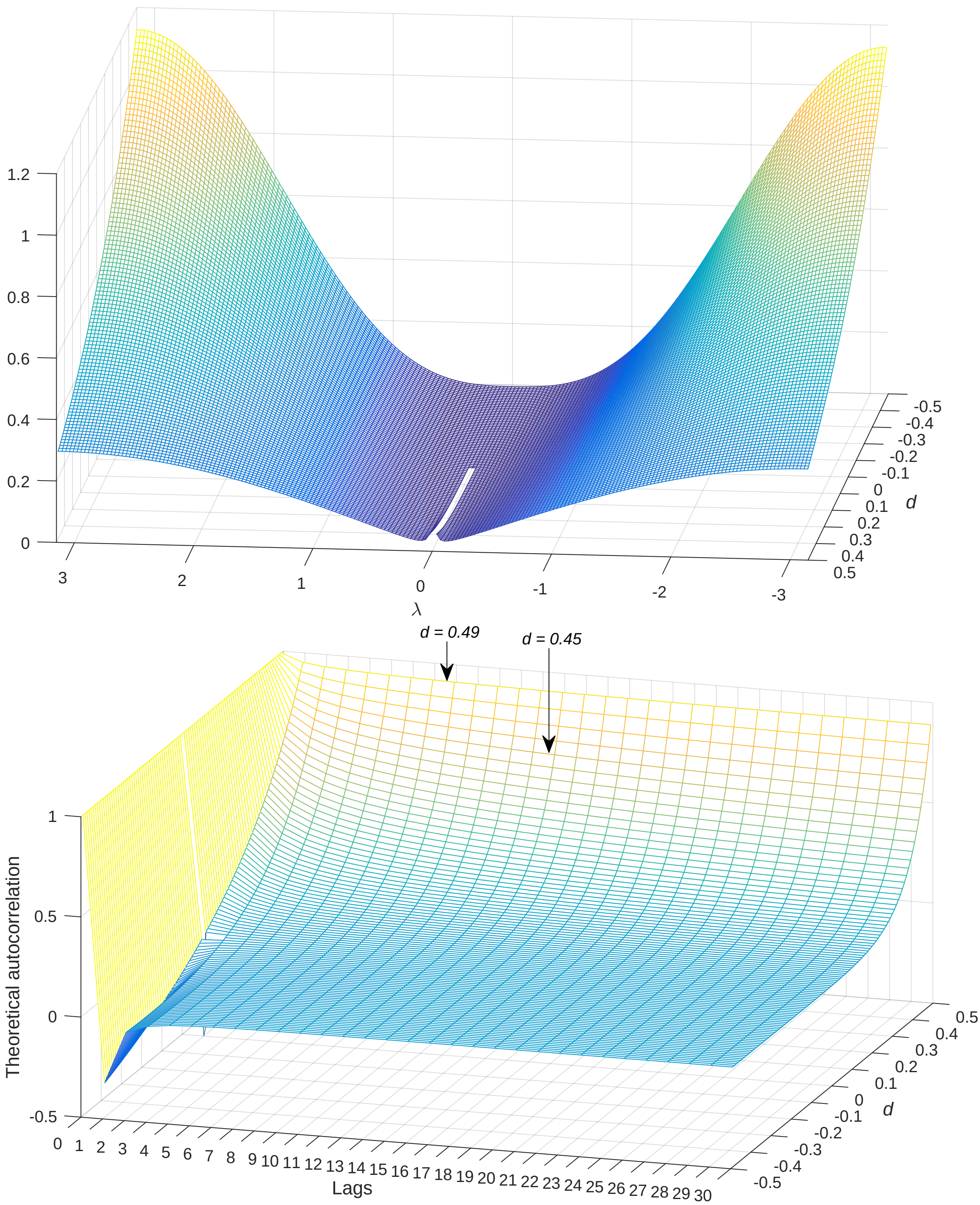

Figure 2.2: Spectral density (displayed in the top panel) and autocorrelation function (displayed in the bottom panel) of ARFIMA$(0, d, 1)$ with the MA coefficient $\theta = -0.9$.

for $0 < d < 0.5$ and the slow decay is clearly visible for values of d closer to 0.5, at small lags. However, the slow decay becomes less obvious when the values of d move towards zero. For $-0.5 < d < 0$, the decay is slightly faster than that of long memory dependence, but not as fast as the exponential rate that occurs for $d = 0$. At higher lags, although the autocorrelations are closer to zero, they are significantly different from zero. The oscillation of the autocorrelations is observed due to the negative AR coefficient (i.e. positive correlation between the observations).

Figure 2.2 shows the spectral density and the correlogram of an ARFIMA$(0, d, 1)$ process with an MA coefficient of $\theta = -0.9$. The spectral density behaves in a similar manner in the neighbourhood of zero frequency to that of the ARFIMA$(1, d, 0)$ process, for $0 < d < 0.5$. When $d \leq 0$, the spectral density has an inverted bell shape showing the continuity on $[-\pi, \pi]$, being bounded above and is zero at zero frequency. Moreover, as the frequency increases to $\pm\pi$, the density curve increases to some point. The correlogram shows similar features as mentioned previously for the ARFIMA$(1, d, 0)$ process, except for the fact that there is no pattern of oscillation observed. A sharp downward peak is observed at lag 1 for the short memory process ($d = 0$) as this simplifies to an MA(1) process.

2.2 Estimation of the location parameter

The first step in the statistical analysis of the class of ARFIMA models defined in (1.1) is estimation of the process mean, μ, together with the scale parameter, σ^2. In this section, we focus on estimation of the mean and the relevant theoretical results established in the literature.

Let $\mathbf{y}^\top = (y_1, y_2, ..., y_n)$ be a random sample of n observations, with a spectral density as given in (2.5). The simplest estimator of the mean is the sample mean, $\widehat{\mu}_{SM} = \overline{y} = \sum_{t=1}^{n} y_t / n$.

Hosking (1996) established theory for the sample mean in the fractional model. It is shown that the sample mean is consistent for μ, but with a rate of convergence, $n^{1/2-d_0}$, that depends on the true value of the fractional differencing parameter, $d_0 \in (-0.5, 0.5)$. Further, $n^{1/2-d_0}\left(\widehat{\mu}_{SM} - \mu_0\right)$ is asymptotically normal with zero mean and asymptotic variance,

$$v_0^2 = \frac{\sigma_0^2 g\left(\boldsymbol{\beta}_0, 0\right)\Gamma\left(1-2d_0\right)}{\left(1+2d_0\right)\Gamma\left(1+d_0\right)\Gamma\left(1-d_0\right)}, \tag{2.8}$$

where $B\left(.\right)$ is the beta function and $g\left(\boldsymbol{\beta}_0, 0\right)$ is the spectral density of the ARMA component evaluated at zero frequency and at the true values of the parameters. Hosking shows that these results are valid under weaker conditions than Gaussianity of the time series.

An alternative unbiased estimator for the mean is the BLUE introduced by Adenstedt (1974). The form of the BLUE is simply the weighted average of the sample observations and is defined by

$$\widehat{\mu}_{BLU,0} = \frac{\mathbf{1}^{\top}\boldsymbol{\Sigma}_0^{-1}\mathbf{y}}{\mathbf{1}^{\top}\boldsymbol{\Sigma}_0^{-1}\mathbf{1}}, \tag{2.9}$$

where $\mathbf{1}$ is the column vector of n ones and $\sigma_0^2\boldsymbol{\Sigma}_0$ is the true variance–covariance matrix of the time series such that $\sigma_0^2\boldsymbol{\Sigma}_0 := \left[\gamma_0\left(i-j\right)\right], i, j = 1, \ldots, n$, with $\gamma_0\left(.\right)$ representing the autocovariance of the true data generating process [TDGP] following the form defined (2.7). The variance of BLUE is given by $\sigma_0^2\left(\mathbf{1}^{\top}\boldsymbol{\Sigma}_0^{-1}\mathbf{1}\right)^{-1}$. Samarov and Taqqu (1988) show that for large n, there is no asymptotic efficiency loss incurred by the sample mean over the BLUE for short range processes, and that this loss remains small for stationary long memory processes. However, it is shown to be much greater for antipersistent processes, such that BLUE is preferred over the sample mean in this case. This may be due to the elements of $\left(\sigma_0^2\boldsymbol{\Sigma}_0\right)^{-1}$ which are dominated by the inversion of the spectral density[1]. Particularly, when $d_0 < 0$, the contribu-

[1]The inversion of the spectral density of an ARFIMA(p, d, q) process is the spectral density of an ARFIMA$(q, -d, p)$.

tion of the inversion of the spectral density leads to significant loss in asymptotic efficiency.

Provided the process is Gaussian, BLUE is simply the maximum likelihood estimator [MLE] of the mean, and is the most efficient estimator in the linear unbiased class. Among the class of linear unbiased estimators, BLUE is the best estimator by the Gauss Markov theorem, even without the assumption of Gaussianity. However, the BLUE as given in (2.9) has a limitation: it is infeasible, as $\sigma_0^2\mathbf{\Sigma}_0$ is unknown in practice.

A large class of estimators for the mean is the M–estimator, introduced by Beran (1991), and defined as the solution of

$$\sum_{t=1}^{n} \psi\left(y_t - \widehat{\mu}\right) = 0,$$

where $\psi\left(.\right)$ is a deterministic function. Special cases of M–estimators are the sample mean and the median. The M–estimators are shown to have the same asymptotic variance as the sample mean, in particular for Gaussian long memory processes. In the case of independent and short range dependent processes, M–estimators have a different asymptotic variance from (2.8).

2.3 Parametric estimation of the dynamic parameters

Parametric estimation of the dynamic parameters is performed either in the time domain or in the frequency domain. Time domain estimation is perhaps the more conventional method. It includes methods such as TML estimation and the CSS estimation method. Time domain estimation involves the direct use of the observed data in the specification of the criterion function, whereas frequency domain estimation (involving spectral analysis) is a two-step procedure. In the first step, one applies the Fourier transformation to the data and then, in the second step, the parameters are estimated by using the transformed data. This includes methods such as FML and EWH.

Several parametric estimators have been established in the literature. We shall define these estimators and discuss their limiting properties under both the correct and incorrect specification of the model, in Sections 2.3.1 and 2.3.2, respectively.

2.3.1 Correct specification of the model

Assume that $\{y_t\}$ is generated from an ARFIMA(p_0, d_0, q_0) process with spectral density given by

$$f_0(\lambda) = \frac{\sigma_0^2}{2\pi} \left[\sin(\lambda/2)\right]^{-2d_0} \frac{\left|\theta_0(\exp(\imath\lambda))\right|^2}{\left|\phi_0(\exp(\imath\lambda))\right|^2}, \quad -\pi \leq \lambda \leq \pi, \tag{2.10}$$

where $\theta_0(z) = 1 + \theta_{10}z + ... + \theta_{q_00}z^{q_0}$ and $\phi(z) = 1 + \phi_{10}z + ... + \phi_{p_00}z^{p_0}$. Under correct specification, the model to be estimated is the ARFIMA(p, d, q) model, where $p = p_0$ and $q = q_0$. Denote by $\boldsymbol{\vartheta}$ the $(p_0 + q_0 + 1) \times 1$ vector of dynamic parameters in the model.

The TML estimator of $\boldsymbol{\vartheta}$ is defined by maximizing the exact Gaussian log-likelihood function

$$\ell\left(\boldsymbol{\vartheta}, \mu, \sigma^2\right) = -\frac{1}{2n}\log\left(2\pi\sigma^2\right) - \frac{1}{2n}\log\left|\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}\right| - \frac{1}{2n\sigma^2}\left(\mathbf{y} - \mu\mathbf{l}\right)^\top \boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}^{-1}\left(\mathbf{y} - \mu\mathbf{l}\right), \tag{2.11}$$

over the parameter space $\mathbb{E}$ as defined in Section 2.1. Here, $\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}$ is the autocovariance of the fitted model. This form of Gaussian log-likelihood function is directly analyzed in Dahlhaus (1989) and Lieberman *et al.* (2012). One can replace μ by the true mean, if it is known. If it is unknown, one can simply replace it by the sample mean or the BLUE defined in the previous section, where the BLUE is itself a function of $\boldsymbol{\vartheta}$.

An alternative to TML is the CSS method proposed by Chung and Baillie (1993), Beran (1995) and Robinson (2006). The CSS estimator is defined by minimizing the function

$$S(\boldsymbol{\vartheta}) = \frac{1}{n}\sum_{t=1}^{n} e_t^2, \tag{2.12}$$

where e_t is the approximation of ε_t in (2.3) for $t = 1, \ldots, n$, given by

$$e_t = \sum_{i=0}^{t-1} \tau_i(\boldsymbol{\vartheta})(y_{t-i} - \mu),$$

where $\tau_i(\boldsymbol{\vartheta})$ is as defined in (2.4). Here too μ can be treated as mentioned for the TML procedure.[2]

Under certain regularity conditions, Grenander and Szego (1958) showed that

$$\frac{1}{n}\log|\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}| \to \frac{1}{2\pi}\int_{-\pi}^{\pi}\log f(\lambda)\, d\lambda, \quad \text{as} \quad n \to \infty. \tag{2.13}$$

Beran (1994, Lemma 5.3) states that the matrix $\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}^{-1}$ can be approximated by the matrix $\mathbf{A} = [\alpha(i-k)], i, j = 1, \ldots, n$, where

$$\alpha(i-k) = \int_{-\pi}^{\pi} \frac{1}{f(\lambda)} \exp[\imath(i-k)\lambda]\, d\lambda. \tag{2.14}$$

Therefore, the approximation in (2.14) immediately gives,

$$\begin{aligned}
\frac{1}{n}(\mathbf{y}-\mu\mathbf{l})^{\top}\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}^{-1}(\mathbf{y}-\mu\mathbf{l}) &\approx \frac{1}{n}(\mathbf{y}-\mu\mathbf{l})^{\top}\mathbf{A}(\mathbf{y}-\mu\mathbf{l}) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n} a(i-k)(y_i-\mu)(y_k-\mu) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}\int_{-\pi}^{\pi}\frac{1}{f(\lambda)}\exp[\imath(i-k)\lambda]\, d\lambda\,(y_i-\mu)(y_k-\mu) \\
&= \frac{1}{n}\int_{-\pi}^{\pi}\frac{1}{f(\lambda)}\sum_{i=1}^{n}\sum_{k=1}^{n}\exp[\imath(i-k)\lambda](y_i-\mu)(y_k-\mu)\, d\lambda
\end{aligned}$$

[2] Another estimation procedure that involves estimating the coefficients of the AR representation, is the Yule-Walker method, which yields closed-form solutions for the coefficients of AR models (see Box and Jenkins, 1970, pages 58-59). However, this method is not feasible for fractional process as the AR representation has infinite number of coefficients and truncation of the infinite series may not be appropriate due to the slow convergence of $\tau_i(\boldsymbol{\eta})$.

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{I(\lambda,\mu)}{f(\lambda)}d\lambda, \tag{2.15}$$

where $I(\lambda,\mu)$ is the periodogram defined as

$$I(\lambda,\mu) = |D(\lambda,\mu)|^2; \quad D(\lambda,\mu) = \frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n}(y_t-\mu)\exp(-\imath\lambda t). \tag{2.16}$$

The approximations in (2.13) and (2.15) lead to the approximation of the log-likelihood in the frequency domain (leaving the constant term in (2.11) out) as

$$-\frac{1}{4\pi}\int_{-\pi}^{\pi}\log f(\lambda)\,d\lambda - \frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{I(\lambda,\mu)}{f(\lambda)}d\lambda, \tag{2.17}$$

where $\imath = \sqrt{-1}$ is the imaginary unit. The function in (2.17) is the exact Whittle log-likelihood that is introduced by Whittle (1952) for the class of short memory models. Later, Dahlhaus (1989) adopted the technique for long memory models[3]. In this thesis we refer to the objective function in (2.17) as it is the representation of the exact Gaussian log-likelihood function in the frequency domain. Thus, the resultant estimator that maximizes (2.17) is defined as the EWH estimator of $\boldsymbol{\vartheta}$.

Replacing the integral in (2.17) by discrete frequencies measured at $\lambda_j = 2\pi j/n$, for $j = 1,\ldots,\lfloor n/2\rfloor$, gives the discrete version of the exact Whittle log-likelihood as,

$$-\frac{1}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\log f(\lambda_j) - \frac{1}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j,\mu)}{f(\lambda_j)}. \tag{2.18}$$

The maximizer of (2.18) is denoted as the DWH estimator of $\boldsymbol{\vartheta}$. Alternatively, Cheung and Diebold (1994) exploit the fact that the first component in the above expression is negligible for large n – as its integral form is exactly zero – and introduce another approximation to (2.11) in the form of

$$-\frac{1}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j,\mu)}{f(\lambda_j)}. \tag{2.19}$$

[3] Fox and Taqqu (1986) analyzed the limiting properties of the exact Whittle log-likelihood objective function without the first component in (2.17).

Denote by the FML estimator of $\boldsymbol{\vartheta}$, the maximizer of the above function.

Remark 2.1 *The FML and DWH objective functions are not functions of μ, as the component, $\sum_{t=1}^{n} \mu \exp\left(-\imath\lambda_j t\right)$ of the DFT given in (4.13) produces zero for $\lambda_j = 2\pi j/n$, $j = 1, \ldots, \lfloor n/2 \rfloor$. Thus, it is obvious that the properties of the FML and DWH estimators of $\boldsymbol{\vartheta}$ do not alter due to the mean being known or estimated. However, the EWH log-likelihood function is a function of $\boldsymbol{\vartheta}$ and μ.*

Remark 2.2 *Various other versions of Whittle estimators have also been analyzed in the time series analysis literature. For example, refer to Whittle (1957) and Whittle (1962). However, in this thesis we use the three forms given in (2.17), (2.18) and (2.19) that are commonly used particularly in the context of fractionally integrated models.*

Given the correct specification of the model, with either choice of mean estimators (that is, the sample mean or the BLUE), the asymptotic properties of the TML, CSS, EWH and DWH estimators will be the same as in the known zero mean case. From the theory established in the literature, it follows that in all cases the estimator, $\widehat{\boldsymbol{\vartheta}}$, is a consistent estimator of $\boldsymbol{\vartheta}_0$, the true value of $\boldsymbol{\vartheta}$, and that $\sqrt{n}\left(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)$ is asymptotically normal with mean zero. The exact form of the limiting covariance matrix is given by $\mathbf{V}^{-1}\left(\boldsymbol{\vartheta}_0\right)$, where, $\mathbf{V}\left(\boldsymbol{\vartheta}_0\right) = \left[v_{ij}\left(\boldsymbol{\vartheta}_0\right)\right], i, j = 1, \ldots, n$, with

$$v_{ij}\left(\boldsymbol{\vartheta}_0\right) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{\partial}{\partial\vartheta_i} \log f\left(\lambda\right) \frac{\partial}{\partial\vartheta_j} \log f\left(\lambda\right)\bigg|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} d\lambda.$$

Further, the above mentioned standard asymptotic theory has been shown to hold for non-Gaussian fractional time series under appropriate moment restrictions on the innovation process (see, Hualde and Robinson, 2011 and Cavaliere *et al.*, 2017).

Although these methods are asymptotically equivalent, they show significant difference in their finite sample performance. The comprehensive simulation exercise of Nielsen and

Frederiksen (2005) for Gaussian fractional time series, illustrates that when estimating the fractional differencing parameter, the TML estimator outperforms the others, in terms of bias and RMSE, provided that the process mean is known. When the mean is estimated by BLUE, then DWH performs better than the others.

Among the five objective functions, the frequency domain methods in (2.17), (2.18) and (2.19) are preferred on grounds of computational ease, as the evaluation of the quantities $\mathbf{\Sigma}_{\boldsymbol{\vartheta}}^{-1}$ and $\tau_i(\boldsymbol{\vartheta})$ in (2.11) and (2.12) is time consuming, especially when $d > 0$. Doornik and Ooms (2003) proposed an algorithm to calculate the elements of $\mathbf{\Sigma}_{\boldsymbol{\vartheta}}^{-1}$. Although their method reduces the computational time to some extent, the frequency domain methods remain the simplest from a computational point of view.

2.3.2 Mis-specification of the model

A fundamental assumption underlying the classical results on the statistical properties of the above estimators is that the model is correctly specified for the process. If one does not assume the correct specification of the model, the next important question is: Does the parametric estimator still converge to some limit? If the estimator does converge to some limit: (1) Does this limit have any meaning?, (2) Does the estimator achieve the usual $\sqrt{n}$ rate of convergence? and, (3) Is the limiting distribution still normal? Specific answers to each of these questions are provided by White (1982) for short range processes. See also Berk (1966, 1970) and Huber, 1967 for earlier work on mis-specification, particularly aimed at addressing the question of consistency.

Under mis-specification, the TDGP and the fitted (mis-specified) model are different. Estimating the parameters of the incorrect model proceeds as described for the correct specification, in Section 2.3.1, except that form of the relevant objective functions will depict the incorrectly specified model, not the TDGP. More details on the objective functions under mis-specification will be provided in Chapter 3.

Only two published papers address the above questions in the context of fractional models. Yajima (1993) analyzes the consequences of fitting an ARMA(p, q) model to a stationary ARFIMA(p_0, d_0, q_0) TDGP where $\{p \neq p_0 \cup q \neq q_0\} \setminus \{p_0 \leq p \cap q_0 \leq q\}$. Yajima shows that under mis-specification, the EWH estimator of $\boldsymbol{\beta}$ – notation we use here to denote the vector of ARMA coefficients – and the error variance, converges to the pseudo-true value $\left(\boldsymbol{\beta}_1^\top, \sigma_1^2\right)^\top$. Further, when the true value of d, d_0, lies in the range $d_0 \in (0.25, 0.5)$, the rate of convergence is slower than $\sqrt{n}$ and it depends on d_0. The limiting distribution is of the form of a Rosenblatt process. When $d_0 = 0.25$, asymptotic normality is achieved with a rate of convergence different from $\sqrt{n}$. Whenever $d_0 \in (0, 0.25)$, both asymptotic normality and the $\sqrt{n}$ rate of convergence are achieved.

As was mentioned in Section 1, Chen and Deo (2006) developed the asymptotic theory for fitting an ARFIMA(p, d, q) model given in (2.5) to a stationary ARFIMA(p_0, d_0, q_0) TDGP. Denote by $\boldsymbol{\eta} = \left(d, \boldsymbol{\beta}^\top\right)^\top$ the $(p + q + 1) \times 1$ vector of dynamic parameters in the fitted model such that $\{p \neq p_0 \cup q \neq q_0\} \setminus \{p_0 \leq p \cap q_0 \leq q\}$. It is shown that the FML estimator of $\boldsymbol{\eta}$ converges to the pseudo-true value $\boldsymbol{\eta}_1$. The rate of convergence and the form limiting distribution are case-specific depending on the degree of mis-specification, d^*, measured by the difference between the true and the pseudo-true values of d (that is, $d^* = d_0 - d_1$). If $d^* > 0.25$, the asymptotic distribution is not normal. In this case, the rate of convergence is slower than $\sqrt{n}$ and the rate declines as the degree of mis-specification increases. If $d^* = 0.25$, the limiting distribution is normal with a rate of convergence different from $\sqrt{n}$. If $d^* < 0.25$, $\sqrt{n}-$ consistency and asymptotic normality are achieved. The exact forms of the rate of convergence and the limiting distributions will be provided in Chapter 3. Chen and Deo focus only on the limiting behaviour of the FML estimator of $\boldsymbol{\eta}$, and they do not consider the properties of the estimator of the error variance, as did Yajima (1993). Neither Yajima nor Chen and Deo evaluate the finite sampling properties of any estimators under mis-specification.

The other closely related article addressing issues related to mis-specification in a long memory setting is Crato and Taylor (1996). They discuss the consequences of mis-specification in three different cases in terms of forecasting errors: (1) fitting an autoregressive integrated moving average [ARIMA] model to a non-stationary ARFIMA process, (2) fitting a non-stationary ARFIMA model to a stationary ARFIMA process, and, (3) mis-specifying a fractional noise process with an ARIMA model. In all three cases, the error variance of the $k-$ step ahead forecast is large under mis-specification. Further, the error variance is much larger when $d_0 > 1$. The error variance also increases as k (the number of steps ahead for forecasting) increases.

2.4 Semi-parametric estimation of d: Local methods

As mentioned in Chapter 1, with the semi-parametric approach, a full parametric model is not specified for the spectral density of the process. The literature on semi-parametric estimation can be viewed in terms of global methods and local methods. The so-called global (or broadband) method yields an estimator of d by constructing an estimator of the spectral density over the whole range $[-\pi, \pi]$. Examples of such global estimators are the broadband log-periodogram regression estimator of Moulines and Soulier (1999), Hurvich and Brodsky (2001), the fractional exponential estimator of Robinson (2004) and the fractional autoregressive estimator of Bhansali *et al.* (2006). Besides estimating d, these articles provide a non-parametric estimator of the spectral density f. For more details on global methods, see, Doukhan *et al.* (2003, pages 251-301).

In the local (or narrowband) methods, an estimator of d is obtained without prior knowledge about the spectral density outside an arbitrarily small neighbourhood of the origin, and therefore with no strong assumptions on $g(\boldsymbol{\beta}, \lambda)$ in (2.6) (hereinafter referred to as $g(\lambda)$ in this chapter for notational simplicity). Local estimators use only periodogram ordinates belonging to this small neighbourhood around zero. This class of estimator was initiated by Geweke and Porter-Hudak (1983) and Robinson (1995a). The local semi-parametric estimators developed thereinafter, such as those of Robinson (1995b), Shimotsu and Phillips (2002), Andrews and Guggenberger (2003) and Andrews and Sun (2004), are extensions of, or improvements on, these two initial methods.

2.4.1 LPR estimator of Geweke and Porter-Hudak (1983)

The approach of Geweke and Porter-Hudak (1983) is motivated by the simple linear regression model,

$$\log I(\lambda_j) = (\log g(0) - C) - 2d \log(2\sin(\lambda_j/2)) + \xi_j, \tag{2.20}$$

where C is the Euler constant, $I(\lambda)$ is the periodogram of the vector of realizations, $\mathbf{y}$, measured at Fourier frequencies, $\lambda_j = 2\pi j/n$; $(j = 1, 2..., N_n)$. Here the error term

$$\xi_j = \log\left(I(\lambda_j)/f(\lambda_j)\right) + C + V_j,$$

where

$$V_j = \log\left(g(\lambda_j)/g(0)\right), \tag{2.21}$$

and the error terms are assumed to be asymptotically independently and identically distributed (*i.i.d.*). The LPR estimator of d is simply the ordinary least squares [OLS] estimator of the slope parameter in (2.20) and is given by

$$\widehat{d}_n = \frac{-0.5\sum_{j=1}^{N_n}\left(x_j - \overline{x}\right) z_j}{\sum_{j=1}^{N_n}(x_j - \overline{x})^2}, \tag{2.22}$$

where $z_j = \log I\left(\lambda_j\right)$, $x_j = \log\left(2\sin\left(\lambda_j/2\right)\right)$, and $\overline{x} = \frac{1}{N_n}\sum_{j=1}^{N_n} x_j$. The bandwidth N_n is chosen such that it tends to zero as $n \to \infty$, but at a slower rate than n.

For stationary Gaussian fractional processes (with this assumption of Gaussianity later relaxed by Velasco, 2000), the limiting properties of the LPR estimator have been derived by Geweke and Porter-Hudak (1983) for the case $d_0 < 0$. Later, Robinson (1995b) established the LPR properties for $-0.5 < d_0 < 0.5$. It is shown that the LPR estimator is $\sqrt{N_n}$ – consistent for d_0, at slower rate than the $\sqrt{n}$ rate that is achieved by parametric estimators under correct specification of the model. Further, $\sqrt{N_n}\left(\widehat{d}_n - d_0\right)$ is asymptotically normal with zero mean and asymptotic variance $\pi^2/24$. Velasco (1999b) and Kim and Phillips (2006) show that consistency holds for the range of $d_0 \in (-0.5, 1]$ and the asymptotic normality is valid for $d_0 \in (-0.5, 0.75]$. Hurvich *et al.* (1998) provide asymptotic expressions for the bias, variance and mean squared error [MSE]. Hurvich *et al.* suggest choosing the bandwidth,

$$N_n^{Opt} = \left(\frac{27}{128\pi^2}\right)^{1/5}\left(\frac{g(0)}{g''(0)}\right)^{2/5} n^{4/5},$$

for the LPR estimator to achieve minimum MSE. We will revisit the LPR estimator in Chapter 5 where we provide both the regularity conditions under which the above mentioned asymptotic properties hold, and the bias and variance expressions. Hence, we omit those details here.

Simulation exercises of Agiakloglou *et al.* (1993) and Nielsen and Frederiksen (2005) illustrate that the LPR estimators exhibit large finite sample bias, for example in the presence of strong autoregressive noise. A detailed finite sample performance of the estimator is documented in Chapter 5. In the next section, we shall review the bias-corrected LPR estimators that have been proposed in the literature.

2.4.2 Bias correction for the LPR estimator: Analytical methods

Andrews and Guggenberger (2003) suggest a bias-reduction method by replacing V_j in (2.21) with the first $2r$ number of terms in the Taylor's expansion of V_j around zero, and considering these $2r$ terms as additional regressors in the regression model in (2.20). The idea behind this approach is to use some approximation of the short memory dynamics of the process rather than ignoring the information completely. Thus, their approach uses a multivariate regression model defined as

$$\log I\left(\lambda_j\right) = (\log g(0) - C) - 2d\log(2\sin(\lambda_j/2)) + \sum_{k=1}^{r}\frac{b_{2k}}{(2k)!}\lambda_j^{2k} + \zeta_j,$$

where $\zeta_j = \xi_j - \sum_{k=1}^{r}\frac{b_{2k}}{(2k)!}\lambda_j^{2k}$ and $b_k = \frac{d^k \log g(0)}{d\lambda^k}$. The term V_j is a function of $\log g\left(\lambda_j\right)$, which is an even continuous function of sine components. Hence, all continuous odd derivatives are equal to zero at zero, and the expansion of V_j is expressed as an even-order polynomial in the frequency. Note that if $r = 0$ then the estimator is simply the LPR estimator.

The bias-reduced LPR estimator of Andrews and Guggenberger (2003) (which we denote by $\widehat{d}_r^{AG}$) is the least squares [LS] estimator of the coefficient on $-2\log(2\sin(\lambda_j/2))$. Given that g is smooth of order $s \geq 2 + 2r$ at $\lambda = 0$ for some non-negative integer r,

$$E\left(\widehat{d}_r^{AG}\right) \quad = \quad d_0 + \tau_r b_{2+2r}\frac{N_n^{2+2r}}{n^{2+2r}}\left(1 + o\left(1\right)\right) + O\left(\frac{N_n^q}{n^q}\right) + O\left(\frac{\ln^3 N_n}{N_n}\right), \qquad (2.23)$$

$$Var\left(\widehat{d}_r^{AG}\right) = \frac{\pi^2}{24N_n}c_r + o\left(\frac{1}{N_n}\right), \tag{2.24}$$

where $q = \min(s, 4+2r)$. Let $\boldsymbol{\nu}_r$ be a vector with k^{th} element $\nu_{r,k}$ and $\mathbf{Y}_r = [\varkappa_{i,k}], i, k = 1, \ldots, r$ such that

$$\nu_{r,k} = \frac{2k}{(2k+1)^2} \text{ and } \varkappa_{i,k} = \frac{4ik}{(2i+2k+1)(2i+1)(2k+1)}.$$

The values of c_r in (2.24) are defined as

$$c_r = \left[1 - \boldsymbol{\nu}_r^\top \mathbf{Y}_r^{-1} \boldsymbol{\nu}_r\right]^{-1}.$$

Suppose $\boldsymbol{\varsigma}_r$ is a vector with k^{th} element $\varsigma_{r,k}$ defined as

$$\varsigma_{r,k} = \frac{2k(3+2r)}{(2r+2k+3)(2k+1)}.$$

The values of τ_r in (2.23) are given by

$$\tau_r = -\frac{(2\pi)^{2+2r}(2+2r)c_r}{2(3+2r)!(3+2r)}\left[1 - \boldsymbol{\nu}_r^\top \mathbf{Y}_r^{-1} \boldsymbol{\varsigma}_r\right].$$

The asymptotic bias of the estimator is of order N_n^{2+2r}/n^{2+2r}, that is of smaller order than that of the LPR estimator. Under certain assumptions, the authors show that $\widehat{d}_r^{AG}$ is consistent for d_0 and that

$$\sqrt{N_n}\left(\widehat{d}_r^{AG} - d_0 - \tau_r b_{2+2r}\frac{N_n^{2+2r}}{n^{2+2r}}\right) \to^D N\left(0, \frac{\pi^2}{24}c_r\right),$$

where $c_r > 1$ for $r \geq 1$ and $c_r = 1$ for $r = 0$. The asymptotic variance of $\widehat{d}_r^{AG}$ is thus increased by a multiplicative constant, relative to that of the LPR estimator. The increase in asymptotic variance is due to the reduction in the asymptotic order of magnitude of the bias. However, the asymptotic variance is of order N_n^{-1}, the same as that of the LPR estimator. Andrews and Guggenberger recommend using small values of r, such as $r = 1$ or $r = 2$, in practice, for better finite sample performance, in terms of bias-reduction and MSE, although the asymptotic results hold for any value of r.

Guggenberger and Sun (2006) subsequently introduced a methodology to retain the bias reduction of the estimator of Andrews and Guggenberger (2003), for any given value of r, but with less variance inflation. They propose a weighted average of LPR estimators based on a different sets of discrete Fourier frequencies. We provide an extensive review of their estimator in Chapter 5. Thus, we omit details here.

Another variant of the LPR estimator is the pooled log-periodogram regression [PLPR] estimator of Shimotsu and Phillips (2002). The estimator is given by

$$\widehat{d}_n^{SP} = \frac{\sum_{j=0}^{L} \sum_{\{j:\lambda_s \in B_j\}} z_{sj} \left(x_{sj} - \overline{x}_{.j}\right)}{\sum_{j=0}^{L} \sum_{\{j:\lambda_s \in B_j\}} \left(x_{sj} - \overline{x}_{.j}\right)^2},$$

where $z_{sj} = \log I(\lambda_s)$, $x_{sj} = 2\log(2\sin(\lambda_s/2))$ for $\lambda_s \in B_j$, $\overline{x}_{.j} = \frac{1}{N_n}\sum_{\{j:\lambda_s \in B_j\}} x_{sj}$, and the B_j's are the frequency bands of width π/M such that

$$B_j = \begin{cases} \left\{\lambda_s | \omega_j - \frac{\pi}{2N_n} < \lambda_s \le \omega_j + \frac{\pi}{2N_n}\right\} & \omega_j = \frac{(2j+1)\pi}{2N_n},\ j = 1, \ldots, M-1 \\ \left\{\lambda_s | 0 < \lambda_s \le \frac{\pi}{N_n}\right\} & \omega_j = 0,\ j = 0, \end{cases},$$

with $M = n/(2N_n)$. The PLPR estimator is consistent for d_0, and $\sqrt{N_n}\left(\widehat{d}_n^{SP} - d_0\right)$ is asymptotically normal with zero mean and variance $\pi^2/24(1+\Xi)$ where Ξ is a positive constant. The asymptotic variance of the estimator is smaller than that of the LPR estimator, but at the cost of larger asymptotic bias. In a Monte-Carlo experiment based on ARFIMA$(1, d, 0)$ and ARFIMA$(0, d, 1)$ processes, Nielsen and Frederiksen (2005) show that PLPR estimator exhibits a slightly greater finite sample bias than the LPR estimator, with a smaller root mean squared error [RMSE]. A larger bandwidth leads to even larger bias and smaller RMSE.

Reisen (1994) suggests replacing the periodogram ordinates in (2.20) by a smoothed periodogram of the form

$$\widehat{f}_s(\lambda) = \frac{1}{2\pi} \sum_{s=-N_n}^{N_n} k\left(\frac{s}{N_n}\right) \widehat{\gamma}(s) \cos(s\lambda), \tag{2.25}$$

where $\widehat{\gamma}(s)$ is the sample auto-covariance at lag s, defined as

$$\widehat{\gamma}(s) = \frac{1}{n}\sum (y_t - \overline{y})(y_{t+s} - \overline{y}),$$

and $k(u)$ is the lag window generator, a fixed continuous even function in the range $-1 < u < 1$, with $k(0) = 1$ and $k(-u) = k(u)$. The author also suggests choosing the Parzen lag window generator. The resultant estimator $\widehat{d}_n^{\mathrm{Re}}$, is of the form of the LPR estimator after replacing $\log I(\lambda_j)$ in (2.22) by the logarithmic transformation of $\widehat{f}_s(\lambda)$ in (2.25). Although the author discusses the consistency and asymptotic normality of the estimator, for antipersistent ARFIMA processes, no rigorous proofs are provided. In the simulation experiment for ARFIMA$(1, d, 0)$ and ARFIMA$(0, d, 1)$ processes, the author shows that the estimator uniformly performs better than the LPR estimator in terms of bias and MSE. Alternative smoothed periodogram-based LPR estimators are discussed by Hassler (1993) and Chen *et al.* (1994), where the smoothed periodograms are defined as an empirical counterpart to the spectrum and as a lag-window estimator of the spectrum similar to (2.25), respectively. Their resultant LPR estimators show smaller bias and mean squared error [MSE] compared to the original LPR estimator. The simulation experiments of Chen *et al.* suggest that the Bartlett-Priestley lag window is a better choice than the Parzen lag window in terms of MSE.

2.4.3 Bias correction for the LPR estimator: Non-parametric methods

A prefiltered sieve bootstrap-based bias-corrected LPR estimator [PFSB] is introduced by Poskitt *et al.* (2016) for long memory processes. The algorithm to obtain the estimator is summarized as follows.

Step 1: Generate the prefiltered series, $\{w_t\}_{t=1}^n$, from the observed data using a pre-determined value of d, $d^f = \widehat{d}_n$:

$$w_t = \sum_{j=0}^{t-1} \frac{\Gamma\left(j - d^f\right)}{\Gamma\left(-d^f\right)\Gamma\left(j+1\right)} y_{t-j}.$$

Step 2: Fit a p^{th} order autoregressive model to $\{w_1, w_2, ..., w_n\}$. Denote by $\widehat{\mathbf{a}}\left(p\right) = \left[\widehat{a}_j\left(p\right)\right]$, $j = 1, \ldots, p$, the Yule-Walker autoregressive parameters, that is, $\widehat{\mathbf{a}}\left(p\right) = \widehat{\mathbf{\Gamma}}\left(p\right)^{-1}\widehat{\gamma}_p$, where

$$\widehat{\gamma}\left(h\right) = \frac{1}{n}\sum_{t=1}^{n-|h|}\left(w_t - \overline{w}\right)\left(w_{t+|h|} - \overline{w}\right), \text{ for } 0 \leq h \leq p,$$

$\overline{w} = \frac{1}{n}\sum_{t=1}^{n} w_t$, $\widehat{\mathbf{\Gamma}}\left(p\right) = \left[\widehat{\gamma}\left(i-j\right)\right]$; $i, j = 1, \ldots, p$ and $\widehat{\gamma}_p = \left[\widehat{\gamma}\left(1\right), \ldots, \widehat{\gamma}\left(p\right)\right]'$.

Step 3: Evaluate the residuals $\widetilde{\varepsilon}_t\left(p\right) = \sum_{j=0}^{p}\widehat{a}_j\left(p\right)w_{t-j}$, $t = 1, 2, \ldots, n$, using $w_{1-j} = w_{n-j+1}$. Calculate $\overline{\sigma}_{\widetilde{\varepsilon}}^2 = n^{-1}\sum_{t=1}^{n}\left(\widetilde{\varepsilon}_t\left(p\right) - \overline{\varepsilon}\right)^2$ where $\overline{\varepsilon} = n^{-1}\sum_{t=1}^{n}\widetilde{\varepsilon}_t\left(p\right)$. Then, set $\varepsilon_t^*\left(p\right) = \overline{\sigma}_{\widetilde{\varepsilon}}e_t$, $t = 1, \ldots, n$, where $\{e_t\}_{t=1}^{n}$ is a simple random sample of *i.i.d* standard normal random variables.

Step 4: Generate sieve bootstrap sample $\{w_1^*, w_2^*, ..., w_n^*\}$, where $w_t^* = \sum_{j=0}^{p}\widehat{a}_j\left(p\right)w_{t-j}^* = \varepsilon_t^*$, initiating at $w_{1-j}^* = w_{\tau-j+1}$, $j = 1, \ldots, p$, where $\tau \sim \textit{discrete uniform}\left[p, n\right]$.

Step 5: Generate pre-filtered sieve bootstrap sample, $\{y_t^*\}_{t=1}^{n}$ where $y_t^* = \sum_{j=0}^{t-1}\frac{\Gamma\left(j+d^f\right)}{\Gamma\left(d^f\right)\Gamma\left(j+1\right)}w_{t-j}^*$, and evaluate the LPR estimate of $\widehat{d}_{n,b}^*$, using bootstrap sample $\left\{y_{t,b}^*\right\}_{t=1}^{n}$.

Step 6: Repeat Steps $1-6$ for $B = 1000$ times and estimate the bias-corrected estimator, $\widehat{d}^{SBS} = \widehat{d}_n - b_{n,B}^*$, where $b_{n,B}^* = \frac{1}{B}\sum_{b=1}^{B}\widehat{d}_{n,b}^* - d^f$.

In addition to establishing the error rates for bootstrap–based estimation of the bias of the LPR estimator, and proving the theoretical validity of highest probability density confidence intervals constructed from the prefiltered bootstrap replications, the authors demonstrate numerically a significant reduction in bias via the prefiltering approach. More details on the finite sample performance of the resultant bias-corrected estimator will be provided in Chapter 5.

An alternative non-parametric approach to the bootstrap is the jackknife. The fundamental nature of jackknifing is to perform bias correction by combining the full sample and one or more sub-samples with some appropriate weights. There are two techniques available for drawing sub-samples; one is non-overlapping and the other is moving-block. For dependent data, the sub-samples are drawn in blocks to preserve the dependence structure of the full sample within the sub-samples too.

Let $\mathbf{y}_i$ denote the i^{th} sub-sample of the full sample $\mathbf{y}$. Denote by l the common sub-sample length and m the number of sub-samples such that $n = m \times l$. If the sub-samples are drawn using the 'non-overlapping' method, $\mathbf{y}_i^\top = \left(y_{(i-1)l+1}, \ldots, y_{il}\right)$ for $i = 1, \ldots, m$; alternatively if the sub-sampling scheme is 'moving-block' then $\mathbf{y}_i^\top = (y_i, \ldots, y_{i+l-1})$ for all i. See Figure 2.3 for a graphical illustration of both types of jackknife method. Suppose one is interested in bias-correcting the LPR estimator. Then the bias-corrected jackknife-based estimator is defined as follows,

$$\widehat{d}_{J,m} = w_n \widehat{d}_n - \sum_{i=1}^{m} w_i \widehat{d}_i,$$

where $\widehat{d}_n$ and $\widehat{d}_i$ are the full and sub-sample LPR estimators evaluated using $\mathbf{y}$ and $\mathbf{y}_i$ respectively. The weights are determined depending on the order of bias correction sought.

In Chapter 5 we make brief mention of earlier work in which the jackknife has been used to bias correct estimators in other time series settings. Ekonomi and Butka (2011) are the only authors, as far as we are aware, to use the technique in a long memory context, adopting the methodology of Chambers to bias-adjust the LPR estimator. A simulation exercise using

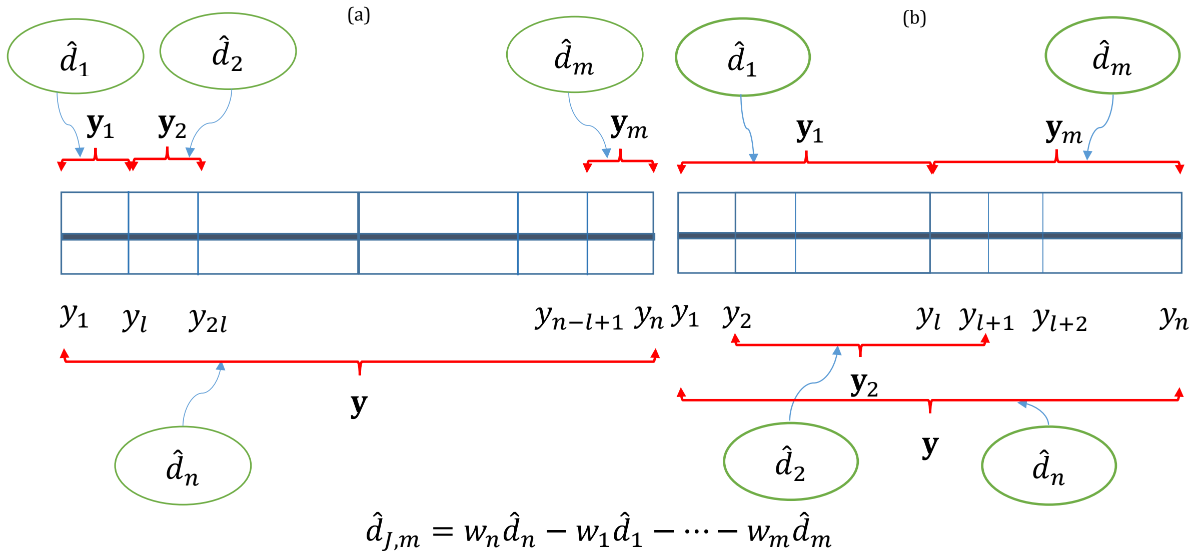

Figure 2.3: Graphical illustration of the jackknife method

an ARFIMA$(0, d, 0)$ process reveals that jackknife estimator helps to reduce the bias of the LPR estimator, relative to an alternative bias reduction method based on the block bootstrap. Their study does not, however, provide any theoretical results for their bias-corrected LPR estimator.

2.4.4 Other local semi-parametric estimators

The local Whittle (Gaussian semi-parametric or LW) estimator (abbreviated by $\widehat{d}_n^{LW}$) analyzed by Robinson (1995a), motivated by the local Whittle approach of Künsch (1987), is defined as the minimizer of the local Whittle likelihood function,

$$Q(G, d) = \frac{1}{N_n}\sum_{j=1}^{N_n}\left\{\log\left(G\lambda_j^{-2d}\right) + \frac{\lambda_j^{-2d}}{G} I(\lambda_j)\right\},$$

where G is some positive quantity that is also estimated simultaneously with d. Robinson (1995a) showed that the estimator is consistent for d_0 and $\sqrt{N_n}\left(\widehat{d}_n^{LW} - d_0\right) \to^D N(0, 1/4)$. These asymptotic results have subsequently been extended to non-stationary fractional processes by Velasco (1999a) and Phillips and Shimotsu (2004).

Andrews and Sun (2004) proposed a generalized version of the LW estimator that minimizes the local polynomial Whittle (LPW) log-likelihood function,

$$\begin{aligned} Q\left(G, d, a_1, \ldots, a_r\right) &= \frac{1}{N_n} \sum_{j=1}^{N_n} \left\{ \log\left(G \lambda_j^{-2d} \exp\left(- \sum_{k=1}^{r} a_k \lambda_j^{2k} \right) \right) \right. \\ &\quad \left. + \frac{\lambda_j^{-2d}}{G \exp\left(- \sum_{k=1}^{r} a_k \lambda_j^{2k} \right)} I\left(\lambda_j\right) \right\}. \end{aligned}$$

The LPW estimator (abbreviated by $\widehat{d}_n^{AS}$) is consistent for true d_0 and $\sqrt{N_n}\left(\widehat{d}_n^{AS} - d_0\right) \rightarrow^D N\left(0, 1/4c_r\right)$. The values of c_r are provided in the paper. Although the LPW estimator does give bias improvement, this comes at the cost of an increase in the asymptotic variance by a multiplicative constant c_r. To this end, Guggenberger and Sun (2006) proposed a weighted average of different LPW estimators with different bandwidths that achieves the same degree of bias reduction as the LPW estimator for any given r, but with less variance inflation.

Poskitt *et al.* (2016) adopted the prefiltered sieve bootstrap approach to bias correct the LW estimator. Their simulation experiments show notable bias reductions, particularly when the methodology is applied to the analytically bias-adjusted version of LW, LPW.

An exact local Whittle estimator was introduced by Shimotsu and Phillips (2005) and is defined by the minimizer of the function

$$Q\left(G, d\right) = \frac{1}{N_n} \sum_{j=1}^{N_n} \left\{ \log\left(G \lambda_j^{-2d} \right) + \frac{\lambda_j^{-2d}}{G} I_{\Delta^d y}\left(\lambda_j\right) \right\},$$

where $I_{\Delta^d y}\left(\lambda_j\right)$ is the periodogram of $\Delta^d y$ – the d difference series of $\{y_t\}_{t=1}^{n}$. The limiting properties of this estimator are the same as those of the LW estimator.

We finish this section with a series of remarks.

Remark 2.3 *The semi-parametric estimators are biased in the presence of the short run dynamics as the low frequencies are contaminated by the higher frequencies of the spectral density, particularly in the case of positive AR noise. Hence, the bias of the estimators increases as the AR noise becomes more persistent. However, MA noise contaminates the long run part less than AR noise does. as MA noise affects the short run part of the spectral density at the higher frequencies. This produces smaller bias regardless of the size of the MA parameter.*

Remark 2.4 *Semi-parametric estimators (in general) are governed by the bandwidth, N_n. It is a common practice to choose the bandwidth such that $N_n = \lfloor n^{\alpha} \rfloor$ where $\alpha = 0.5, 0.65$ or 0.7. Here $\lfloor x \rfloor$ denotes the integer part of x. Nielsen and Frederiksen (2005) state that when no short run dynamics are present in the data, it is preferable to use the larger bandwidth, and the opposite is typically the case in the presence of short run dynamics.*

Remark 2.5 *The following drawbacks of the semi-parametric estimators. i.e. that they: (i) have slower rates of convergence than the usual $\sqrt{n}$, (ii) exhibit large finite sample bias, and, (iii) produce estimates of only the fractional differencing parameter, with the short memory parameters not estimated explicitly, motivate us to investigate the convergence and distribution properties of parametric estimators under mis-specification of short memory dynamics; as these estimators achieve a $\sqrt{n}$ rate of convergence under certain conditions and do allow us to estimate the short run parameters.*

2.5 Conclusion

In this chapter, we have briefly discussed existing parametric techniques for estimating the mean, fractional differencing parameter and short memory dynamics of the ARFIMA model. In particular, we have reviewed the asymptotic properties of the parametric estimators in the case of both correct and incorrect specification of the model for a given DGP.

Semi-parametric estimation methods also play an important role in estimating the fractional differencing parameter. We have provided a brief review of the existing methods. Further, we reviewed bias-correction procedures for semi-parametric estimators. We also identified certain topics in the area of ARFIMA models that deserve further study. As was indicated in Chapter 1, the objective of the thesis is to make advances in the estimation of fractionally integrated models in some of these areas related to parametric estimation of dynamic parameters using mis-specified model and bias correction of a semi-parametric estimator.

Chapter 3

Issues in the estimation of mis-specified models of fractionally integrated processes

3.1 Introduction

Let $\{y_t\}$, $t \in \mathbb{Z}$, be a (strictly) stationary process with mean μ_0 and spectral density $f_0(\lambda)$, $\lambda \in [-\pi, \pi]$, that is such that

$$f_0(\lambda) \sim |\lambda|^{-2d_0} L_0(\lambda) \quad \text{as} \quad \lambda \to 0\,,$$

where $0 \leq |d_0| < 0.5$ and $L_0(\lambda)$ is a positive function that is slowly varying at 0. Prototypical examples of processes of this type are fractional noise, obtained as the increments of self-similar processes, and fractional autoregressive moving average processes. The process $\{y_t\}$ is said to exhibit long memory (or long range dependence) when $0 < d_0 < 0.5$, short memory (or short range dependence) when $d_0 = 0$, and antipersistence when $-0.5 < d_0 < 0$, and in this chapter we undertake an extensive examination of the consequences for estimation of such processes of mis-specifying the short run dynamics.[1] In so doing we provide a significant extension of earlier work on this particular form of mis-specification in Yajima (1993) and Chen and Deo (2006), as well as complementing work that focuses on other types

[1] This chapter is virtually identical to a paper submitted to *Journal of Econometrics,* which is now under first-round revision for that journal. The paper is jointly authored by Gael M Martin, Kanchana Nadarajah and Donald S Poskitt. Some repetition of material appearing other chapters has been retained, so as to minimize the extent to which the paper has been modified.

of mis-specification in fractional settings, as in Hassler (1994) and Crato and Taylor (1996). Our work also complements that of Robinson (2014), where mis-specification of the local-to zero characterization of long memory is examined, and that in Cavaliere *et al.* (2017), where a comprehensive treatment of inference in fractional models under very general forms of heteroscedasticity is provided. Whilst mis-specification *per se* is not the focus of the latter paper, the proof of convergence to a pseudo-true parameter of the conditional sum of squares [CSS] estimator under the imposition of incorrect linear restrictions bears some relationship with our more general results on mis-specified estimators in the fractional setting. Our results also generalize the existing literature on the properties of various parametric estimators - including their asymptotic equivalence - in *correctly specified* long memory models; see Fox and Taqqu (1986), Dahlhaus (1989), Giraitis and Surgailis (1990), Sowell (1992), Beran (1995), Robinson (2006) and Hualde and Robinson (2011), among others.

We begin by showing that four alternative parametric techniques – frequency domain maximum likelihood [FML], Whittle, time domain maximum likelihood [TML] and CSS – converge to a common pseudo-true parameter value when the short memory component is mis-specified.[2] Convergence is established for all three forms of dependence in the true data generating process [TDGP] - long memory, short memory and antipersistence. We establish convergence by demonstrating that when the mis-specified model is evaluated at points in the parameter space where the fractional index d exceeds $d_0 - 0.5$ the FML criterion function has a deterministic limit, but that the FML criterion function is *divergent* otherwise. The difference in the behaviour of the FML criterion function on subsets of the parameter space implies that the objective function does not behave uniformly. (See Robinson, 1995a; Hualde and Robin-

[2]Given that each of these estimators can be derived from a Gaussian likelihood, but we do not presuppose Gaussianity, each could be designated as a 'quasi' maximum likelihood estimator in the usual way; however for the sake of notational simplicity we avoid this qualifying term.

son, 2011; Cavaliere *et al.*, 2017, for related discussion.) This lack of uniformity makes proofs of convergence across the whole parameter space more complex than usual, but solutions presented in the previously cited references can be tailored to the current situation. We then show that under common mis-specification the criterion functions that define all three alternative estimators behave in a manner similar to that of the FML criterion. All four estimators are, accordingly, shown to converge to the same pseudo-true parameter value – by definition the common value that optimizes all four limiting objective functions.

Secondly, we derive closed-form representations for the first-order conditions that define the pseudo-true parameters for completely general autoregressive fractionally integrated moving average [ARFIMA] model structures – both true and mis-specified. This represents a substantial extension of the analysis in Chen and Deo (2006), in which the FML estimator under mis-specification was first investigated, but with expressions for the relevant first-order conditions provided for certain special specifications only, and with convergence established solely for *long* memory Gaussian processes.

Thirdly, we extend the asymptotic theory established by Chen and Deo (2006) for the FML estimator in the long memory Gaussian process case to the other three estimators, under long memory, short memory and antipersistence for both the TDGP and the estimated model, and without the imposition of Gaussianity. We show that all four methods are asymptotically equivalent in that they converge in distribution under common mis-specification. The convergence rate and nature of the asymptotic distribution is determined by the deviation of the pseudo-true value of the fractional index, d_1 say, from the true value, d_0, with three critical ranges for $d^* = d_0 - d_1 < 0.5$ given by $d^* > 0.25$, $d^* = 0.25$ and $d^* < 0.25$. This nonstandard distributional behaviour for all four parametric estimators introduces a further degree of complexity into the analysis, and contrasts sharply with earlier results established in correctly specified models, where separate modeling of the short run and fractional dynamics results in the asymptotic distribution of the parameter estimates being normal and free of the fractional indices.

A fourth contribution of this chapter is the provision of a closed-form representation of the (common) asymptotic distribution that obtains under the most extreme type of mis-specification – whereby both a $\sqrt{n}$ rate of convergence and limiting Gaussianity is lost – together with a demonstration of how to implement the distribution numerically using appropriate truncation of the series expansion that characterizes the distribution. This then enables us to illustrate graphically the differences in the rates at which the finite sample distributions of the four different estimators approach the (common) asymptotic distribution. Notably, for $d^* \geq 0.25$, there is a distinct grouping into frequency domain and time domain techniques; with the latter tending to replicate the asymptotic distribution more closely than the former in small samples.

Finally, we perform an extensive set of simulation experiments in which the relative finite sample performance of all four mis-specified estimators is assessed. The experiments are first conducted assuming a known (zero) mean, in line with the theoretical derivations in the chapter, and then re-run with the mean estimated. The ranking of the estimators, in terms of bias and MSE, is shown to depend heavily on whether the mean is specified or estimated, a conclusion that parallels results documented previously for correctly specified ARFIMA models (see, for example, Sowell, 1992; Cheung and Diebold, 1994 and Nielsen and Frederiksen, 2005).

We note that in defining the Whittle estimator we focus on a particular approximation to the frequency domain Gaussian (negative) log likelihood, in which *sums* over Fourier frequencies are used to approximate the relevant integrals. Despite the analytical equivalence of this estimator with the FML estimator for large n, the small sample performances of the two procedures will be seen to differ systematically. In common with the FML approach, this form of Whittle estimator is invariant to the mean of the process. For interest, we also present selected numerical results pertaining to the integral-based form of the Whittle estimator (referred to hereafter as 'exact Whittle'), both for the known mean case and when the mean is unknown, the lack of invariance of this estimator to the mean rendering this latter exercise of particular interest.

The chapter is organized as follows. In Section 3.2 we define the estimation problem, namely producing an estimate of the parameters of a fractionally integrated model when the component of the model that characterizes the short term dynamics is mis-specified. The criterion functions that define the Whittle, TML and CSS estimators, as well as the FML estimator, are specified, and we demonstrate that all four estimators possess a common probability limit under mis-specification. The limiting form of the criterion function for a mis-specified ARFIMA model is presented in Section 3.3, under complete generality for the short memory dynamics in the true process and estimated model, and closed-form expressions for the first-order conditions that define the pseudo-true values of the parameters are then given. The asymptotic equivalence of all four estimation methods is proved in Section 3.4. The finite sample performance of the alternative estimators of d in the mis-specified model – with reference to estimating the pseudo-true value of d – is documented in Section 3.5. The form of the sampling distribution is recorded, as is the bias and mean squared error [MSE], under different degrees of mis-specification, for all four estimators. Bias and MSE results are also documented for the exact Whittle estimator. The experiments are first conducted assuming a known (zero) mean, in line with the theoretical derivations in the chapter. In this case, the CSS estimator exhibits superior performance, in terms of bias and mean squared error, across a range of mis-specification settings, whilst the performance of the FML estimator is notably inferior. We then re-run the simulations using demeaned data. Only the time domain estimators, plus the exact Whittle estimator, are affected by this change, and with the rate of convergence of the sample mean being slow under long memory we find that the superiority of the time domain estimators is diminished - the (sums-based) Whittle estimator now being the best performer overall. The chapter concludes in Section 3.6 with a brief summary and some discussion of several issues that arise from the work. The proofs of the results presented in the chapter are assembled in Appendix 3.A. Appendix 3.B contains certain technical derivations referenced in the text.

3.2 Estimation under mis-specification of the short run dynamics

Assume that $\{y_t\}$ is generated from a TDGP that is a purely-nondeterministic stationary and ergodic process with spectral density given by

$$\frac{\sigma_0^2}{2\pi} f_0(\lambda) = \frac{\sigma_0^2}{2\pi} g_0\left(\lambda\right)\left(2\sin(\lambda/2)\right)^{-2d_0}, \tag{3.1}$$

where σ_0^2 is the innovation variance, $g_0\left(\lambda\right)$ is a real valued symmetric function of λ defined on $[-\pi, \pi]$ that is bounded above and bounded away from zero, and $-0.5 < d_0 < 0.5$. Then there exists a zero mean process $\{\varepsilon_t\}$ of uncorrelated random variables with variance σ_0^2 such that $\{y_t\}$ has the moving average representation

$$y_t = \mu_0 + \sum_{j=0}^{\infty} b_{0j}\varepsilon_{t-j}\,, \quad t \in \mathbb{Z} = 0, \pm 1, \ldots, \tag{3.2}$$

where $\{b_{0j}\}$ is a sequence of constants satisfying $b_{00} = 1$ and $\sum_{j=0}^{\infty} b_{0j}^2 < \infty$, and $f_0(\lambda) = |b_0(\exp(\imath\lambda))|^2$, $\lambda \in [-\pi, \pi]$, with $(1-z)^{d_0} b_0(z) = c_0(z) = \sum_{j=0}^{\infty} c_{0j} z^j$ and $0 < |c_0(z)|$, $|z| \leq 1$. We will suppose that $c(\exp(\imath\lambda))$ is differentiable in λ for all $\lambda \neq 0$ with a derivative that is of order $O(|\lambda|^{-1})$ as $\lambda \to 0$,and that

$(A.1)$ For all $t \in \mathbb{Z}$ we have $E_0[\varepsilon_t|\mathbb{F}_{t-1}] = 0$ and $E_0[\varepsilon_t^2|\mathbb{F}_{t-1}] = \sigma_0^2$, a.s. where $\mathbb{F}_{t-1}$ in the conditional expectations is the sigma-field of events generated by ε_s, $s \leq t-1$. Here, and in what follows, the zero subscript denotes that the moments are defined with respect to the TDGP.

The conditions imposed on $c_0(z)$ imply that $g_0(\lambda)$ corresponds to the spectrum of an invertible short memory process that is bounded and bounded away from zero for all $\lambda \in [-\pi, \pi]$ and the TDGP satisfies *Conditions A* of Hannan (1973, page 131). Assumption $(A.1)$ was introduced into time series analysis by Hannan (1973) and has been employed by several authors in investigations of both short memory and fractional linear processes since. The assumption that $\{\varepsilon_t\}$ is a conditionally homoscedastic martingale difference process circumvents the need to assume independence or identical distributions for the innovations, but rules out heteroscedasticity (see Cavaliere *et al.*, 2017, pages 5-6).

The model to be estimated is a parametric specification for the spectral density of $\{y_t\}$ of the form

$$\frac{\sigma^2}{2\pi} f_1(\boldsymbol{\eta}, \lambda) = \frac{\sigma^2}{2\pi} g_1(\boldsymbol{\beta}, \lambda)\,(2\sin(\lambda/2))^{-2d}, \tag{3.3}$$

where $g_1(\boldsymbol{\beta},\lambda)$ is a real valued symmetric function of λ defined on $[-\pi, \pi]$. The parameter of interest will be taken as $\boldsymbol{\eta} = \left(d, \boldsymbol{\beta}^\top\right)^\top$, where $d \in (-0.5, 0.5)$ and $\boldsymbol{\beta} \in \mathbb{B}$, where $\mathbb{B}$ is an l-dimensional compact convex set in $\mathbb{R}^l$. The variance σ^2 will be viewed as a supplementary or nuisance parameter. The model is to be estimated from a realization y_t, $t = 1, \ldots, n$, of $\{y_t\}$ and, in order that the structure of the model should parallel the assumed properties of the TDGP, it will be assumed that the model is specified in such a way that:

$(A.2)$ For all $\boldsymbol{\beta} \in \mathbb{B}$, $\int_{-\pi}^{\pi} \log g_1(\boldsymbol{\beta}, \lambda) d\lambda = 0$, and $\boldsymbol{\beta} \neq \boldsymbol{\beta}'$ implies that $g_1(\boldsymbol{\beta}, \lambda) \neq g_1(\boldsymbol{\beta}', \lambda)$ on a set of positive Lebesgue measure.

$(A.3)$ The function $g_1(\boldsymbol{\beta}, \lambda)$ is differentiable with respect to λ, with derivative $\partial g_1(\boldsymbol{\beta}, \lambda)/\partial\lambda$ continuous at all $(\boldsymbol{\beta}, \lambda)$, $\lambda \neq 0$, and $|\partial g_1(\boldsymbol{\beta}, \lambda)/\partial\lambda| = O(|\lambda|^{-1})$ as $\lambda \to 0$. Furthermore, $\inf_{\boldsymbol{\beta}} \inf_{\lambda} g_1(\boldsymbol{\beta}, \lambda) > 0$ and $\sup_{\boldsymbol{\beta}} \sup_{\lambda} g_1(\boldsymbol{\beta}, \lambda) < \infty$.

If there exists a subset of $[-\pi, \pi]$ with non-zero Lebesgue measure in which $g_1(\boldsymbol{\beta},\lambda) \neq g_0(\lambda)$ for all $\boldsymbol{\beta} \in \mathbb{B}$ then the model will be referred to as a mis-specified model (MisM).

The above TDGP and modelling assumptions encompass the standard parametric models, such as fractional noise, and fractional exponential and ARFIMA processes. (A detailed outline of the properties of such processes is provided in Beran, 1994.) We will return to a discussion of these regularity conditions later, where a strengthening of these conditions – detailed below – will be required in order to derive our asymptotic distribution theory. Meanwhile we note (for future reference) that an ARFIMA model for a time series $\{y_t\}$ may be defined as follows,

$$\phi(L)(1-L)^d\{y_t - \mu\} = \theta(L)\varepsilon_t, \tag{3.4}$$

where $\mu = E(y_t)$, L is the lag operator such that $L^k y_t = y_{t-k}$, and $\phi(z) = 1 + \phi_1 z + ... + \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + ... + \theta_q z^q$ are the autoregressive and moving average operators respectively, where it is assumed that $\phi(z)$ and $\theta(z)$ have no common roots and that the roots lie outside the unit circle. The errors $\{\varepsilon_t\}$ are assumed to be a white noise sequence with finite variance $\sigma^2 > 0$. For $|d| < 0.5$, $\{y_t\}$ can be represented as an infinite-order moving average of $\{\varepsilon_t\}$ with square-summable coefficients and, hence, on the assumption that the specification in 3.4 is correct, $\{y_t\}$ is defined as the limit in mean square of a covariance-stationary process. When $0 < d < 0.5$ neither the moving average coefficients nor the autocovariances of the process are absolutely summable, declining at a hyperbolic rate rather than the exponential rate typical of an ARMA process, with the term 'long memory' invoked accordingly. Thus, for an ARFIMA model we have $g_1(\boldsymbol{\beta}, \lambda) = |\theta(e^{i\lambda})|^2/|\phi(e^{i\lambda})|^2$ where $\boldsymbol{\beta} = (\phi_1, \phi_2, ..., \phi_p, \theta_1, \theta_2, ..., \theta_q)^\top$ and an $ARFIMA(p, d, q)$ model will be mis-specified if the realizations are generated from a true $ARFIMA(p_0, d_0, q_0)$ process and any of $\{p \neq p_0 \cup q \neq q_0\} \setminus \{p_0 \leq p \cap q_0 \leq q\}$ obtain.

We consider estimators of the parameter of interest, $\boldsymbol{\eta} = \left(d, \boldsymbol{\beta}^\top\right)^\top$, that are obtained by minimizing a criterion function $Q_n(\boldsymbol{\eta})$ over a user-assigned compact subset of the parameter space $(-0.5, 0.5) \times \mathbb{B}$,

$$\mathbb{E}_\delta = \mathbb{D}_\delta \times \mathbb{B} \quad \text{where} \quad \mathbb{D}_\delta = \{d : |d| \ \leq 0.5 - \delta\}, \text{ for some } 0 < \delta \ll 0.5 . \tag{3.5}$$

The bound on $|d|$ must be set by the practitioner via some criterion that reflects numerical precision. Under mis-specification the generic estimator, denoted by $\widehat{\boldsymbol{\eta}}_1$ for the time being, is obtained by minimizing $Q_n(\boldsymbol{\eta})$ assuming that $\{y_t\}$ follows the MisM.[3] In Section 3.2.1 we specify the form of $Q_n(\boldsymbol{\eta})$ associated with the FML estimator considered in Chen and Deo (2006) and outline its relationship with the criterion functions underlying two alternative versions of the frequency domain estimator introduced by Whittle, making it clear which form of Whittle estimator is the focus of our theoretical investigations. In Section 3.2.2 we define

[3] We follow the usual convention by denoting the estimator obtained under mis-specification as $\widehat{\boldsymbol{\eta}}_1$ rather than simply by $\widehat{\boldsymbol{\eta}}$, say. This is to make it explicit that the estimator is obtained under mis-specification and does not correspond to the estimator produced under the correct specification of the model, which could be denoted by $\widehat{\boldsymbol{\eta}}_0$.

the two time domain estimators that we consider here, TML and CSS, and their associated criterion functions.

Anticipating the convergence results that follow later in this section, for any given $Q_n(\boldsymbol{\eta})$ a law of large numbers can be combined with standard arguments to establish that on compact subsets of $\{\mathbb{D}_\delta \cap \{d : (d_0 - d) < 0.5\}\} \times \mathbb{B}$, i.e. subsets of $\mathbb{E}_\delta$ where $\mathbb{D}_\delta$ intersects with $\{d : (d_0 - d) < 0.5\}$, the criterion $Q_n(\boldsymbol{\eta})$ will converge uniformly to the non-stochastic limiting objective function

$$Q(\boldsymbol{\eta}) = \lim_{n\to\infty} E_0\left[Q_n(\boldsymbol{\eta})\right] = \frac{\sigma_0^2}{2\pi}\int_0^\pi \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta},\lambda)}d\lambda\,. \tag{3.6}$$

If, on the other hand, $Q_n(\boldsymbol{\eta})$ is evaluated on a subset of $\mathbb{E}_\delta$ where $\mathbb{D}_\delta$ intersects with $\{d : (d_0 - d) \geq 0.5\}$, then the criterion function is divergent. The latter corresponds to the integral on the right hand side in (3.6) being assigned the value ∞ if $(d_0 - d) \geq 0.5$ (see the comment by Hannan on his Lemma 2 in Hannan, 1973, page 134). This difference in behaviour of the criterion function about the point $d_0 - d = 0.5$ implies that $Q_n(\boldsymbol{\eta})$ does not converge uniformly on subsets of the parameter space that include this point. Nevertheless, as will be demonstrated below, provided that $\boldsymbol{\eta}_1 \in \mathbb{E}_\delta$, where $\boldsymbol{\eta}_1$ is the minimizer of $Q(\boldsymbol{\eta})$, $Q_n(\widehat{\boldsymbol{\eta}}_1)$ will converge to $Q(\boldsymbol{\eta}_1)$ and $\widehat{\boldsymbol{\eta}}_1$ will converge to $\boldsymbol{\eta}_1$ as a consequence.

In Section 3.2.3 we derive our asymptotic results pertaining to the convergence of $Q_n(\boldsymbol{\eta})$ and demonstrate the relationships between the limiting criterion functions of the Whittle, TML and CSS estimators to the limiting criterion function of the FML estimator. The value that minimizes the limiting criterion function of all four estimators is shown to be identical, and the asymptotic convergence of all four estimators to the common pseudo-true parameter, $\boldsymbol{\eta}_1$, is thereby established. In the theoretical derivations we adopt the assumption of a known mean for both the true and estimated models, with a zero value specified without loss of generality.

3.2.1 Frequency domain estimators

In their paper Chen and Deo (2006) focus on the estimator of $\boldsymbol{\eta} = \left(d, \boldsymbol{\beta}^\top\right)^\top$ defined as the value of $\boldsymbol{\eta}$ that minimizes the objective function

$$Q_n^{(1)}(\boldsymbol{\eta}) = \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{f_1(\boldsymbol{\eta}, \lambda_j)}, \tag{3.7}$$

where $I(\lambda_j)$ is the periodogram, defined as $I(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} y_t \exp(-i\lambda t) \right|^2$ evaluated at the Fourier frequencies $\lambda_j = 2\pi j/n$; $(j = 1, ..., \lfloor n/2 \rfloor)$, $\lfloor x \rfloor$ is the largest integer not greater than x. We have labeled this the FML estimator. The objective function in (3.7) is an approximation to the frequency domain Gaussian (negative) log-likelihood introduced initially by Whittle (1952) for short range dependent processes, namely

$$W_n(\sigma^2, \boldsymbol{\eta}) = \int_{-\pi}^{\pi} \left\{ \log \frac{\sigma^2}{2\pi} f_1(\boldsymbol{\eta}, \lambda) + \frac{2\pi I(\lambda)}{\sigma^2 f_1(\boldsymbol{\eta}, \lambda)} \right\} d\lambda, \tag{3.8}$$

and it coincides with the frequency domain objective function considered in Hannan (1973). Concentrating out σ^2 in 3.8 and minimizing the associated profile function with respect to $\boldsymbol{\eta}$ produces what we refer to as the exact Whittle estimator.

An alternative approximation to the Whittle criterion function in (3.8), considered for example in Beran (1994), is

$$Q_n^{(2)}(\sigma^2, \boldsymbol{\eta}) = \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \log \left[\frac{\sigma^2}{2\pi} f_1(\boldsymbol{\eta}, \lambda_j) \right] + \frac{(2\pi)^2}{\sigma^2 n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{f_1(\boldsymbol{\eta}, \lambda_j)}. \tag{3.9}$$

Taking $\boldsymbol{\eta}$ as the parameter of interest and concentrating $Q_n^{(2)}(\sigma^2, \boldsymbol{\eta})$ with respect to σ^2 indicates that the value of σ^2 that minimizes 3.9 is given by $\widehat{\sigma}^2(\boldsymbol{\eta}) = 2Q_n^{(1)}(\boldsymbol{\eta})$. Substituting back in to

3.9 yields the (negative) profile likelihood,

$$Q_n^{(2)}(\boldsymbol{\eta}) = \frac{2\pi}{2} \log\left(\frac{\widehat{\sigma}^2(\boldsymbol{\eta})}{2\pi}\right) + \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \log f_1(\boldsymbol{\eta}, \lambda_j) + \pi.$$

Minimization of $Q_n^{(2)}(\boldsymbol{\eta})$ with respect to $\boldsymbol{\eta}$ yields what we call (simply) the Whittle estimator, and which is the form of Whittle procedure that features in our theoretical derivations. Since $\lim_{n\to\infty} \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \log f_1(\boldsymbol{\eta}, \lambda_j) = 0$ (see Appendix 3.A) it follows that this estimator is equivalent to the FML estimator for large n. However, as indicated in Boes *et al.* (1989), and as will be seen in the simulation results documented in Section 3.5, the finite sample performance of these two estimators differs. For interest (and as prompted by a referee) we also report selected numerical results on the finite sample performance of the exact Whittle estimator described above.

3.2.2 Time domain estimators

The criterion functions of the two alternative time domain estimators are defined as follows:

- Let $\mathbf{Y}^\top = (y_1, y_2, ..., y_n)$ and denote the variance covariance matrix of $\mathbf{Y}$ derived from the mis-specified model by $\sigma^2 \boldsymbol{\Sigma}_\eta = [\gamma_1(i-j)]$, $i, j = 1, 2, ..., n$, where

$$\gamma_1(\tau) = \gamma_1(-\tau) = \frac{\sigma^2}{2\pi} \int_{-\pi}^{\pi} f_1(\boldsymbol{\eta}, \lambda) e^{i\lambda\tau} d\lambda\,.$$

The Gaussian log-likelihood function for the TML estimator is

$$-\frac{1}{2}\left(n \log(2\pi\sigma^2) + \log|\boldsymbol{\Sigma}_\eta| + \frac{1}{\sigma^2}(\mathbf{Y} - \mu\mathbf{l})^\top \boldsymbol{\Sigma}_\eta^{-1} (\mathbf{Y} - \mu\mathbf{l})\right), \qquad (3.10)$$

where $\mathbf{l}^\top = (1, 1, ..., 1)$, and maximizing (3.10) is equivalent to minimizing the criterion function

$$Q_n^{(3)}(\sigma^2, \boldsymbol{\eta}) = \log \sigma^2 + \frac{1}{n} \log|\boldsymbol{\Sigma}_\eta| + \frac{1}{n\sigma^2}(\mathbf{Y} - \mu\mathbf{l})^\top \boldsymbol{\Sigma}_\eta^{-1} (\mathbf{Y} - \mu\mathbf{l})\,. \qquad (3.11)$$

- To construct the CSS estimator note that we can expand $(1-z)^d$ in a binomial expansion as

$$(1-z)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} z^j , \tag{3.12}$$

where $\Gamma(\cdot)$ is the gamma function. Furthermore, since $g_1(\boldsymbol{\beta}, \lambda)$ is bounded, by Assumption $(A.3)$, we can employ the method of Whittle Whittle (1984, Section 2.8) to construct an autoregressive operator $\alpha(\boldsymbol{\beta}, z) = \sum_{i=0}^{\infty} \alpha_i(\boldsymbol{\beta}) z^i$ such that $g_1(\boldsymbol{\beta}, \lambda) = |\alpha(\boldsymbol{\beta}, e^{i\lambda})|^{-2}$. The objective function of the CSS estimation method then becomes

$$Q_n^{(4)}(\boldsymbol{\eta}) = \frac{1}{n} \sum_{t=1}^{n} e_t^2 , \tag{3.13}$$

where

$$e_t = \sum_{i=0}^{t-1} \tau_i(\boldsymbol{\eta}) \left(y_{t-i} - \mu \right) \tag{3.14}$$

and the coefficients $\tau_j(\boldsymbol{\eta})$, $j = 0, 1, 2, \ldots$, are given by $\tau_0(\boldsymbol{\eta}) = 1$ and

$$\tau_j(\boldsymbol{\eta}) = \sum_{s=0}^{j} \frac{\alpha_{j-s}(\boldsymbol{\beta})\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} , \quad j = 1, 2, \ldots . \tag{3.15}$$

As with the FML estimator, the CSS estimate of σ^2 is given implicitly by the minimum value of the criterion function.

We can think of the CSS estimator as providing an approximation to the TML estimator that parallels the approximation of the FML and (sums-based) Whittle estimators to the exact Whittle estimator.

3.2.3 Convergence properties

In Chen and Deo (2006) it is shown that if $\{y_t\}$ is a long range dependent Gaussian process, then on subsets of the parameter space of the form $(\delta, 0.5-\delta) \times \Phi$, where $0 < \delta < 0.25$ and Φ

is a compact convex set, we have (for $Q_n^{(1)}(\boldsymbol{\eta})$ defined in (3.7)) $\operatorname{plim}_{n\to\infty}\left|Q_n^{(1)}(\boldsymbol{\eta}) - Q(\boldsymbol{\eta})\right| = 0$ (Chen and Deo, 2006, Lemma 2). The minimum of the limiting objective function $Q(\boldsymbol{\eta})$ then defines a pseudo-true parameter value to which the FML estimator will converge, since with the addition of the assumption that there exists a unique vector $\boldsymbol{\eta}_1 = \left(d_1, \boldsymbol{\beta}_1^\top\right)^\top \in (\delta, 0.5 - \delta) \times \Phi$ that minimizes $Q(\boldsymbol{\eta})$, it follows that the FML estimator will converge to $\boldsymbol{\eta}_1$.

Because Chen and Deo assumed that the TDGP was a long memory process and that in the MisM the fractional index was similarly confined to the long memory region, they did not explicitly consider the case where $(d_0 - d) \geq 0.5$. In contrast, as noted with reference to the TDGP in (3.1), our work allows for $0 \leq |d_0| < 0.5$, and involves the specification of the appropriate user-assigned compact subset for $\boldsymbol{\eta} = \left(d, \boldsymbol{\beta}^\top\right)^\top$ in 3.5. This implies a wider range of values for $(d_0 - d)$ and, hence, the need for our analysis to deal with the differing behaviour of $Q_n^{(1)}(\boldsymbol{\eta})$ about the point $d_0 - d = 0.5$ alluded to above. To achieve this, we divide the parameter space $\mathbb{E}_\delta$ into three disjoint sub-sets:

1. $\mathbb{E}_\delta^0 = \mathbb{D}_\delta^0 \times \mathbb{B}$ where $\mathbb{D}_\delta^0 = \mathbb{D}_\delta \cap \{d : -(1-2\delta) \leq (d_0 - d) \leq 0.5 - \delta\}$,
2. $\overline{\mathbb{E}}_{\delta 1}^0 = \overline{\mathbb{D}}_{\delta 1}^0 \times \mathbb{B}$ where $\overline{\mathbb{D}}_{\delta 1}^0 = \mathbb{D}_\delta \cap \{d : 0.5 - \delta < (d_0 - d) < 0.5\}$, and,
3. $\overline{\mathbb{E}}_{\delta 2}^0 = \overline{\mathbb{D}}_{\delta 2}^0 \times \mathbb{B}$ where $\overline{\mathbb{D}}_{\delta 2}^0 = \mathbb{D}_\delta \cap \{d : 0.5 \leq (d_0 - d) \leq 1 - 2\delta\}$.

The superscript '0' is used to indicate that the relevant subspaces relate to the deviation $(d_0 - d)$ assuming that $d_0 \in \mathbb{D}_\delta$. The notation in 2. and 3. is used to denote the breakdown of the complement of the set in 1, $\mathbb{E}_\delta^0$, into two disjoint subsets, $\overline{\mathbb{E}}_{\delta 1}^0$ and $\overline{\mathbb{E}}_{\delta 2}^0$. This division of the parameter space of $(d_0 - d)$ is depicted graphically in Figure 3.1.

We will establish that on the subset $\mathbb{E}_\delta^0$ we have $\lim_{n\to\infty} Q_n^{(1)}(\boldsymbol{\eta}) = Q(\boldsymbol{\eta})$ almost surely and

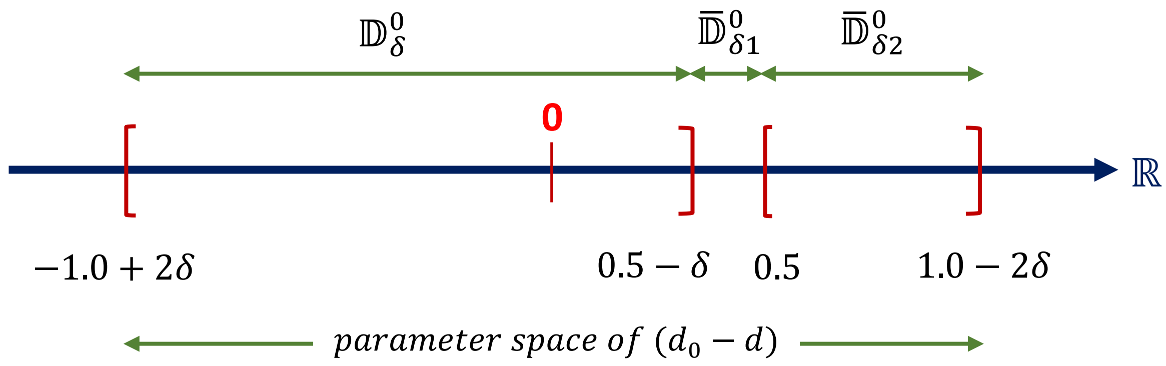

Figure 3.1: Graphical illustration of the division of the parameter space of $(d_0 - d)$

uniformly in $\boldsymbol{\eta}$, where $Q(\boldsymbol{\eta})$ is defined as in 3.6, whereas $Q_n^{(1)}(\boldsymbol{\eta})$ is of order $O(\delta^{-1})$ on $\overline{\mathbb{E}}^0_{\delta 1}$ and is divergent as $n \to \infty$ on $\overline{\mathbb{E}}^0_{\delta 2}$. This is the content of Lemmas 3.1, 3.2, 3.3 and 3.4 below. Proposition 3.1 then establishes that the FML estimator converges to $\boldsymbol{\eta}_1 = \arg\min_{\boldsymbol{\eta}} Q(\boldsymbol{\eta})$. We will also establish that the convergence/divergence properties of each of the three alternative estimators, Whittle, TML and CSS, is the same as that of the FML estimator. The upshot of this is summarized in Theorem 3.1.

Lemma 3.1 *Suppose that the TDGP of $\{y_t\}$ is as prescribed in equations (3.1) and (3.2) and that the MisM is specified as in (3.3). Assume also that Assumptions $(A.1) - (A.3)$ are satisfied. Then for any constant $\nu_f > 0$,*

$$\left| \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j) + \nu_f} - \frac{\sigma_0^2}{2\pi} \int_0^\pi \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta},\lambda) + \nu_f} d\lambda \right|$$

converges to zero almost surely and uniformly in $\boldsymbol{\eta}$ on $\mathbb{E}^0_\delta$.

Since, obviously, $f_1(\boldsymbol{\eta},\lambda) < f_1(\boldsymbol{\eta},\lambda) + \nu_f$ it follows from Lemma 3.1 that,

$$\begin{aligned}
\liminf_{n\to\infty} Q_n^{(1)}(\boldsymbol{\eta}) &\geq \lim_{n\to\infty} \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j) + \nu_f} \\
&= \frac{\sigma_0^2}{2\pi} \int_0^\pi \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta},\lambda) + \nu_f} \quad a.s.
\end{aligned}$$

uniformly in $\boldsymbol{\eta}$ on $\mathbb{E}_\delta^0$. Letting $\delta_f \to 0$ and applying Lebegue's monotone convergence theorem gives

$$\liminf_{n\to\infty} Q_n^{(1)}(\boldsymbol{\eta}) \geq Q(\boldsymbol{\eta}) = \frac{\sigma_0^2}{2\pi}\int_0^\pi \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta},\lambda)}d\lambda \quad a.s.$$

To establish that $Q(\boldsymbol{\eta})$ also provides a limit superior for $Q_n^{(1)}(\boldsymbol{\eta})$ when $\boldsymbol{\eta} \in \mathbb{E}_\delta^0$ we will use the following lemma.

Lemma 3.2 *Suppose that the conditions of Lemma 3.1 hold. Set*

$$h_1(\boldsymbol{\eta},\lambda) = \begin{cases} f_1(\boldsymbol{\eta},\lambda), & f_1(\boldsymbol{\eta},\lambda) \geq \nu_f \\ \nu_f, & f_1(\boldsymbol{\eta},\lambda) < \nu_f\,, \end{cases}$$

where $\nu_f > 0$. Then for all $\nu_f > 0$,

$$\left| \frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{h_1(\boldsymbol{\eta},\lambda_j)} - \frac{\sigma_0^2}{2\pi}\int_0^\pi \frac{f_0(\lambda)}{h_1(\boldsymbol{\eta},\lambda)}d\lambda \right|$$

converges to zero almost surely uniformly in $\boldsymbol{\eta}$ on $\mathbb{E}_\delta^0$.

The following lemma shows that the limiting form of the FML criterion function presented by Chen and Deo (2006), for Gaussian processes (specifically) and only in the case where both d and d_0 lie in the interval $(0, 0.5)$, holds more generally, and can incorporate all three forms of memory - long memory, short memory and antipersistence - in both the true and estimated models.

Lemma 3.3 *Suppose that the conditions of Lemmas 3.1 and 3.2 hold. Then,*

$$\lim_{n\to\infty} \sup_{\boldsymbol{\eta}\in\mathbb{E}_\delta^0} |Q_n^{(1)}(\boldsymbol{\eta}) - Q(\boldsymbol{\eta})| = 0\,.$$

Lemma 3.4 then indicates that for points in $\mathbb{E}_\delta$ where $(d_0 - d) > 0.5 - \delta$, $0 < \delta < 0.5$, uniform convergence of the criterion function $Q_n^{(1)}(\boldsymbol{\eta})$ fails.

Lemma 3.4 *Suppose that the TDGP of $\{y_t\}$ is as prescribed in equations (3.1) and (3.2) and that the MisM is specified as in (3.3). Assume also that Assumptions $(A.1)-(A.3)$ are satisfied. Then for all $\boldsymbol{\eta} \in \overline{\mathbb{E}}^0_{\delta 1}$ we have $\liminf_{n\to\infty} Q_n^{(1)}(\boldsymbol{\eta}) = O(\delta^{-1})$ and for $\boldsymbol{\eta} \in \overline{\mathbb{E}}^0_{\delta 2}$*

$$\liminf_{n\to\infty} Q_n^{(1)}(\boldsymbol{\eta}) \geq C > 0$$

almost surely for all C, no matter how large.

Note that Lemma 3.4 implies that as n increases, and for all δ sufficiently small, $\widehat{\boldsymbol{\eta}}_1^{(1)} = \arg\min_{\boldsymbol{\eta}} Q_n^{(1)}(\boldsymbol{\eta})$ cannot lie in $\overline{\mathbb{E}}^0_{\delta 1} \cup \overline{\mathbb{E}}^0_{\delta 2}$. Proposition 3.1 now follows as an almost immediate corollary of the previous developments if we suppose that the following additional assumption holds:

$(A.4)$ There exists a unique pseudo-true parameter vector $\boldsymbol{\eta}_1 = \left(d_1, \boldsymbol{\beta}_1^\top\right)^\top$ belonging to the subset $\mathbb{E}^0_\delta$ that satisfies $\boldsymbol{\eta}_1 = \arg\min_{\boldsymbol{\eta}} Q(\boldsymbol{\eta})$.

Proposition 3.1 establishes the convergence of the FML estimator to $\boldsymbol{\eta}_1$ under the same generality for both the TDGP and MisM as highlighted above (*cf.* Chen and Deo, 2006, Corollary 1).

Proposition 3.1 *Suppose that the TDGP of $\{y_t\}$ is as prescribed in equations (3.1) and (3.2) and that the MisM is specified as in (3.3). Assume also that Assumptions $(A.1)-(A.4)$ are satisfied. Let $\widehat{\boldsymbol{\eta}}_1^{(1)}$ denote the FML estimator obtained by minimizing the criterion function $Q_n^{(1)}(\boldsymbol{\eta})$ over $\mathbb{E}_\delta$. Then $\lim_{n\to\infty} Q_n^{(1)}(\widehat{\boldsymbol{\eta}}_1^{(1)}) = Q(\boldsymbol{\eta}_1)$ and $\widehat{\boldsymbol{\eta}}_1^{(1)} \to \boldsymbol{\eta}_1$ almost surely.*

Index now by $i = 2, 3$ and 4 the estimators associated with the Whittle, TML and CSS criterion functions respectively; that is $\widehat{\boldsymbol{\eta}}_1^{(i)}$ minimizes $Q_n^{(i)}(\cdot)$, $i = 2, 3, 4$, with each viewed as a function of $\boldsymbol{\eta}$. Given the relationships between $Q_n^{(1)}(\cdot)$ and $Q_n^{(i)}(\cdot)$, $i = 2, 3, 4$, as outlined in the appendix, it follows that $\widehat{\boldsymbol{\eta}}_1^{(i)}$, $i = 1, 2, 3, 4$, must share the same convergence properties. Thus we can state the following theorem:

Theorem 3.1 *Suppose that the TDGP of $\{y_t\}$ is as prescribed in equations (3.1) and (3.2) and that the MisM is specified as in (3.3). Assume also that Assumptions $(A.1) - (A.4)$ are satisfied. Let $\widehat{\boldsymbol{\eta}}_1^{(i)}$, $i = 1, 2, 3, 4$, denote, respectively, the FML, Whittle, TML and CSS estimators of the parameter vector $\boldsymbol{\eta} = \left(d, \boldsymbol{\beta}^\top\right)^\top$ of the MisM. Then $\lim_{n\to\infty} \|\widehat{\boldsymbol{\eta}}_1^{(i)} - \widehat{\boldsymbol{\eta}}_1^{(j)}\| = 0$ almost surely for all $i, j = 1, 2, 3, 4$, where the common limiting value of $\widehat{\boldsymbol{\eta}}_1^{(i)}$, $i = 1, 2, 3, 4$, is $\boldsymbol{\eta}_1 = \arg\min_{\boldsymbol{\eta}} Q(\boldsymbol{\eta})$.*

Having established that the four parametric estimators converge towards a common $\boldsymbol{\eta}_1$, we can as a consequence now broaden the applicability of the asymptotic distributional results derived by Chen and Deo (2006) for the FML estimator. This we do in Section 3.4 by establishing that all four alternative parametric estimators converge in distribution for all three forms of memory - long memory, short memory and antipersistence.

Prior to doing this, however, we note that Cavaliere *et al.* (2017) have shown that if $g_0(\lambda)$ has a parametric form that is known, but the parameter values that characterize it are not, then the CSS estimator will convergence to a pseudo-true value if it is evaluated whilst imposing incorrect linear parameter constraints (Cavaliere *et al.*, 2017, Theorem 5(ii)). Theorem 3.1 provides a generalization of this result by extending it to the FML, Whittle and TML estimators, and by also allowing for the possibility that the parametric form of the model, $g_1(\lambda)$, may itself be mis-specified. In what follows we indicate the precise form of the limiting objective function $Q(\boldsymbol{\eta})$, and the associated first-order conditions that define the pseudo-true value $\boldsymbol{\eta}_1$ of the four estimation procedures, in the ARFIMA case.

3.3 Pseudo-true parameters under ARFIMA mis-specification

Under Assumptions $A.1 - A.4$, the value of $\boldsymbol{\eta}_1 = \arg\min_{\boldsymbol{\eta}} Q(\boldsymbol{\eta})$ can be determined as the solution of the first-order condition $\partial Q(\boldsymbol{\eta})/\partial \boldsymbol{\eta} = 0$, and Chen and Deo (2006) illustrate the relationship between $\partial \log Q(\boldsymbol{\eta})/\partial d$ and the deviation $d^* = d_0 - d_1$ for the simple special case in which the TDGP is an $ARFIMA(0, d_0, 1)$ and the MisM is an $ARFIMA(0, d, 0)$. They then cite (without providing detailed derivations) certain results that obtain when the MisM is an $ARFIMA(1, d, 0)$. Here we provide a significant generalization, by deriving expressions for both $Q(\boldsymbol{\eta})$ and the first-order conditions that define the pseudo-true parameters, under the full $ARFIMA(p_0, d_0, q_0)$ / $ARFIMA(p, d, q)$ dichotomy for the true process and the estimated model. Representations of the associated expressions via polynomial and power series expansions suitable for the analytical investigation of $Q(\boldsymbol{\eta})$ are presented. It is normally not possible to solve the first order conditions $\partial Q(\boldsymbol{\eta})/\partial \boldsymbol{\eta} = 0$ exactly as they are both nonlinear and (in general) defined as infinite sums. Instead one would determine the estimate numerically, via a Newton iteration for example, with the series expansions replaced by finite sums. An evaluation of the magnitude of the approximation error produced by any power series truncation that might arise from such a numerical implementation is given. The results are then illustrated in the special case where $p_0 = q = 0$, in which case true MA short memory dynamics of an arbitrary order are mis-specified as AR dynamics of an arbitrary order. In this particular case, as will be seen, no truncation error arises in the computations.

To begin, denote the spectral density of the TDGP, a general $ARFIMA(p_0, d_0, q_0)$ process, by

$$\frac{\sigma_0^2}{2\pi} f_0(\lambda) = \frac{\sigma_0^2}{2\pi} \frac{\left|1 + \theta_{10} e^{i\lambda} + ... + \theta_{q_0 0} e^{iq_0\lambda}\right|^2}{\left|1 + \phi_{10} e^{i\lambda} + ... + \phi_{p_0 0} e^{ip_0\lambda}\right|^2} |2\sin(\lambda/2)|^{-2d_0},$$

and that of the MisM, an $ARFIMA(p,d,q)$ model, by

$$\frac{\sigma^2}{2\pi}f_1(\boldsymbol{\eta},\lambda) = \frac{\sigma^2}{2\pi}\frac{\left|1+\theta_1 e^{i\lambda}+...+\theta_q e^{iq\lambda}\right|^2}{\left|1+\phi_1 e^{i\lambda}+...+\phi_p e^{ip\lambda}\right|^2}|2\sin(\lambda/2)|^{-2d}.$$

Substituting these expressions into the limiting objective function we obtain the representation

$$Q(\boldsymbol{\eta}) = \frac{\sigma_0^2}{2\pi}\int_0^{\pi}\frac{f_0(\lambda)}{f_1(\boldsymbol{\eta},\lambda)}d\lambda = \frac{\sigma_0^2}{2\pi}\int_0^{\pi}\frac{|A_\beta(e^{i\lambda})|^2}{|B_\beta(e^{i\lambda})|^2}|2\sin(\lambda/2)|^{-2(d_0-d)}d\lambda\,, \tag{3.16}$$

where

$$A_\beta(z) = \sum_{j=0}^{\underline{q}} a_j z^j = \theta_0(z)\phi(z) = \left(1+\theta_{10}z+...+\theta_{q_0 0}z^{q_0}\right)\left(1+\phi_1 z+...+\phi_p z^p\right), \tag{3.17}$$

with $\underline{q} = q_0 + p$ and

$$B_\beta(z) = \sum_{j=0}^{\underline{p}} b_j z^j = \phi_0(z)\theta(z) = \left(1+\phi_{10}z+...+\phi_{p_0 0}z^{p_0}\right)\left(1+\theta_1 z+...+\theta_q z^q\right), \tag{3.18}$$

with $\underline{p} = p_0 + q$. The expression for $Q(\boldsymbol{\eta})$ in (3.16) takes the form of the variance of an ARFIMA process with MA operator $A_\beta(z)$, AR operator $B_\beta(z)$ and fractional index $d_0 - d$. It follows that $Q(\boldsymbol{\eta})$ could be evaluated using the procedures presented in Sowell (1992). Sowell's algorithms are based upon series expansions in gamma and hypergeometric functions however, and although they are suitable for numerical calculations, they do not readily lend themselves to the analytical investigation of $Q(\boldsymbol{\psi})$. We therefore seek an alternative formulation.

Let $C(z) = \sum_{j=0}^{\infty} c_j z^j = A_\beta(z)/B_\beta(z)$ where $A_\beta(z)$ and $B_\beta(z)$ are as defined in (3.17) and (3.18) respectively. Then (3.16) can be expanded to give

$$Q(\boldsymbol{\eta}) = 2^{1-2(d_0-d)}\frac{\sigma_0^2}{2\pi}\left[\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}c_j c_k\int_0^{\pi/2}\cos\left(2\left(j-k\right)\lambda\right)\sin(\lambda)^{-2(d_0-d)}d\lambda\right].$$

Using standard results for the integral $\int_0^{\pi}(\sin x)^{v-1}\cos(ax)dx$ from Gradshtein and Ryzhik

(2007, page 397) yields, after some algebraic manipulation,

$$Q\left(\boldsymbol{\eta}\right)=\frac{\sigma_0^2}{2(1-2(d_0-d))}\left[\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{c_jc_k\cos\left(\left(j-k\right)\pi\right)}{\mathcal{B}\left(1-(d_0-d)+\left(j-k\right),1-(d_0-d)-\left(j-k\right)\right)}\right],$$

where $\mathcal{B}(a,b)$ denotes the Beta function. This expression can in turn be simplified to

$$Q\left(\boldsymbol{\eta}\right)=\frac{\sigma_0^2\Gamma(1-2(d_0-d))}{2\Gamma^2(1-(d_0-d))}\}K(\boldsymbol{\eta}), \tag{3.19}$$

where

$$K(\boldsymbol{\eta})=\sum_{j=0}^{\infty}c_j^2+2\sum_{k=0}^{\infty}\sum_{j=k+1}^{\infty}c_jc_k\rho(j-k)$$

and

$$\rho(h)=\prod_{i=1}^{h}\left(\frac{(d_0-d)+i-1}{i-(d_0-d)}\right),\quad h=1,2,\ldots.$$

Using (3.19) we now derive the form of the first-order conditions that define $\boldsymbol{\eta}_1$, namely $\partial Q(\boldsymbol{\eta})/\partial\boldsymbol{\eta}=0$. Differentiating $Q\left(\boldsymbol{\eta}\right)$ first with respect to β_r, $r=1,\ldots,l$, and then d gives:

$$\frac{\partial Q\left(\boldsymbol{\eta}\right)}{\partial\beta_r}=\{\frac{\sigma_0^2\Gamma(1-2(d_0-d))}{2\Gamma^2(1-(d_0-d))}\}\frac{\partial K\left(\boldsymbol{\eta}\right)}{\partial\beta_r},\quad r=1,2,...,l,$$

where

$$\frac{\partial K\left(\boldsymbol{\eta}\right)}{\partial\beta_r}=\sum_{j=1}^{\infty}2c_j\frac{\partial c_j}{\partial\beta_r}+2\sum_{k=0}^{\infty}\sum_{j=k+1}^{\infty}(c_k\frac{\partial c_j}{\partial\beta_r}+\frac{\partial c_k}{\partial\beta_r}c_j)\rho(j-k),$$

and

$$\frac{\partial Q\left(\boldsymbol{\eta}\right)}{\partial d}=\{\frac{\sigma_0^2\Gamma(1-2(d_0-d))}{2\Gamma^2(1-(d_0-d))}\}\left\{2\left(\Psi[1-2(d_0-d)]-\Psi[1-(d_0-d)]\right)K(\boldsymbol{\eta})+\frac{\partial K\left(\boldsymbol{\eta}\right)}{\partial d}\right\},$$

where $\Psi(\cdot)$ denotes the digamma function and

$$\begin{aligned}\frac{\partial K\left(\boldsymbol{\eta}\right)}{\partial d}=&2\sum_{k=0}^{\infty}\sum_{j=k+1}^{\infty}c_jc_k\rho(j-k)\left\{2\Psi[1-(d_0-d)]\right.\\&\left.-\Psi[1-(d_0-d)+(j-k)]-\Psi[1-(d_0-d)-(j-k)]\right\}.\end{aligned}$$

Eliminating the common (non-zero) factor $\{\pi \frac{\sigma_0^2}{\sigma^2} \frac{\Gamma(1-2(d_0-d))}{\Gamma^2(1-(d_0-d))}\}$ from both $\partial Q(\boldsymbol{\eta})/\partial \boldsymbol{\beta}$ and $\partial Q(\boldsymbol{\eta})/\partial d$, it follows that the pseudo-true parameter values of the $ARFIMA(p,d,q)$ MisM can be obtained by solving

$$\frac{\partial K(\boldsymbol{\eta})}{\partial \beta_r} = 0, \quad r = 1, 2, ..., l, \tag{3.20}$$

and

$$2(\Psi[1-2(d_0-d)] - \Psi[1-(d_0-d)])K(\boldsymbol{\eta}) + \frac{\partial K(\boldsymbol{\eta})}{\partial d} = 0 \tag{3.21}$$

for β_{r1}, $r = 1, \ldots, l$, and d_1 using appropriate algebraic and numerical procedures. A corollary of the following theorem is that $\boldsymbol{\eta}_1$ can be calculated to any desired degree of numerical accuracy by truncating the series expansions in the expressions for $K(\boldsymbol{\eta})$, $\partial K(\boldsymbol{\eta})/\partial \boldsymbol{\beta}$ and $\partial K(\boldsymbol{\eta})/\partial d$ after a suitable number of N terms before substituting into (3.20) and (3.21) and solving (numerically) for ϕ_{i1}, $i = 1, 2, ..., p$, θ_{j1}, $j = 1, 2, ..., q$, and d_1.

Theorem 3.2 *Set* $C_N(z) = \sum_{j=0}^{N} c_j z^j$ *and let* $Q_N(\boldsymbol{\eta}) = \left(\sigma_0^2/\sigma^2\right) I_N$ *where the integral* $I_N = \int_0^{\pi} |C_N(\exp(-i\lambda))|^2 |2\sin(\lambda/2)|^{-2(d_0-d)} d\lambda$. *Then,*

$$Q(\boldsymbol{\eta}) = Q_N(\boldsymbol{\eta}) + R_N = \left\{ \frac{\sigma_0^2 \Gamma(1-2(d_0-d))}{2\Gamma^2(1-(d_0-d))} \right\} K_N(\boldsymbol{\eta}) + R_N$$

where

$$K_N(\boldsymbol{\eta}) = \sum_{j=0}^{N} c_j^2 + 2\sum_{k=0}^{N-1} \sum_{j=k+1}^{N} c_j c_k \rho(j-k)$$

and there exists a ζ, $0 < \zeta < 1$, *such that* $R_N = O(\zeta^{(N+1)}) = o(N^{-1})$. *Furthermore,* $\partial Q_N(\boldsymbol{\eta})/\partial \boldsymbol{\eta} = \partial Q(\boldsymbol{\eta})/\partial \boldsymbol{\eta} + o(N^{-1})$.

By way of illustration, consider the case of mis-specifying a true $ARFIMA(0, d_0, q_0)$ process by an $ARFIMA(p, d, 0)$ model. When $p_0 = q = 0$ we have $B_\beta(z) \equiv 1$ and $C(z)$ is polynomial,

$C(z) = 1 + \sum_{j=1}^{q} c_j z^j$ where $c_j = \sum_{r=\max\{0,j-p\}}^{\min\{j,p\}} \theta_{(j-r)0}\phi_r$. Abbreviating the latter to $\sum_r \theta_{(j-r)0}\phi_r$, this then gives us:

$$
\begin{aligned}
K(d, \phi_1, \ldots, \phi_p) = & \sum_{j=0}^{q} (\sum_r \theta_{(j-r)0}\phi_r)^2 + \\
& 2\sum_{k=0}^{q-1} \sum_{j=k+1}^{q} (\sum_r \theta_{(j-r)0}\phi_r)(\sum_r \theta_{(k-r)0}\phi_r)\rho(j-k)\,;
\end{aligned}
$$

and setting $\theta_{s0} \equiv 0$, $s \ni [0, 1, \ldots, q_0]$,

$$
\begin{aligned}
\frac{\partial K\left(d, \phi_1, \ldots, \phi_p\right)}{\partial \phi_r} = & \sum_{j=1}^{q} 2(\sum_r \theta_{(j-r)0}\phi_r)\theta_{(j-r)0} + \\
& 2\sum_{k=0}^{q-1} \sum_{j=k+1}^{q} \left\{ (\sum_r \theta_{(j-r)0}\phi_r)\theta_{(k-r)0} + \theta_{(j-r)0}(\sum_r \theta_{(k-r)0}\phi_r) \right\} \rho(j-k)\,,
\end{aligned}
$$

$r = 1, \ldots, p$, and

$$
\begin{aligned}
\frac{\partial K\left(d, \phi_1, \ldots, \phi_p\right)}{\partial d} = & \; 2\sum_{k=0}^{q-1} \sum_{j=k+1}^{q} (\sum_r \theta_{(j-r)0}\phi_r)(\sum_r \theta_{(k-r)0}\phi_r)\rho(j-k) \\
& \times (2\Psi[1-(d_0-d)] - \Psi[1-(d_0-d)+(j-k)] \\
& -\Psi[1-(d_0-d)-(j-k)])
\end{aligned}
$$

for the required derivatives. The pseudo-true values ϕ_{r1}, $r = 1, \ldots, p$, and d_1 can now be obtained by solving (3.20) and (3.21) having inserted these exact expressions for $K\left(d, \phi_1, \ldots, \phi_p\right)$, $\partial K\left(d, \phi_1, \ldots, \phi_p\right)/\partial \phi_r$, $r = 1, \ldots, p$, and $\partial K\left(d, \phi_1, \ldots, \phi_p\right)/\partial d$ into the equations.

Let us further highlight some features of this special case by focussing on the example where the TDGP is an $ARFIMA(0, d_0, 1)$ and the MisM an $ARFIMA(1, d, 0)$. In this example $\underline{q} = 2$ and $C(z) = 1 + c_1 z + c_2 z^2$ where, neglecting the first order MA and AR coefficient subscripts, $c_1 = (\theta_0 + \phi)$ and $c_2 = \theta_0\phi$. The second factor of the criterion function in (3.19) is now

$$
K(d, \phi) = 1 + (\theta_0 + \phi)^2 + (\theta_0\phi)^2
$$

$$+\frac{2\left[\theta_0\phi(d_0-d+1)-(1+\theta_0\phi)(\theta_0+\phi)(d_0-d-2)\right](d_0-d)}{(d_0-d-1)(d_0-d-2)}. \tag{3.22}$$

The derivatives $\partial K(d,\phi)/\partial\phi$ and $\partial K(d,\phi)/\partial d$ can be readily determined from (3.22) and hence the pseudo-true values d_1 and ϕ_1 evaluated.

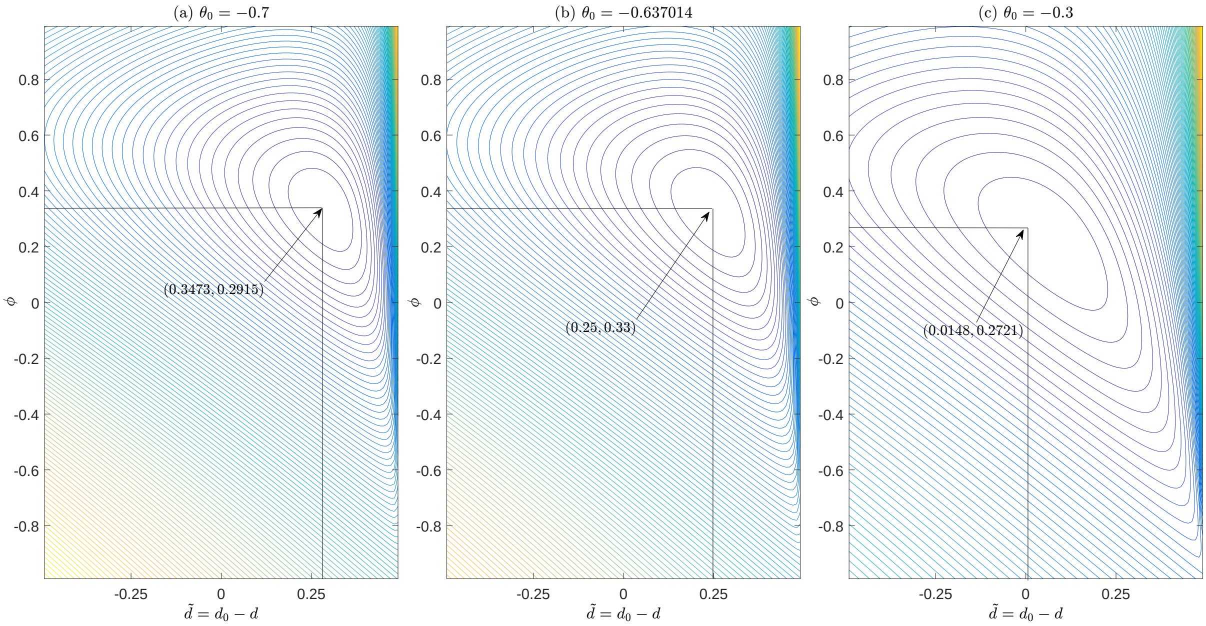

Figure 3.2: Contour plot of $Q(d,\phi)$ against $\widetilde{d}= d_0{-}d$ and ϕ for the mis-specification of an ARFIMA$(0,d_0,1)$ TDGP by an ARFIMA$(1,d,0)$ MisM; $\widetilde{d}\in(-0.5,0.5)$, $\phi\in(-1,1)$. Pseudo-true coordinates $(d_0{-}d_1,\phi_1)$ are (a) $(0.2915, 0.3473)$, (b) $(0.25, 0.33)$ and (c) $(0.0148, 0.2721)$.

It is clear from (3.22) that for given values of $|\theta_0| < 1$ we can treat $K(d,\phi)$ as a function of $\widetilde{d} = (d_0 - d)$ and ϕ, and hence treat $Q(d,\phi) = Q(\boldsymbol{\eta})$ similarly. Figure 3.2 depicts the contours of $Q(d,\phi)$ graphed as a function of $\widetilde{d}$ and ϕ for the values of $\theta_0 = \{-0.7, -0.637014, -0.3\}$ when $\sigma^2 = \sigma_0^2$. Pre-empting the discussion to come in the following section, the values of θ_0 are deliberately chosen to coincide with $d^* = d_0 - d_1$ being respectively greater than, equal to and less than 0.25.

The three graphs in Figure 3.2 clearly demonstrate the divergence in the asymptotic criterion function that occurs as $\widetilde{d} = (d_0 - d)$ approaches 0.5 and they illustrate that although the location of (d_1, ϕ_1) may be unambiguous, the sensitivity of $Q(d, \phi)$ to perturbations in (d, ϕ) can be very different depending on the value of $d^* = d_0 - d_1$.[4] In Figure 2(a) the contours indicate that when $d^* > 0.25$ the limiting criterion function has hyperbolic profiles in a small neighbourhood of the pseudo-true parameter point (d_1, ϕ_1), with similar but more locally quadratic behaviour exhibited in Figure 2(b) when $d^* = 0.25$. The contours of $Q(d, \phi)$ in Figure 2(c), corresponding to $d^* < 0.25$, are more elliptical and suggest that in this case the limiting criterion function is far closer to being globally quadratic around (d_1, ϕ_1). It turns out that these three different forms of $Q(d, \phi)$, reflecting the most, intermediate, and the least mis-specified cases, correspond to the three different forms of asymptotic distribution presented in the following section.

3.4 Asymptotic distributions

In this section we show that the asymptotic distribution of the FML estimator derived in Chen and Deo (2006) in the context of long range dependence is also applicable to the Whittle, TML and CSS estimators, and that all four estimators are, hence, asymptotically equivalent under mis-specification. As was highlighted by Chen and Deo, the rate of convergence and the nature of the asymptotic distribution of the FML estimator is determined by the deviation of the pseudo-true value d_1 from the true value d_0.[5]. Theorem 3.3 shows that in the event that any one of the FML, Whittle, TML or CSS estimators possesses one of the asymptotic distributions as described in the theorem, then all four estimators will share the same asymptotic distribution, and this will hold for all three forms of memory in the TDGP and the mis-specified

[4] All the numerical results presented in this chapter have been produced using MATLAB *2011b*, version *7.13.0.564 (R2011b)*.

[5] As already noted, the results in Chen and Deo presupposed that the parameter space of the estimated model coincided with the long memory region assumed for the TDGP. Since d_1 is only defined for $(d_0 - d_1) < 0.5$ it follows that the distributional results they presented for the FML estimator were only valid for this region, something that was not explicitly mentioned in their original derivation.

model. We comment further on this matter below.

For each of the estimators the asymptotic distributions are obtained via the usual Taylor series expansion of the score function, having first established convergence, and consequently stronger smoothness conditions are required to establish the asymptotic distribution theory and to ensure that the asymptotic variance-covariance matrix of the estimators is well defined. We will therefore suppose:

$(A.5)$ The function $g_1(\boldsymbol{\beta},\lambda)$ of the MisM is thrice differentiable with continuous third derivatives. Furthermore, the derivatives satisfy;

$(A.5.1)$ $\sup_{\lambda}\sup_{\boldsymbol{\beta}}\left|\frac{\partial g_1(\boldsymbol{\beta},\lambda)}{\partial \beta_i}\right| < \infty$, $1 \leqslant i \leqslant l$,

$(A.5.2)$ $\sup_{\lambda}\sup_{\boldsymbol{\beta}}\left|\frac{\partial^2 g_1(\boldsymbol{\beta},\lambda)}{\partial \beta_i \partial \beta_j}\right| < \infty$, $\sup_{\lambda}\sup_{\boldsymbol{\beta}}\left|\frac{\partial^2 g_1(\boldsymbol{\beta},\lambda)}{\partial \beta_i \partial \lambda}\right| < \infty$, $1 \leqslant i, j \leqslant l$, and

$(A.5.3)$ $\sup_{\lambda}\sup_{\boldsymbol{\beta}}\left|\frac{\partial^3 g_1(\boldsymbol{\beta},\lambda)}{\partial \beta_i \partial \beta_j \partial \beta_k}\right| < \infty$, $1 \leqslant i, j, k \leqslant l$.

Assumptions $(A.2) - (A.5)$ are similar to the assumptions adopted by Fox and Taqqu (1986) and Dahlhaus (1989) in the context of correct specification, and they are in essence equivalent to the conditions used in the work of Chen and Deo (2006) on the mis-specified case. In order to derive the asymptotic distribution we will assume that $\{\varepsilon_t\}$ is a strictly stationary, regular process that satisfies the following weak dependence and moment conditions.

$(A.1')$ The innovation $\{\varepsilon_t\}$ satisfies Assumption $(A.1)$. Moreover, $E_0[|\varepsilon_t|^{4+p}] < \infty$ for some $p \in (0, \infty)$ and there exist finite constants μ_3 and μ_4 such that $E_0[\varepsilon_t^3|\mathbb{F}_{t-1}] = \mu_3$ and $E_0[\varepsilon_t^4|\mathbb{F}_{t-1}] = \mu_4$ a.s. for all $t \in \mathbb{Z}$.

Assumption $(A.1')$ states that $\{\varepsilon_t\}$ is a martingale difference sequence, a not unreasonable assumption that has almost become standard in the asymptotic analysis of time series, and it is typical to assume finite bounds on the first four moments of the innovation process (see Cavaliere *et al.*, 2017, for a detailed explanation of the importance of such bounds). Assumption $(A.1')$ implies that $\{\varepsilon_t\}$ is completely regular, and that $\{\varepsilon_t^p\}$ is a uniformly integrable sequence for any $p \leq 4$.[6]

Theorem 3.3 *Suppose that the TDGP of $\{y_t\}$ is as prescribed in equations (3.1) and (3.2) and that the MisM is specified as in (3.3), and assume that Assumptions $(A.1')$ and $(A.2)-(A.5)$ hold. Let*

$$\mathbf{B} = -\frac{\sigma_0^2}{\pi}\int_{-\pi}^{\pi}\frac{f_0(\lambda)}{f_1^3(\boldsymbol{\eta}_1,\lambda)}\frac{\partial f_1(\boldsymbol{\eta}_1,\lambda)}{\partial\boldsymbol{\eta}}\frac{\partial f_1(\boldsymbol{\eta}_1,\lambda)}{\partial\boldsymbol{\eta}^T}d\lambda + \frac{\sigma_0^2}{2\pi}\int_{-\pi}^{\pi}\frac{f_0(\lambda)}{f_1^2(\boldsymbol{\eta}_1,\lambda)}\frac{\partial^2 f_1(\boldsymbol{\eta}_1,\lambda)}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T}d\lambda\,, \tag{3.23}$$

and set $\boldsymbol{\mu}_n = \mathbf{B}^{-1}E_0\left(\frac{\partial Q_n(\boldsymbol{\eta}_1)}{\partial\boldsymbol{\eta}}\right)$ where $Q_n(\cdot)$ denotes the objective function that defines $\widehat{\boldsymbol{\eta}}_1$.[7] *Let $\widehat{\boldsymbol{\eta}}_1$ denote the estimator obtained by minimizing $Q_n(\boldsymbol{\eta})$ over the compact set $\mathbb{E}_\delta$ where $\boldsymbol{\eta}_1 \in \mathbb{E}_\delta$ and assume that $\boldsymbol{\eta}_1 \ni \partial\mathbb{E}_\delta$ where $\partial\mathbb{E}_\delta$ is the boundary of the set $\mathbb{E}_\delta$. Then the FML, Whittle, TML or CSS estimators are asymptotically equivalent with a common limiting distribution as delineated in Cases 3.1, 3.2 and 3.3:*

Case 3.1 *When $d^* = d_0 - d_1 > 0.25$,*

$$\frac{n^{1-2d^*}}{\log n}\left(\widehat{\boldsymbol{\eta}}_1 - \boldsymbol{\eta}_1 - \boldsymbol{\mu}_n\right) \rightarrow^D \mathbf{B}^{-1}\left[\sum_{j=1}^{\infty} W_j, 0, ..., 0\right]^{\top}, \tag{3.24}$$

where $\sum_{j=1}^{\infty} W_j$ is the mean square limit of the random sequence $\sum_{j=1}^{s} W_j$ as $s \rightarrow \infty$ wherein

$$W_j = \frac{(2\pi)^{1-2d^*}g_0(\boldsymbol{\eta}_0,0)}{j^{2d^*}g_1(\boldsymbol{\eta}_1,0)}\left[U_j^2 + V_j^2 - E_0\left(U_j^2 + V_j^2\right)\right],$$

[6] This assumption is closely related to Assumption $(A.1)$ of Lahiri (2003), which specifies a set of weak dependence and moment conditions on $\{\varepsilon_t\}$ based on $\alpha-$mixing.

[7] Heuristically, $\boldsymbol{\mu}_n$ measures the bias associated with the estimator $\widehat{\boldsymbol{\eta}}_1$. That is, $\boldsymbol{\mu}_n \approx E_0(\widehat{\boldsymbol{\eta}}_1) - \boldsymbol{\eta}_1$. Note that the expression for $\boldsymbol{\mu}_n$ given in Chen and Deo (2006, page 263) contains a typographical error; the proofs in that paper use the correct expression. The derivation of $\boldsymbol{\mu}_n$ for all four estimation methods considered in this chapter is provided in Appendix 3.B.

where, $\{U_j, V_k\}$ are a sequence of random normal variables with zero mean and the covariance structure of $\{U_j, V_k\}$ denoted by $Cov_0(.)$ is as follows,

$$\begin{aligned} Cov_0\left(U_j, V_k\right) &= \iint\limits_{[0,1]^2} \{\sin(2\pi jx)\sin(2\pi ky) + \sin(2\pi kx)\sin(2\pi jy)\} \, |x-y|^{2d_0-1}\, dxdy, \\ Cov_0\left(U_j, U_k\right) &= Cov_0\left(U_j, V_k\right) = Cov_0\left(V_j, V_k\right), \forall j,k \in \mathbb{N}. \end{aligned} \tag{3.25}$$

Case 3.2 *When $d^* = d_0 - d_1 = 0.25$,*

$$n^{1/2}\overline{\Lambda}^{-1/2}\left(\widehat{\boldsymbol{\eta}}_1 - \boldsymbol{\eta}_1\right) \to^D \mathbf{B}^{-1}\left(Z, 0, ..., 0\right)^\top, \tag{3.26}$$

where

$$\overline{\Lambda} = \frac{1}{n}\sum_{j=1}^{n/2}\left(\frac{f_0(\lambda_j)}{f_1(\boldsymbol{\eta}_1,\lambda_j)}\frac{\partial \log f_1(\boldsymbol{\eta}_1,\lambda_j)}{\partial d}\right)^2, \tag{3.27}$$

and Z is a standard normal random variable.

Case 3.3 *When $d^* = d_0 - d_1 < 0.25$,*

$$\sqrt{n}\left(\widehat{\boldsymbol{\eta}}_1 - \boldsymbol{\eta}_1\right) \to^D N(0, \Xi), \tag{3.28}$$

where $\Xi = \mathbf{B}^{-1}\boldsymbol{\Lambda}\mathbf{B}^{-1}$, and

$$\boldsymbol{\Lambda} = 2\pi\int_0^\pi \left(\frac{f_0(\lambda)}{f_1(\boldsymbol{\eta}_1,\lambda)}\right)^2 \left(\frac{\partial \log f_1(\boldsymbol{\eta}_1,\lambda)}{\partial \boldsymbol{\eta}}\right)\left(\frac{\partial \log f_1(\boldsymbol{\eta}_1,\lambda)}{\partial \boldsymbol{\eta}}\right)^\top d\lambda. \tag{3.29}$$

A key point to note from the three cases delineated in Theorem 3.3 is that when the deviation between the true and pseudo-true values of d is sufficiently large ($d^* \geq 0.25$) – something that is related directly to the degree of mis-specification of $g_0(\lambda)$ by $g_1(\boldsymbol{\beta}, \lambda)$ – the $\sqrt{n}$ rate of convergence is lost, with the rate being arbitrarily close to zero depending on the value of d^*. For d^* strictly greater than 0.25, asymptotic Gaussianity is also lost, with the limiting distribution being a function of an infinite sum of non-Gaussian variables. For the $d^* \geq 0.25$ case, the limiting distribution – whether Gaussian or otherwise – is degenerate in the sense that the limiting distribution for each element of $\widehat{\boldsymbol{\eta}}_1$ is a different multiple of the same random variable ($\sum_{j=1}^{\infty} W_j$ in the case of $d^* > 0.25$ and Z in the case of $d^* = 0.25$).

For the form of limiting distribution that obtains in Cases 1, 2 and 3 we refer to Theorems 1, 3 and 2 of Chen and Deo (2006), wherein these distributions were produced specifically for the FML estimator in the context of long range dependence. Their proofs depend on the Fourier sine and cosine transformations of the observed series being normally distributed with a given covariance structure. In Chen and Deo (2006) the latter properties are derived by assuming that $\{x(t)\}$ is a Gaussian process. Here we achieve the same outcome by employing Assumption $A.1'$ and appealing to results of Lahiri (2003) which imply that the Fourier sine and cosine transformations are asymptotically normal and hence that lemmas of Moulines and Soulier (1999) used by Chen and Deo can be applied in a more general setting.

To prove that these same limiting distributions hold for the Whittle, TML and CSS estimators we establish that $R_n(\widehat{\boldsymbol{\eta}}_1^{(i)} - \widehat{\boldsymbol{\eta}}_1^{(1)}) \to^D 0$ for $i = 2, 3$ and 4, where R_n denotes the convergence rate applicable in the three different cases outlined in the theorem. We use a first-order Taylor expansion of $\partial Q_n^{(\cdot)}(\boldsymbol{\eta}_1)/\partial\boldsymbol{\eta}$ about $\partial Q_n^{(\cdot)}(\widehat{\boldsymbol{\eta}}_1^{(\cdot)})/\partial\boldsymbol{\eta} = \mathbf{0}$. This gives

$$\frac{\partial Q_n^{(\cdot)}(\boldsymbol{\eta}_1)}{\partial\boldsymbol{\eta}} = \frac{\partial^2 Q_n^{(\cdot)}(\grave{\boldsymbol{\eta}}_1^{(\cdot)})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top}\left(\boldsymbol{\eta}_1 - \widehat{\boldsymbol{\eta}}_1^{(\cdot)}\right)$$

and

$$R_n(\widehat{\boldsymbol{\eta}}_1^{(i)} - \widehat{\boldsymbol{\eta}}_1^{(j)}) = \left[\frac{\partial^2 Q_n^{(j)}(\grave{\boldsymbol{\eta}}_1^{(j)})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top}\right]^{-1} R_n \frac{\partial Q_n^{(j)}(\boldsymbol{\eta}_1)}{\partial\boldsymbol{\eta}} - \left[\frac{\partial^2 Q_n^{(i)}(\grave{\boldsymbol{\eta}}_1^{(i)})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top}\right]^{-1} R_n \frac{\partial Q_n^{(i)}(\boldsymbol{\eta}_1)}{\partial\boldsymbol{\eta}},$$

where $\|\boldsymbol{\eta}_1 - \grave{\boldsymbol{\eta}}_1^{(\cdot)}\| \leq \|\boldsymbol{\eta}_1 - \widehat{\boldsymbol{\eta}}_1^{(\cdot)}\|$. Since $\operatorname{plim}\widehat{\boldsymbol{\eta}}_1^{(\cdot)} = \boldsymbol{\eta}_1$ it is therefore sufficient to show that there exists a scalar, possibly constant, function $\mathcal{C}_n(\boldsymbol{\eta})$ such that

$$\left\| \frac{\partial^2 \{\mathcal{C}_n(\boldsymbol{\eta}_1) \cdot Q_n^{(i)}(\boldsymbol{\eta}_1) - Q_n^{(j)}(\boldsymbol{\eta}_1)\}}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top} \right\| = o_p(1) \tag{3.30}$$

and

$$\underset{n\to\infty}{\text{plim}} \quad R_n \left\| \mathcal{C}_n(\boldsymbol{\eta}_1) \cdot \frac{\partial Q_n^{(i)}(\boldsymbol{\eta}_1)}{\partial \boldsymbol{\eta}} - \frac{\partial Q_n^{(j)}(\boldsymbol{\eta}_1)}{\partial \boldsymbol{\eta}} \right\| = 0\,. \tag{3.31}$$

The condition in (3.30) is established by showing that $\partial^2\{Q_n^{(1)}(\boldsymbol{\eta}_1)\}/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top$ converges in probability to $\mathbf{B}$, as defined in (3.23), and that for each $i = 2, 3$ and 4 the corresponding Hessian is proportional to $\partial^2\{Q_n^{(1)}(\boldsymbol{\eta}_1)\}/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top$ with probability approaching one. The proof of (3.30) is fairly conventional, whereas the proof of (3.31) – which implicitly invokes the Cramér-Wold device since the moments (cumulants) of the asymptotically normal gradient vector are convergence determining for the limiting distributions in Theorem 3.3 – is more involved because of the presence of the scaling factor R_n. In Appendix 3.A we present the steps necessary to prove (3.30) and (3.31) for each estimator, and for TDGPs with fractional indices in the range $-0.5 < d_0 < 0.5$.

Finally, we highlight the fact that the FML and Whittle estimators are mean invariant by virtue of being defined on the non-zero fundamental Fourier frequencies. As a consequence, all convergence results presented in both this and the previous section for the FML and Whittle estimators also hold for a process that has an arbitrary (non-zero) mean, which may be unknown, thereby broadening the applicability of the theoretical results as they pertain to these particular estimators. The same is not true, however, either for the two time domain based methods or for the exact (integral-based) Whittle estimator, as will be demonstrated in Section 3.5.4 below.

3.5 Finite sample performance of the mis-specified parametric estimators of the pseudo-true parameter

3.5.1 Experimental design

In this section we explore the finite sample performance of the alternative methods, as it pertains to estimation of the pseudo-true value of the long memory parameter, d_1, under specific types of mis-specification. We give particular focus to the four estimators: $\widehat{d}_1^{(1)}$ (FML), $\widehat{d}_1^{(2)}$ (Whittle), $\widehat{d}_1^{(3)}$ (TML) and $\widehat{d}_1^{(4)}$ (CSS), for which the proceeding theoretical results have been produced. We first document the form of the finite sample distributions for each estimator by plotting the distribution of the standardized versions of the estimators, for which the asymptotic distributions are given in Cases 1, 2 and 3 respectively in Theorem 3.3. As part of this exercise we develop a method for obtaining the limiting distribution for $d^* > 0.25$, as the distribution does not have a closed form in this case, as well as a method for estimating the bias-adjustment term, $\boldsymbol{\mu}_n$, which is relevant for this distribution. In the figures that follow the 'Limit' curve depicts the limiting distribution of the relevant statistic. Supplementing these graphical results, we then tabulate the bias and MSE of the four different techniques, as estimators of the pseudo-true parameter d_1, again under specific types of mis-specification and, hence, for different values of d^*. These results are supplemented by bias and MSE values for the exact Whittle estimator, which we refer to as $\widehat{d}_1^{(5)}$.

Data are simulated from a zero-mean Gaussian $ARFIMA(p_0, d_0, q_0)$ process, with the method of Sowell (1992), as modified by Doornik and Ooms (2003), used to compute the exact autocovariance function for the TDGP for any given values of p_0, d_0 and q_0. We have produced results for $n = 100$, 200, 500 and 1000 and for two versions of mis-specification nested in the general case for which the analytical results are derived in Section 3.3. However, we report selected results (only) from the full set due to space constraints. The bias and MSE results, plus certain computations needed for the numerical specification for the limiting distribution in the $d^* > 0.25$ case, are produced from $R = 1000$ replications of samples of size n from the relevant TDGP. The two forms of mis-specification considered are:

Example 3.1 *: An ARFIMA$(0, d_0, 1)$ TDGP, with parameter values $d_0 = \{-0.2, 0.2, 0.4\}$ and $\theta_0 = \{-0.7, -0.444978, -0.3\}$; and an ARFIMA$(0, d, 0)$ MisM. The value $\theta_0 = -0.7$ corresponds to the case where $d^* > 0.25$ and $\widehat{d}_1^{(i)}$, $i = 1, 2, 3, 4$, have the slowest rate of convergence, $n^{1-2d^*} / \log n$, and to a non-Gaussian distribution. The value $\theta_0 = -0.444978$ corresponds to the case where $d^* = 0.25$, in which case asymptotic Gaussianity is preserved but the rate of convergence is of order $(n / \log^3 n)^{1/2}$. The value $\theta_0 = -0.3$ corresponds to the case where $d^* < 0.25$, with $\sqrt{n}$-convergence to Gaussianity obtaining.*

Example 3.2 *: An ARFIMA$(0, d_0, 1)$ TDGP, with parameter values $d_0 = \{-0.2, 0.2, 0.4\}$ and $\theta_0 = \{-0.7, -0.637014, -0.3\}$; and an ARFIMA$(1, d, 0)$ MisM. In this example the value $\theta_0 = -0.7$ corresponds to the case where $d^* > 0.25$, the value $\theta_0 = -0.637014$ corresponds to the case where $d^* = 0.25$, and the value $\theta_0 = -0.3$ corresponds to the case where $d^* < 0.25$.*

In Section 3.5.2 we document graphically the form of the finite sampling distributions of all four estimators of d under the first type of mis-specification described above, and for $d_0 = 0.2$ only. The corresponding graphs under the different values of d_0 (and for all three cases) are qualitatively equivalent to those reported here and, hence, are not included.[8] In Section 3.5.3 we then report the bias and MSE of all four estimators (in terms of estimating the pseudo-true value d_1) under both forms of mis-specification and for all three values of d_0. To supplement

[8]These additional graphical results are presented in Appendix 3.C.

these results, all of which are based on the assumption that the mean is known, in Section 3.5.4 we reproduce corresponding bias and MSE results for the two time domain estimators, plus the exact Whittle estimator, using data in which the unknown mean is estimated along with the other parameters.

3.5.2 Finite sample distributions

In this section we consider in turn the three cases listed under Theorem 3.3. For notational ease and clarity we use $\widehat{d_1}$ to denote the (generic) estimator obtained under mis-specification, remembering that this estimator may be produced by any one of the four estimation methods: $\widehat{d}_1^{(1)}$ to $\widehat{d}_1^{(4)}$. Similarly, we use $Q_n(\cdot)$ to denote the criterion associated with a generic estimator. Only when contrasting the (finite sample) performances of the alternative estimators do we re-introduce the superscript notation.

Case 1: $d^* > 0.25$

The limiting distribution for $\widehat{d_1}$ in this case is

$$\frac{n^{1-2d^*}}{\log n}\left(\widehat{d_1} - d_1 - \mu_n\right) \xrightarrow{D} b^{-1}\sum_{j=1}^{\infty} W_j\,, \tag{3.32}$$

where $\mu_n = b^{-1}E_0\left(\frac{\partial Q_n(\eta_1)}{\partial d}\right)$,

$$\begin{aligned} b &= -2\frac{\sigma_0^2}{2\pi}\int_{-\pi}^{\pi}\frac{f_0(\lambda)}{f_1^3(\boldsymbol{\eta}_1,\lambda)}\left(\frac{\partial f_1(\boldsymbol{\eta}_1,\lambda)}{\partial d}\right)^2 d\lambda + \frac{\sigma_0^2}{2\pi}\int_{-\pi}^{\pi}\frac{f_0(\lambda)}{f_1^2(\boldsymbol{\eta}_1,\lambda)}\frac{\partial^2 f_1(\boldsymbol{\eta}_1,\lambda)}{\partial d^2}d\lambda \\ &= -2\int_0^{\pi}(1+\theta_0^2+2\theta_0\cos(\lambda))(2\sin(\lambda/2))^{-2d^*}(2\log(2\sin(\lambda/2)))^2 d\lambda\,, \end{aligned} \tag{3.33}$$

and $W_j = \frac{(2\pi)^{1-2d^*}(1+\theta_0^2)}{j^{2d^*}}\left[U_j^2 + V_j^2 - E_0(U_j^2+V_j^2)\right]$, with $\{U_j\}$ and $\{V_k\}$ as defined in Theorem 3.3. (With reference to Theorem 3.3, both $\mathbf{B}$ and $\boldsymbol{\mu}_n$ in 3.24 are here scalars since in Example 1 there is only one parameter to estimate under the MisM, namely d. Hence the obvious changes made to notation. All other notation is as defined in the theorem.)

Given that the distribution in (3.32) is non-standard and does not have a closed form representation, consideration must be given to its numerical evaluation. In finite samples the bias-adjustment term μ_n (which approaches zero as $n \to \infty$) also needs to be calculated. We tackle each of these issues in turn, beginning with the computation of μ_n.

(1) From Theorem 3.3 it is apparent that in general the formula for $\mathbf{B}$ is independent of the estimation method, but the calculation of $\boldsymbol{\mu}_n$ requires separate evaluation of $E_0(\partial Q_n(\boldsymbol{\eta}_1)/\partial\boldsymbol{\eta})$ for each estimator. In Appendix 3.B we provide expressions for $E_0(\partial Q_n(\boldsymbol{\eta}_1)/\partial\boldsymbol{\eta})$ for each of the four estimation methods. These formulae are used to evaluate the scalar μ_n here. Each value is then used in the specification of the standardized estimator $\frac{n^{1-2d^*}}{\log n}\left(\widehat{d}_1 - d_1 - \mu_n\right)$ in the simulation experiments.

(2) Quantification of the distribution of $\sum_{j=1}^{\infty} W_j$ requires the approximation of the infinite sum of the W_j, plus the use of simulation to represent the (appropriately truncated) sum. We truncate the series $\sum_{j=1}^{\infty} W_j$ after s terms where the truncation point s is chosen such that $1 \leqslant s < \lfloor n/2 \rfloor$ with $s \to \infty$ as $n \to \infty$ (*cf.* Lemma 6 of Chen and Deo, 2006). The value of s is determined using the following criterion function. Let

$$S_n = \widehat{Var}_0\left[\frac{n^{1-2d^*}}{\log n}\left(\widehat{d}_1 - d_1 - \mu_n\right)\right] \tag{3.34}$$

denote the empirical finite sample variation observed across the R replications and for each m, $1 \leqslant m < \lfloor n/2 \rfloor$, let

$$T_m = S_n - b^{-2}\Omega_m,$$

where $\Omega_m = Var_0\left(\sum_{j=1}^{m} W_j\right)$. Now set

$$s = \arg\min_{1 \leqslant m < \lfloor n/2 \rfloor} T_m. \tag{3.35}$$

Given s, we generate random draws of $\sum_{j=1}^{s} W_j$ via the underlying Gaussian random variables from which the W_j are constructed, and produce an estimate of the limiting distribution using kernel methods.

To determine s we need to evaluate

$$Var_0\left(\sum_{j=1}^{m} W_j\right) = \sum_{j=1}^{m} Var_0\left(W_j\right) + 2\sum_{j=1}^{m}\sum_{\substack{k=1\\ j\neq k}}^{m} Cov_0\left(W_j, W_k\right) \ . \tag{3.36}$$

The variance of W_j in this case is

$$\begin{aligned} & Var_0\left\{\frac{(2\pi)^{1-2d^*}(1+\theta_0^2)}{j^{2d^*}}\left[U_j^2+V_j^2-E_0\left(U_j^2+V_j^2\right)\right]\right\} \\ &= \frac{(2\pi)^{2-4d^*}(1+\theta_0^2)^2}{j^{4d^*}}\left\{E_0\left(U_j^2+V_j^2\right)^2-\left[E_0\left(U_j^2+V_j^2\right)\right]^2\right\}. \end{aligned}$$

As $\{U_j\}$ and $\{V_k\}$ are normal random variables with a covariance structure as specified in Theorem 3.3, standard formulae for the moments of Gaussian random variables yield the result that

$$\begin{aligned} E_0\left(U_j^2+V_j^2\right)^2 &= E_0\left(U_j^4\right)+2E_0\left(U_j^2V_j^2\right)+E_0\left(V_j^4\right) \\ &= 3\left[Var_0\left(U_j\right)\right]^2+2\left[Var_0\left(U_j\right)Var_0\left(V_j\right)+2Cov_0(U_j,V_j)\right] \\ &\quad +3\left[Var_0\left(V_j\right)\right]^2 \\ &= 12\left[Var_0\left(U_j\right)\right]^2 \end{aligned}$$

and

$$\begin{aligned} \left[E_0\left(U_j^2+V_j^2\right)\right]^2 &= \left[E_0\left(U_j^2\right)+E_0\left(V_j^2\right)\right]^2 \\ &= \left[Var_0\left(U_j\right)+Var_0\left(V_j\right)\right]^2 \\ &= 4\left[Var_0\left(U_j\right)\right]^2. \end{aligned}$$

Thus,

$$Var_0(W_j) = \frac{8\,(2\pi)^{2-4d^*}\,(1+\theta_0^2)^2}{j^{4d^*}}\left[Var_0\,(U_j)\right]^2 .$$

Similarly, the covariance between W_j and W_k when $j \neq k$ can be shown to be equal to

$$\begin{aligned}\frac{(2\pi)^{2-4d^*}\,(1+\theta_0^2)^2}{(jk)^{2d^*}}&Cov_0\left(U_j^2+V_j^2, U_k^2+V_k^2\right)\\ &=\frac{4\,(2\pi)^{2-4d^*}\,(1+\theta_0^2)^2}{(jk)^{2d^*}}\left[Var_0\,(U_j)\;Var_0\,(V_k)+2Cov_0(U_j,V_k)\right] .\end{aligned}$$

The expression in (3.36) can therefore be evaluated numerically using the formula for $Cov_0(U_j, V_k)$ to calculate the necessary moments required to determine s from (3.35).

The idea behind the use of T_m is simply to minimize the difference between the second-order sample and population moments. The value of S_n in (3.34) will vary with the estimation method of course; however, we choose s based on S_n calculated from the FML estimates and maintain this choice of s for all other methods. The terms in (3.36) are also dependent on the form of both the TDGP and the MisM and hence T_m needs to be determined for any specific case. The values of s for the sample sizes used in the particular simulation experiment underlying the results in this section are provided in Table 3.1.

Table 3.1: Truncation values s corresponding to the case d^*= 0.3723 for the Example 1: ARFIMA $(0, d_0, 1)$ TDGP with d_0= 0.2 and θ_0= -0.7 vis-à-vis MisM: ARFIMA $(0, d, 0)$.

n	100	200	500	1000
s	36	75	162	230

Each panel in Figure 3.3 provides the kernel density estimate of $\frac{n^{1-2d^*}}{\log n}(\widehat{d_1} - d_1 - \mu_n)$ under the four estimation methods, for a specific n as labeled above each plot, plus the limiting distribution for the given s. The particular parameter values employed in the specification of the TGDP are $d_0 = 0.2$ and $\theta_0 = -0.7$, with $d^* = 0.3723$ in this case, and the values

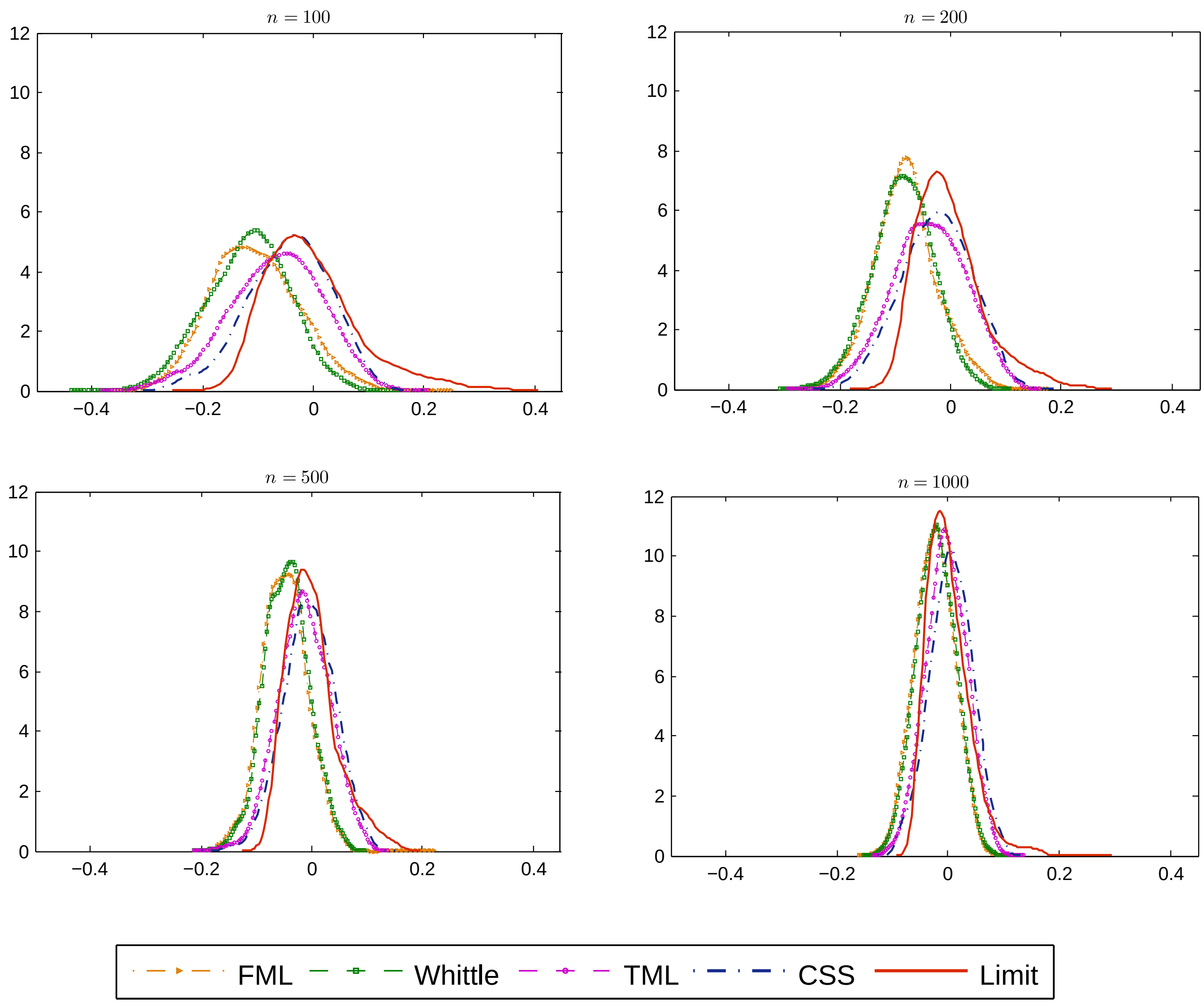

Figure 3.3: Kernel density of $\frac{n^{1-2d^*}}{\log n}\left(\widehat{d}_1 - d_1 - \mu_n\right)$ for an $ARFIMA(0, d_0, 1)$ TDGP with $d_0 = 0.2$ and $\theta_0 = -0.7$, and an $ARFIMA(0, d, 1)$ MisM; $d^* > 0.25$.

of s used are those given in Table 3.1. From Figure 3.3 we see that $\frac{n^{1-2d^*}}{\log n}(\widehat{d}_1 - d_1 - \mu_n)$ is centered away from zero for all sample sizes, for all estimation methods. However, as the sample size increases the point of central location of $\frac{n^{1-2d^*}}{\log n}(\widehat{d}_1 - d_1 - \mu_n)$ approaches zero and all distributions of the standardized statistics go close to matching the asymptotic ('limit') distributions. The salient feature to be noted is the clustering that occurs, in particular for $n \leqslant 500$; that is, TML and CSS form one cluster and FML and Whittle form the other, with the time domain estimators being closer to the asymptotic distribution for all three (smaller) sample sizes.

Case 2: $d^* = 0.25$

The limiting distribution for $\widehat{d}_1$ in the case of $d^* = 0.25$ is

$$n^{1/2}[\overline{\Lambda}_{dd}]^{-1/2}\left(\widehat{d}_1 - d_1\right) \xrightarrow{D} N(0, b^{-2}), \tag{3.37}$$

where

$$\overline{\Lambda}_{dd} = \frac{1}{n}\sum_{j=1}^{n/2}(1+\theta_0^2+2\theta_0\cos(\lambda_j))^2(2\sin(\lambda_j/2))^{-1}(2\log(2\sin(\lambda_j/2)))^2 \tag{3.38}$$

and b is as in (3.33). In both (3.38) and (3.33) $\theta_0 = -0.444978$, as $d^* = 0.25$ occurs at this particular value. Once again, $d_0 = 0.2$ in the TDGP.

Each panel of Figure 3.4 provides the densities of $n^{1/2}[\overline{\Lambda}_{dd}]^{-1/2}\left(\widehat{d}_1 - d_1\right)$ under the four estimation methods, for a specific n as labeled above each plot, plus the limiting distribution given in (3.37). Once again we observe a disparity between the time domain and frequency domain kernel estimates, with the pair of time domain methods yielding finite sample distributions that are closer to the limiting distribution, for all sample sizes considered. The discrepancy between the two types of methods declines as the sample size increases, with the distributions of all methods being reasonably close both to one another, and to the limiting distribution, when $n = 1000$.

Case 3: $d^* < 0.25$

In this case we have

$$\sqrt{n}\left(\widehat{d}_1 - d_1\right) \xrightarrow{D} N(0, v^2), \tag{3.39}$$

where

$$v^2 = \Lambda_{11}/b^{-2}, \tag{3.40}$$

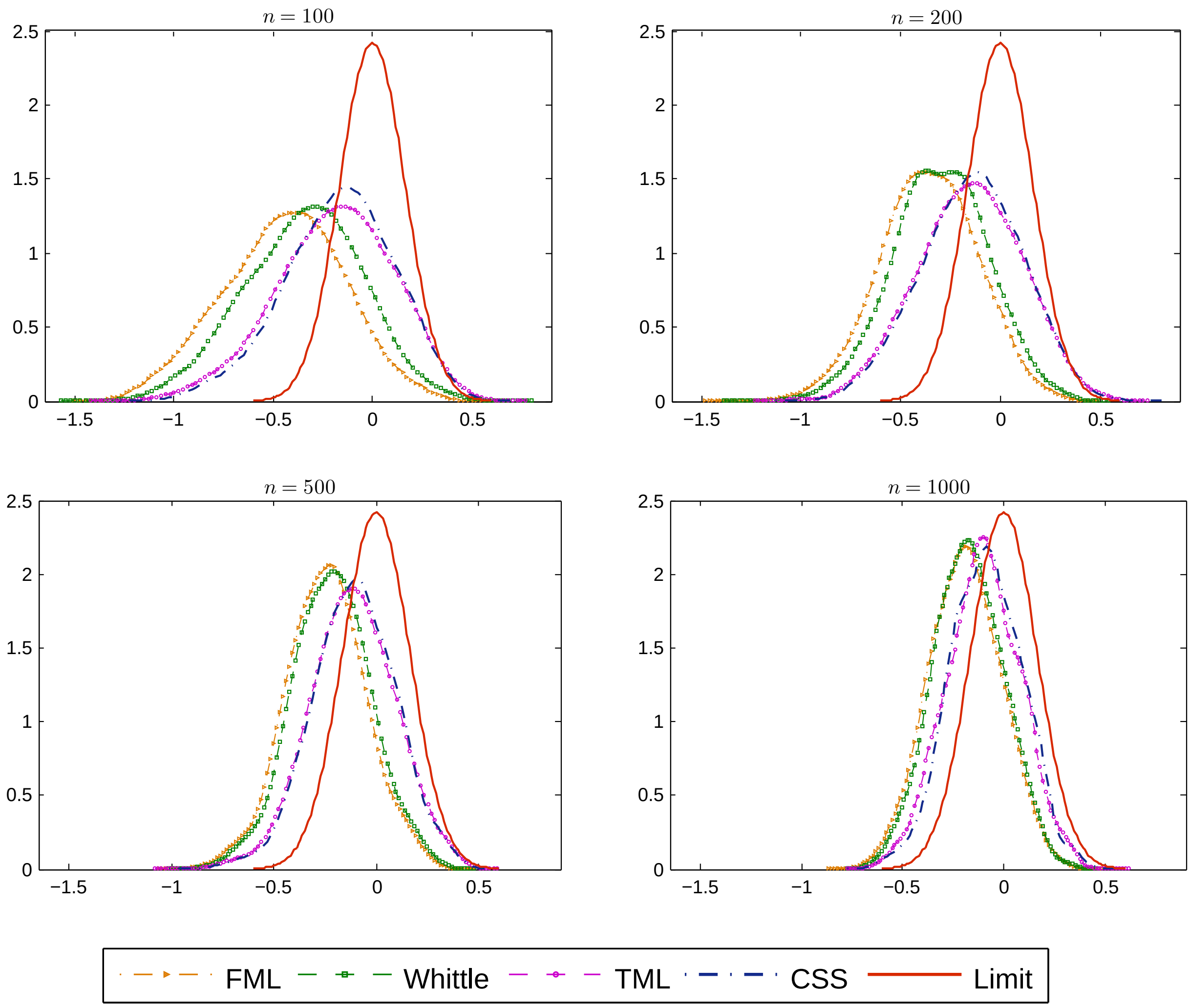

Figure 3.4: Kernel density of $n^{1/2}[\overline{\Lambda}_{dd}]^{-1/2}\left(\widehat{d}_1 - d_1\right)$ for an $ARFIMA(0,d_0,1)$ TDGP with d_0= 0.2 and θ_0= −0.444978, and an $ARFIMA(0,d,1)$ MisM; d^*= 0.25.

with

$$\begin{aligned}\Lambda_{11} &= 2\pi\frac{\sigma_0^2}{2\pi}\int_0^\pi\left(\frac{f_0(\lambda)}{f_1(d_1,\lambda)}\right)^2\left(\frac{\partial\log f_1(d_1,\lambda)}{\partial d}\right)^2 d\lambda\\ &= 2\pi\int_0^\pi(1+\theta_0^2+2\theta_0\cos(\lambda))^2(2\sin(\lambda/2))^{-4d^*}(2\log(2\sin(\lambda/2)))^2 d\lambda\,,\end{aligned}$$

and b as given in (3.33) evaluated at $\theta_0 = -0.3$ and $d^* = 0.1736$. Each panel in Figure 3.5 provides the kernel density estimate of the standardized statistic $\sqrt{n}(\widehat{d}_1 - d_1)$, under the four estimation methods, for a specific n as labeled above each plot, plus the limiting distribution

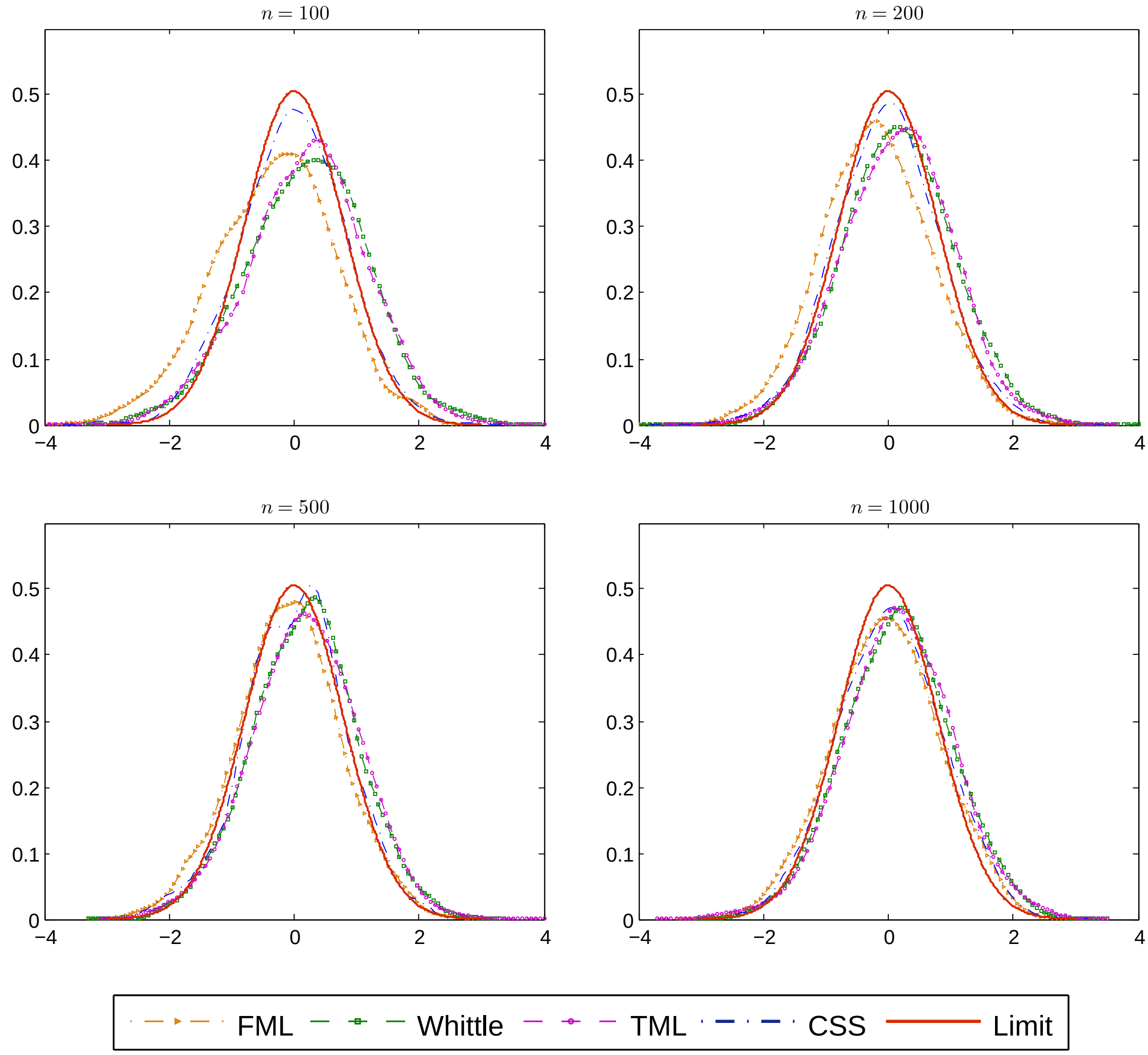

Figure 3.5: Kernel density of $\sqrt{n}\left(\widehat{d}_1 - d_1\right)$ for an $ARFIMA(0, d_0, 1)$ TDGP with $d_0 = 0.2$ and $\theta_0 = -0.3$, and an $ARFIMA(0, d, 1)$ MisM; $d^* < 0.25$.

given in (3.39). In this case there is no clear visual differentiation between the time domain and frequency domain methods, for any sample size, and perhaps not surprisingly given the faster convergence rate in this case, all the methods produce finite sample distributions that match the limiting distribution reasonably well by the time $n = 1000$.

The characteristics observed with the finite sampling distributions of the four parametric estimators for the case with $d_0 = 0.2$ in the TDGP are very similar to the cases with the other two values of $d_0 = -0.2$ (refer Figures 3.6 – 3.8) and 0.4 (refer Figures 3.9 – 3.11).

3.5.3 Finite sample bias and MSE of estimators of the pseudo-true parameter d_1 : known mean case

We supplement the graphical results in the previous section by documenting the finite sample bias and MSE of the four alternative estimators discussed in the previous section, in addition to the exact Whittle estimator, as estimators of the pseudo-true parameter d_1. The following standard formulae,

$$\begin{aligned}
\widehat{\text{Bias}}_0\left(\widehat{d}_1^{(i)}\right) &= \frac{1}{R}\sum_{r=1}^{R}\widehat{d}_r^{(i)} - d_1 && (3.41)\\
\widehat{Var}_0\left(\widehat{d}_1^{(i)}\right) &= \frac{1}{R}\sum_{r=1}^{R}\left(\widehat{d}_{1,r}^{(i)}\right)^2 - \left(\frac{1}{R}\sum_{r=1}^{R}\widehat{d}_{1,r}^{(i)}\right)^2 && (3.42)\\
\widehat{\text{MSE}}_0\left(\widehat{d}_1^{(i)}\right) &= \widehat{\text{Bias}}_0^2 + \widehat{Var}_0\left(\widehat{d}_1^{(i)}\right) && (3.43)\\
\widehat{r.eff}_0\left(\widehat{d}_1^{(i)},\widehat{d}_1^{(j)}\right) &= \frac{\widehat{\text{MSE}}_0\left(\widehat{d}_1^{(i)}\right)}{\widehat{\text{MSE}}_0\left(\widehat{d}_1^{(j)}\right)}, && (3.44)
\end{aligned}$$

are applied to all five estimators $i, j = 1, ..., 5$. Since all empirical expectations and variances are evaluated under the TDGP, we make this explicit with appropriate subscript notation. Results for known mean are produced for Example 1 and Example 2 in Table 3.2 and 3.3 respectively, with selected additional results relevant to both examples recorded in Table 3.4. Values of $d^* = d_0 - d_1$ are documented across the key ranges, $d^* \lesseqgtr 0.25$, along with associated values for the MA coefficient in the TDGP, θ_0. The minimum values of bias and MSE for each parameter setting are highlighted in bold face in all tables for each sample size, n.[9] Corresponding results for the case in which the mean is estimated are recorded in Section 3.5.4.

Consider first the bias and MSE results for Example 1 with $d_0 = 0.2$, as displayed in the middle panel of Table 3.2. As is consistent with the theoretical results (and the graphical

[9] Only that number which is smallest at the precision of 8 decimal places is bolded. Values highlighted with a '$*$' are equally small to 4 decimal places.

Table 3.2: Estimates of the bias and MSE of $\widehat{d_1}$ for the FML, Whittle, Exact Whittle, TML and CSS estimators corresponding to Example 1 - TDGP: ARFIMA $(0, d_0, 1)$ vis-à-vis MisM: ARFIMA $(0, d, 0)$. Process mean, $\mu = 0$, is known.

				FML		Whittle		Exact Whittle		TML		CSS	
d^*	d_0	θ_0	n	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.3723	-0.2	-0.7	100	-0.2416	0.0698	-0.2085	0.0578	-0.1317	0.0388	-0.2055	0.0542	**-0.0941**	**0.0216**
			500	-0.1332	0.0209	-0.1243	0.0185	-0.1107	0.0128	-0.1175	0.0168	**-0.0631**	**0.0081**
			1000	-0.1053	0.0129	-0.1018	0.0121	-0.1053	0.0115	-0.0984	0.0113	**-0.0465**	**0.0050**
0.2500		-0.44	100	-0.1425	0.0308	-0.1068	0.0227	-0.0823	0.0170	-0.0837	0.0169	**-0.0401**	**0.0109**
			500	-0.0539	0.0048	-0.0502	0.0045	-0.0460	0.0039	-0.0414	0.0037	**-0.0192**	**0.0026**
			1000	-0.0378	0.0024	-0.0371	0.0024	-0.0308	0.0022	-0.0325	0.0021	**-0.0092**	**0.0014**
0.1736		-0.3	100	-0.1071	0.0219	-0.0645	0.0147	-0.0468	0.0156	-0.0464	0.0111	**-0.0215**	**0.0087**
			500	-0.0345	0.0028	-0.0272	0.0025	-0.0154	0.0020	-0.0212	0.0021	**-0.0087**	**0.0018**
			1000	-0.0218	0.0013	-0.0191	0.0012	-0.0167	0.0011	-0.0156	0.0010	**-0.0023**	**0.0009**
0.3723	0.2	-0.7	100	-0.1781	0.0915	-0.2466	0.0691	-0.1968	0.0550	-0.1748	0.0481	**-0.1427**	**0.0315**
			500	-0.1354	0.0211	-0.1308	0.0178	-0.1017	0.0141	-0.0916	0.0138	**-0.0798**	**0.0097**
			1000	-0.1019	0.0141	-0.0996	0.0127	-0.0789	0.0097	-0.0776	0.0103	**-0.0670**	**0.0065**
0.2500		-0.44	100	-0.1515	0.0393	-0.1184	0.0298	-0.0795	0.0192	-0.0650	0.0170	**-0.0577**	**0.0119**
			500	-0.0544	0.0048	-0.0487	0.0042	-0.0307	0.0035	-0.0257	0.0027	**-0.0241**	**0.0021**
			1000	-0.0351	0.0023	-0.0323	0.0021	-0.0233	0.0018	-0.0188	0.0015	**-0.0162**	**0.0012**
0.1736		-0.3	100	-0.1082	0.0217	-0.0712	0.0146	-0.0428	0.0109	-0.0340	0.0095	**-0.0330**	**0.0087**
			500	-0.0318	0.0026	-0.0251	0.0022	-0.0184	0.0018	**-0.0106**	0.0017*	-0.0188	**0.0017**
			1000	-0.0184	0.0011	-0.0149	0.0010	-0.0077	0.0010	**-0.0065**	0.0009*	-0.0180	**0.0009**
0.3723	0.4	-0.7	100	-0.2786	0.0995	-0.2456	0.0724	-0.2165	0.0586	-0.2210	0.0515	**-0.1957**	**0.0489**
			500	-0.1598	0.0213	-0.1287	0.0181	-0.1142	0.0140	-0.1347	0.0137	**-0.0871**	**0.0118**
			1000	-0.1123	0.0157	-0.0939	0.0143	-0.0821	0.0113	-0.0812	0.0121	**-0.0648**	**0.0117**
0.2500		-0.44	100	-0.1903	0.0475	-0.1659	0.0383	-0.1086	0.0214	-0.0911	0.0201	**-0.0550**	**0.0138**
			500	-0.0796	0.0095	-0.0550	0.0082	-0.0303	0.0062	-0.0249	0.0059	**-0.0224**	**0.0038**
			1000	-0.0360	0.0048	-0.0295	0.0042	-0.0202	0.0038	-0.0180	0.0035	**-0.0175**	**0.0025**
0.1736		-0.3	100	-0.0990	0.0228	-0.0843	0.0152	-0.0509	0.0122	-0.0422	0.0102	**-0.0321**	**0.0092**
			500	-0.0407	0.0031	-0.0276	0.0025	-0.0188	0.0025	-0.0129	0.0022	**-0.0087**	**0.0019**
			1000	-0.0172	0.0011	-0.0163	0.0010	-0.0089	0.0009	-0.0077	0.0009	**-0.0052**	**0.0008**

illustration in the previous section) the bias and MSE of the four parametric estimators FML, Whittle, TML and CSS, show a clear tendency to decline as the sample size increases, for a fixed value of θ_0. In addition, as θ_0 declines in magnitude, and the MisM becomes closer to the TDGP, there is a tendency for the MSE values and the absolute values of the bias to decline. Importantly, the bias is *negative* for all four estimators, with the (absolute) bias of the two frequency domain estimators (FML and Whittle) being larger than that of the two time domain estimators. These results are consistent with the tendency of the standardized sampling distributions illustrated above to cluster, and for the frequency domain estimators to sit further to the left of zero than those of the time domain estimators, at least for the $d^* \geq 0.25$ cases. Again, as is consistent with the theoretical results, the rate of decline in the (absolute) bias and MSE of all estimators, as n increases, is slower for $d^* \geq 0.25$ than for $d^* < 0.25$. The performance of the exact Whittle estimator (as we term it) is (almost) uniformly better than that of the Whittle estimator, but remains inferior to that of the two time domain estimators, with the CSS being clearly the superior estimator overall. The exact Whittle procedure mimics the other four methods in terms of the decline in both (absolute) bias and MSE as n increases, providing numerical evidence that this frequency domain estimator is also consistent for d_1.[10]

As indicated by the results in the bottom panel of Table 3.2 for $d_0 = 0.4$, the impact of an increase in d_0 (for any given value of d^* and n) is to often (but not uniformly) increase the bias and MSE of all (five) estimator of d_1. That is, the ability of the estimators to accurately estimate the pseudo-true parameter tends to decline (overall) as the long memory in the TDGP increases. In contrast, and with reference to the results in the top panel of the table for the

[10]The tendency for the exact Whittle estimator to have (in particular) smaller finite sample bias than its inexact counterpart would seem to confirm the speculation in Chen and Deo (2006, §2.8) that the bias term (μ_n in Theorem 3) may converge to zero more quickly for the first estimator than for the second.

antipersistent case, estimation accuracy tends to increase (as a general rule) as the memory in the TDGP declines. Nevertheless, the results in Table 3.2 show that the *relativities* between the estimators remain essentially the same for the different values of d_0, with the CSS estimator remaining preferable overall to all other estimators under mis-specification, and the FML estimator performing the worst of all.[11]

The results recorded in Table 3.3 for Example 2 illustrate that the presence of an AR term in the MisM means that more severe mis-specification can be tolerated. More specifically, in all (comparable) cases and for all estimators, the finite sample bias and MSE recorded in Table 3.3 tend to be smaller in (absolute) value than the corresponding values in Table 3.2. Results not presented here suggest, however, that when the value of θ_0 is near zero, estimation under the MisM with an extraneous AR parameter causes an increase in (absolute) bias and MSE, relative to the case where the MisM is fractional noise (see also the following remark). With due consideration taken of the limited nature of the experimental design, these results suggest that the inclusion of some form of short memory dynamics in the estimated model – even if those dynamics are not of the correct form – acts as an insurance against more extreme mis-specification, but at the possible cost of a decline in performance when the consequences of mis-specification are not severe.

Remark 3.1 *When the parameter θ_0 of the $ARFIMA(0, d_0, 1)$ TDGP equals zero the TDGP coincides with the $ARFIMA(0, d, 0)$ model and is nested within the $ARFIMA(1, d, 0)$ model. Thus the value $\theta_0 = 0$ is associated with a match between the TDGP and the model, at which point $d^* = 0$ and there is no mis-specification. That is, neither the $ARFIMA(0, d, 0)$ model estimated in Example 1, nor the*

[11] A slight caveat to this statement is that the superiority of the TML estimator over the exact Whittle is slightly less uniform when either $d = -0.2$ or $d = 0.4$. The overall similarity of the performance of these two estimators across the whole parameter space is, however, not surprising.

Table 3.3: Estimates of the bias and MSE of $\widehat{d}_1$ for the FML, Whittle, Exact Whittle, TML and CSS estimators corresponding to Example 2 - TDGP: ARFIMA $(0, d_0, 1)$ vis-à-vis MisM: ARFIMA $(1, d, 0)$. Process mean, $\mu = 0$, is known.

				FML		Whittle		Exact Whittle		TML		CSS	
d^*	d_0	θ_0	n	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2915	-0.2	-0.7	100	-0.1174	0.0503	-0.0809	0.0216	-0.0756	0.0197	-0.0632	0.0163	**-0.0526**	**0.0138**
			500	-0.0210	0.0115	-0.0179	0.0109	-0.0153	0.0098	-0.0134	0.0085	**-0.0122**	**0.0082**
			1000	-0.0105	0.0054	-0.0088	0.0049	-0.0079	0.0043	-0.0070	0.0044	**-0.0053**	**0.0042**
0.25		-0.64	100	-0.0868	0.0198	-0.0770	0.0153	-0.0641	0.0129	-0.0507	0.0106	**-0.0420**	**0.0096**
			500	-0.0274	0.0039	-0.0167	0.0026	-0.0135	0.0031	**-0.0113**	**0.0021**	-0.0121	0.0023
			1000	-0.0166	0.0012	-0.0128	0.0011	-0.0101	0.0010	-0.0094	0.0010	**-0.0085**	**0.0010**
0.0148		-0.3	100	-0.0320	0.0123	-0.0242	0.0099	-0.0192	0.0103	-0.0156	0.0069	**-0.0095**	**0.0065**
			500	-0.0161	0.0025	-0.0137	0.0031	-0.0089	0.0029	-0.0064	0.0013	**-0.0030**	**0.0013**
			1000	-0.0094	0.0009	-0.0075	0.0007	-0.0055	0.0007	-0.0037	0.0007*	**-0.0011**	**0.0007**
0.2915	0.2	-0.7	100	-0.1612	0.0541	-0.1169	0.0342	-0.0982	0.0304	-0.0950	0.0295	**-0.0671**	**0.0236**
			500	-0.0679	0.0165	-0.0604	0.0125	-0.0557	0.0113	-0.0454	0.0110	**-0.0369**	**0.0089**
			1000	-0.0469	0.0089	-0.0432	0.0071	-0.0339	0.0070	**-0.0250**	0.0067	-0.0303	**0.0059**
0.25		-0.64	100	-0.1339	0.0279	-0.0899	0.0175	-0.0764	0.0145	-0.0655	0.0138	**-0.0457**	**0.0110**
			500	-0.0490	0.0041	-0.0415	0.0030	-0.0373	0.0032	-0.0323	0.0026	**-0.0230**	**0.0022**
			1000	-0.0316	0.0019	-0.0281	0.0015	-0.0228	0.0016	-0.0181	0.0013	**-0.0176**	**0.0011**
0.0148		-0.3	100	-0.0508	0.0139	-0.0256	0.0086	-0.0206	0.0075	0.0190	0.0067	**-0.0082**	**0.0054**
			500	-0.0093	0.0027	-0.0080	0.0019	-0.0078	0.0018	0.0073	0.0016	**-0.0004**	**0.0014**
			1000	-0.0036	0.0010	-0.0023	0.0008	-0.0020	0.0007	0.0067	0.0006*	**0.0003**	**0.0006**
0.2915	0.4	-0.7	100	-0.2299	0.0639	-0.1805	0.0419	-0.1483	0.0389	-0.1279	0.0372	**-0.0699**	**0.0140**
			500	-0.1294	0.0197	-0.1039	0.0150	-0.0948	0.0134	-0.0816	0.0126	**-0.0294**	**0.0101**
			1000	-0.1089	0.0125	-0.0632	0.0099	-0.0521	0.0084	-0.0462	0.0081	**-0.0109**	**0.0069**
0.25		-0.64	100	-0.1396	0.0257	-0.0979	0.0155	-0.0804	0.0148	-0.0692	0.0145	**-0.0508**	**0.0103**
			500	-0.0455	0.0065	-0.0342	0.0046	-0.0315	0.0052	-0.0294	0.0041	**-0.0216**	**0.0033**
			1000	-0.0316	0.0027	-0.0192	0.0021	-0.0156	0.0020	-0.0177	0.0018	**-0.0122**	**0.0014**
0.0148		-0.3	100	-0.0650	0.0162	-0.0422	0.0115	-0.0368	0.0105	0.0246	0.0082	**-0.0132**	**0.0067**
			500	-0.0205	0.0042	-0.0133	0.0034	-0.0099	0.0031	0.0079	0.0026	**-0.0035**	**0.0023**
			1000	-0.0136	0.0021	-0.0088	0.0018	-0.0065	0.0015	0.0053	0.0014	**-0.0017**	**0.0013**

Table 3.4: Estimates of the bias and MSE of $\widehat{d_1}$ for the FML, Whittle, Exact Whittle, TML and CSS estimators corresponding to TDGP: ARFIMA $(0, d_0, 0)$ with $d_0 = 0.2$, $d^* = 0.0$ with the known process mean, $\mu = 0$.

	FML		Whittle		Exact Whittle		TML		CSS	
n	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
					Correct $ARFIMA(0,d,0)$model					
100	-0.0502	0.0113	-0.0173	0.0102	-0.0089	0.0099	**0.0066**	**0.0087**	0.0094	0.0096
500	-0.0089	0.0015	-0.0062	0.0014	-0.0033	0.0014	**0.0026**	**0.0013**	0.0031	0.0014
1000	-0.0045	0.0006*	-0.0037	0.0006*	-0.0028	0.0009	**0.0016**	**0.0006**	0.0025	0.0006*
					Over-Parameterized $ARFIMA(1,d,0)$model					
100	-0.0455	0.0177	0.0371	0.0121	-0.0364	0.0122	0.0255	0.0107	**0.0158**	**0.0087**
500	-0.0120	0.0065	0.0091	0.0049	0.0083	0.0039	0.0078	0.0043	**0.0055**	**0.0037**
1000	-0.0074	0.0027	0.0055	0.0021	0.0040	0.0020	0.0034	0.0019	**0.0028**	**0.0016**

$ARFIMA(1,d,0)$ model estimated in Example 2, is mis-specified (according to our definition) when applied to an ARFIMA $(0,d_0,0)$ TDGP, although the $ARFIMA(1,d,0)$ model is incorrect in the sense of being over-parameterized. Table 3.4 presents the bias and MSE observed when there is such a lack of mis-specification. Under the correct specification of the $ARFIMA(0,d,0)$ model the TML estimator is now superior, in terms of both bias and MSE. The relative accuracy of the TML estimator seen here is consistent with certain results recorded in Sowell (1992) and Cheung and Diebold (1994), in which the performance of the TML method (under a known mean, as is the case considered here) is assessed against that of various comparators under correct model specification. For the over-parameterized $ARFIMA(1,d,0)$ model, however, the CSS estimator dominates once more.

The results in Tables 3.2, 3.3 and 3.4 highlight that, in all but one case, the CSS estimator has the smallest MSE of all five estimators under mis-specification, and when there is no mis-specification but the model is over-parameterized, and that this result holds for all sample sizes considered. Indeed, the MSE results indicate that the CSS estimator is between about two and three times as efficient as the FML estimator (in particular) in the region of the parameter space ($d^* \geq 0.25$) in which both (absolute) bias and MSE are at their highest for all estimators. The absolute value of the bias of CSS is also the smallest in the vast majority of such cases, for all values of d^*. This almost universal superiority of the CSS method presumably reflects a certain in-built robustness of least squares methods.

3.5.4 Finite sample bias and MSE of estimators of the pseudo-true parameter d_1 : unknown mean case

Since the FML and Whittle estimators are both mean invariant the results recorded in the previous two sections for these two estimators are applicable to the unknown (zero) mean case without change. What will potentially alter, however, will be the performance of these two frequency domain estimators relative to that of the exact Whittle and time domain estimators when the unknown mean is also estimated, and it is that possibility that we explore in this section.

In Table 3.5 we record the bias and MSE obtained for the exact Whittle, TML and CSS estimators when the true mean of the process, μ, is estimated using the sample mean; for both mis-specified examples, all three values of d_0, and all three sample sizes. Properties observed in the previous section, such as the decline in bias and MSE with an increase in sample size, for a given θ_0, and the overall decline in MSE and (absolute) bias as the estimated model becomes less mis-specified, continue to obtain in Table 3.5. However, the magnitudes of the bias and MSE figures for all three estimators are now virtually always higher than the corresponding figures in Tables 3.2 and 3.3. As a consequence of this, the time domain estimators lose their relative superiority and no longer uniformly dominate the frequency domain techniques. Instead, the Whittle estimator outperforms all four of the other estimators overall (including its

Table 3.5: Estimates of the bias and MSE of $\widehat{d}_1$ for the Exact Whittle, TML and CSS estimators when the mean is also estimated using the sample mean. The values of d^* for each design are given in Table 2.

		Example 1: TDGP: $ARFIMA(0,d_0,1)$ vis-à-vis MisM: $ARFIMA(0,d,0)$							Example 2: TDGP: $ARFIMA(0,d_0,1)$ vis-à-vis MisM: $ARFIMA(1,d,0)$						
			Exact Whittle		TML		CSS			Exact Whittle		TML		CSS	
n	d_0	θ_0	Bias	MSE	Bias	MSE	Bias	MSE	θ_0	Bias	MSE	Bias	MSE	Bias	MSE
100	-0.2	-0.7	-0.1824	0.0468	-0.2139	0.0581	-0.1543	0.0318	-0.7	-0.1078	0.0356	-0.0951	0.0364	-0.0704	0.0269
500			-0.1223	0.0254	-0.1286	0.0193	-0.0965	0.0122		-0.0487	0.0113	-0.0332	0.0101	-0.0265	0.0096
1000			-0.0823	0.0148	-0.1086	0.0133	-0.0814	0.0084		-0.0224	0.0055	-0.0168	0.0048	-0.0099	0.0055
100		-0.44	-0.1115	0.0202	-0.1223	0.0247	-0.0977	0.0186	-0.64	-0.0763	0.0212	-0.0642	0.0184	-0.0523	0.0142
500			-0.0526	0.0051	-0.0542	0.0049	-0.0444	0.0039		-0.0285	0.0066	-0.0173	0.0055	-0.0119	0.0059
1000			-0.0419	0.0025	-0.0399	0.0026	-0.0325	0.0021		-0.0155	0.0029	-0.0116	0.0022	-0.0094	0.0018
100		-0.3	-0.0461	0.0168	-0.0835	0.0165	-0.0734	0.0143	-0.3	-0.0365	0.0128	-0.0250	0.0110	-0.0186	0.0099
500			-0.0233	0.0030	-0.0314	0.0027	-0.0275	0.0024		-0.0125	0.0031	-0.0127	0.0023	-0.0092	0.0020
1000			-0.0170	0.0016	-0.0209	0.0013	-0.0184	0.0011		-0.0092	0.0013	-0.0072	0.0011	-0.0045	0.0009
100	0.2	-0.7	-0.2850	0.0970	-0.2765	0.0877	-0.2527	0.0739	-0.7	-0.1567	0.0547	-0.1478	0.0501	-0.1216	0.0564
500			-0.1509	0.0398	-0.1425	0.0235	-0.1363	0.0222		-0.0794	0.0222	-0.0650	0.0155	-0.0631	0.0146
1000			-0.1326	0.0222	-0.1210	0.0156	-0.1121	0.0145		-0.0637	0.0096	-0.0562	0.0082	-0.0516	0.0080
100		-0.44	-0.1728	0.0528	-0.1677	0.0462	-0.1527	0.0449	-0.64	-0.1245	0.0430	-0.1171	0.0381	-0.1037	0.0313
500			-0.0748	0.0150	-0.0605	0.0051	-0.0582	0.0049		-0.0510	0.0087	-0.0463	0.0072	-0.0474	0.0068
1000			-0.0433	0.0055	-0.0425	0.0025	-0.0389	0.0024		-0.0319	0.0035	-0.0291	0.0015	-0.0289	0.0015
100		-0.3	-0.1242	0.0188	-0.1168	0.0370	-0.1018	0.0191	-0.3	-0.0585	0.0218	-0.0473	0.0161	-0.0382	0.0153
500			-0.0462	0.0037	-0.0357	0.0026	-0.0300	0.0026		-0.0149	0.0029	0.0085	0.0019	-0.0082	0.0019
1000			-0.0244	0.0018	-0.0166	0.0011	-0.0153	0.0011		-0.0085	0.0012	0.0058	0.0008	-0.0030	0.0008
100	0.4	-0.7	-0.2729	0.0927	-0.2667	0.0838	-0.2597	0.0781	-0.7	-0.2078	0.0819	-0.1992	0.0732	-0.2067	0.0643
500			-0.1462	0.0258	-0.1382	0.0244	-0.1312	0.0199		-0.1321	0.0363	-0.1246	0.0283	-0.1106	0.0278
1000			-0.1166	0.0194	-0.0945	0.0162	-0.1047	0.0155		-0.0832	0.0297	-0.0712	0.0214	-0.0664	0.0201
100		-0.44	-0.1828	0.0578	-0.1750	0.0479	-0.1734	0.0393	-0.64	-0.1141	0.0453	-0.1064	0.0362	-0.1261	0.0324
500			-0.0791	0.0148	-0.0692	0.0108	-0.0585	0.0099		-0.0491	0.0096	-0.0349	0.0071	-0.0388	0.0052
1000			-0.0567	0.0057	-0.0358	0.0056	-0.0285	0.0049		-0.0282	0.0039	-0.0163	0.0016	-0.0202	0.0022
100		-0.3	-0.0903	0.0192	-0.0892	0.0182	-0.0922	0.0183	-0.3	-0.0542	0.0262	-0.0513	0.0172	-0.0583	0.0167
500			-0.0385	0.0028	-0.0325	0.0027	-0.0333	0.0029		-0.0257	0.0054	0.0172	0.0038	-0.0152	0.0036
1000			-0.0232	0.0012	-0.0129	0.0009	-0.0177	0.0010		-0.0178	0.0026	0.0101	0.0018	-0.0090	0.0018

Table 3.6: Estimates of the bias and MSE of $\widehat{d}_1$ for the FML, Whittle, Exact Whittle, TML and CSS estimators corresponding to TDGP: ARFIMA $(0, d_0, 0)$ $d_0 = 0.2$, $d^* = 0.0$. The unknown mean is estimated using the sample mean.

	Exact Whittle		TML		CSS	
n	Bias	MSE	Bias	MSE	Bias	MSE
	Correct $ARFIMA(0, d, 0)$ model					
100	-0.0639	0.0184	-0.0542	0.0149	-0.0585	0.0158
500	-0.0163	0.0036	-0.0111	0.0027	-0.0156	0.0032
1000	-0.0052	0.0018	-0.0047	0.0006	-0.0048	0.0006
	Over-Parameterized $ARFIMA(1, d, 0)$ model					
100	-0.0825	0.0282	-0.0758	0.0224	-0.0701	0.0022
500	-0.0267	0.0081	-0.0195	0.0077	-0.0188	0.0075
1000	-0.0102	0.0031	-0.0087	0.0029	-0.0085	0.0029

exact counterpart), and almost uniformly in the (true) long memory cases ($d_0 = 0.2, 0.4$). Table 3.6 records the outcomes obtained for the exact Whittle, TML and CSS estimators under the correct and over-parameterized specifications when the mean is estimated. Comparing Table 3.6 with Table 3.4 we find (once again) that the Whittle estimator now dominates all other estimators. As the sample size increases the differences between all comparable results for the known and estimated mean cases become less marked, in accordance with the consistency of the estimated mean for the true (zero) mean.[12]

3.6 Discussion

This chapter presents theoretical and simulation-based results relating to the estimation of mis-specified models for fractionally integrated processes. We show that under mis-specification four classical parametric estimation methods, frequency domain maximum likelihood [FML],

[12]The results recorded here regarding the performance of the different estimators in the unknown mean case parallel the qualitative conclusions drawn by Nielsen and Frederiksen (2005) for correctly specified models. Note also that very similar results are obtained if the sample mean is replaced by a feasible (plug in) version of the (asymptotically) best linear unbiased estimator (BLUE) of μ.

Whittle, time domain maximum likelihood [TML] and conditional sum of squares [CSS] converge to the same pseudo-true parameter value. Consistency of the four estimators for the pseudo-true value is proved for fractional exponents of both the true and estimated models in the long memory, short memory and antipersistent ranges. A general closed-form solution for the limiting criterion function for the four alternative estimators is derived in the case of ARFIMA models. This enables us to demonstrate the link between any form of mis-specification of the short memory dynamics and the difference between the true and pseudo-true values of the fractional index, d, and, hence, to the resulting (asymptotic) distributional properties of the estimators, having proved that the estimators are asymptotically equivalent.

The finite sample performance of all four estimators is then documented. The extent to which the finite sample distributions mimic the (numerically specified) asymptotic distributions is displayed. In the case of more extreme mis-specification, and conditional on the mean of the process being known, the pairs of time domain and frequency domain estimators tend to cluster together for smaller sample sizes, with the former pair mimicking the asymptotic distributions more closely. Further, bias and mean squared error [MSE] calculations demonstrate the superiority overall of the CSS estimator, under mis-specification, and the distinct inferiority of the FML estimator – as estimators of the pseudo-true parameter for which they are both consistent. Numerical results for the time domain estimators in the case where the unknown mean is estimated tell a different story, however, with the Whittle estimator being the superior finite sample performer overall. Numerical results presented for an exact version of the Whittle estimator show a slight superiority over the (approximate) version of the Whittle procedure, in the case where the mean is known; however, the overall ranking of the two methods is reversed when the mean is estimated, with the exact Whittle method not sharing the mean-invariance property of its inexact counterpart.

There are several interesting issues that arise from the results that we have established, including the following: First, although the known (zero) mean assumption is inconsequential for the FML and Whittle estimators, this is not the case for the exact Whittle and time domain estimators, as our bias and mean squared error experimental results obtained using demeaned data show. The deterioration in the overall performance of the exact Whittle and time domain estimators once the estimation of μ plays a role in their computation might have been anticipated since the rate of convergence of the sample average to μ is $n^{1/2-d_0}$ (Hosking, 1981, Theorem 8), and thus slower the larger the value of d_0. Similarly, estimation of μ will impact on the limiting distribution of the time domain estimators – because the rate of convergence of the estimators when the true mean is known is $n^{1-2d^*} / \log n$ when $d^* = d_0 - d_1 > 0.25$, $(n/\log\log\log n)^{1/2}$ when $d^* = 0.25$, and $\sqrt{n}$ otherwise – something that we have not pursued for the current chapter, but is the subject of other ongoing research. Second, the extension of our results to non-stationary cases will facilitate the consideration of a broader range of circumstances. To some extent non-stationary values of d might be covered by means of appropriate pre-filtering, for example, the use of first-differencing when $d_0 \in [0.5, 1.5)$, but this would require prior knowledge of the structure of the process and opens up the possibility of a different type of mis-specification from the one we have considered here. Explicit consideration of the interval $d \in [0, 1.5)$, say, allowing for both stationary and non-stationary cases perhaps offers a better approach as prior knowledge of the characteristics of the process would then be unnecessary. The latter also seems particularly relevant given that estimates near the boundaries $d = 0.5$ and $d = 1$ are not uncommon in practice. Previous developments in the analysis of non-stationary fractional processes (see, inter alios, Beran, 1995; Tanaka, 1999; Velasco, 1999a; Velasco and Robinson, 2000) might offer a sensible starting point for such an investigation. Third, our limiting distribution results can be used in practice to conduct inference on the long memory and other parameters after constructing obvious smoothed periodogram consistent estimates of $\mathbf{B}$, $\boldsymbol{\mu}_n$, $\overline{\Lambda}_{dd}$ and $\boldsymbol{\Lambda}$. But which situation should be assumed in any particular instance, $d^* > 0.25$, $d^* = 0.25$ or $d^* < 0.25$, may be a moot point. Fourth, the relationships between the bias and MSE of the parametric estimators of d_1 (denoted respectively below by Bias_d_1 and MSE_d_1), and the bias and MSE as estimators of the *true* value d_0, (Bias_d_0 and MSE_d_0 respectively) can be expressed simply as follows:

$$\begin{aligned} \text{Bias\_}d_0 &= E_0(\widehat{d}_1) - d_0 \\ &= \left[E_0(\widehat{d}_1) - d_1\right] + (d_1 - d_0) \\ &= Bias\_d_1 - d^* , \end{aligned}$$

where we recall, $d^* = d_0 - d_1$, and

$$\begin{aligned} \text{MSE\_}d_0 &= E_0\left(\widehat{d}_1 - d_0\right)^2 \\ &= E_0\left(\widehat{d}_1 - E_0(\widehat{d}_1)\right)^2 + \left[E_0(\widehat{d}_1) - d_0\right]^2 \\ &= E_0\left(\widehat{d}_1 - E_0(\widehat{d}_1)\right)^2 + \left[[E_0(\widehat{d}_1) - d_1] - d^*\right]^2 \\ &= E_0\left(\widehat{d}_1 - E_0(\widehat{d}_1)\right)^2 + \left[E_0(\widehat{d}_1) - d_1\right]^2 + d^{*2} - 2d^*\left[E_0(\widehat{d}_1) - d_1\right] \\ &= MSE\_d_1 + d^{*2} - 2d^* Bias\_d_1. \end{aligned}$$

Hence, if Bias_d_1 is the same sign as d^* at any particular point in the parameter space, then the bias of a mis-specified parametric estimator *as an estimator of* d_0, may be less (in absolute value) than its bias as an estimator of d_1, depending on the magnitude of the two quantities. Similarly, MSE_d_0 may be less than MSE_d_1 if Bias_d_1 and d^* have the same sign, with the final result again depending on the magnitude of the two quantities. These results imply that it is possible for the ranking of mis-specified parametric estimators to be altered, once the reference point changes from d_1 to d_0. This raises the following questions: Does the dominance of the CSS estimator (within the parametric set of estimators) – and in the known mean case – still obtain when the true value of d is the reference value? And more critically from a practical perspective; Are there circumstances where a mis-specified parametric estimator out-performs semi-parametric alternatives in finite samples, the lack of consistency (for d_0) of the former notwithstanding? Such topics remain the focus of current and ongoing research.

3.A Appendix: Proofs

Proof of Lemma 3.1.

The proof of the lemma uses a method that parallels that employed by Fox and Taqqu in the proof of their Lemma 1 (see Fox and Taqqu, 1986, pages 523 − 524), which in turn employs an argument first developed by Hannan in the proof of his Lemma 1 (see Hannan, 1973, pages 133 − 134). To describe the approach, set

$$c_n(\tau) = c_n(-\tau) = \frac{1}{n}\sum_{t=1}^{n-\tau} y_t y_{t+\tau}\,, \quad \tau \geq 0\,,$$

and let

$$k_M(\boldsymbol{\eta}, \lambda) = \sum_{r=-M}^{M} \kappa(r)\left(1 - \frac{|r|}{M}\right)\exp(i\lambda r),$$

denote the Cesaro sum of the first M terms of the Fourier series of $(f_1(\boldsymbol{\eta}, \lambda) + \nu_f)^{-1}$ where M is chosen such that $|(f_1(\boldsymbol{\eta}, \lambda) + \nu_f)^{-1} - k_M(\boldsymbol{\eta}, \lambda))| < \epsilon$ uniformly in $\boldsymbol{\eta} \in \mathbb{E}^0_\delta$. Then following the same steps as in the derivation presented in Hannan (1973, pages 133-134) we have

$$\left| \frac{4\pi}{n}\sum_{j=1}^{[n/2]} I(\lambda_j)\left\{(f_1(\boldsymbol{\eta}, \lambda_j) + \nu_f)^{-1} - k_M(\boldsymbol{\eta}, \lambda_j)\right\}\right| < \epsilon c_n(0)\,,$$

and

$$\lim_{n\to\infty}\left|\frac{4\pi}{n}\sum_{j=1}^{[n/2]} I(\lambda_j)k_M(\boldsymbol{\eta},\lambda_j) - \sum_{r=-M}^{M}\kappa(r)\left(1-\frac{|r|}{M}\right)\gamma_0(r)\right| = 0,$$

almost surely, the latter result since $I(\lambda) = (2\pi)^{-1}\sum_{r=1-n}^{n-1} c_n(r)\exp(-i\lambda r)$ and $c_n(r)$ converges to $\gamma_0(r)$ almost surely by ergodicity. Moreover,

$$\sum_{r=-M}^{M}\kappa(r)\left(1-\frac{|r|}{M}\right)\gamma_0(r) = \frac{\sigma_0^2}{2\pi}\int_{-\pi}^{\pi} f_0(\lambda)k_M(\boldsymbol{\eta},\lambda)d\lambda,$$

differs from the required limiting value by a quantity bounded by $\epsilon\gamma_0(0)$, from which the desired result follows because ϵ is arbitrary.

An alternative proof of this lemma can be obtained by extending the arguments adopted by Brockwell and Davis (1991, §10.8.2, pages 378-379), in the proof of their Proposition 10.8.2, to the stationary fractional case, as suggested in Brockwell and Davis (1991, page 528). ■

Proof of Lemma 3.2. The proof parallels the proof of Lemma 3.1, only now we use the Cesaro sum of M terms of the Fourier series of $h_1(\boldsymbol{\eta},\lambda)^{-1}$. Denote this sum by $c_M(\boldsymbol{\eta},\lambda) > 0$. Since by construction $h_1(\boldsymbol{\eta},\lambda) > 0$, M can be chosen so that $|h_1(\boldsymbol{\eta},\lambda)^{-1} - c_M(\boldsymbol{\eta},\lambda)| < \varepsilon$ uniformly on $\mathbb{E}_\delta^0$ since the Cesaro sum converges uniformly in $(\boldsymbol{\eta},\lambda)$ for $\boldsymbol{\eta}\in\mathbb{E}_\delta^0$. Once again the detailed steps follow Hannan (1973, pages 133-134), as above, or Brockwell and Davis (1991, §10.8.2, pages 378-379). ■

Proof of Lemma 3.3.

Observe that $f_1(\boldsymbol{\eta},\lambda) > 0$ when $d \geq 0$ and hence for δ sufficiently small we have $h_1(\boldsymbol{\eta},\lambda) = f_1(\boldsymbol{\eta},\lambda)$ for all $\lambda\in[-\pi,\pi]$. It follows immediately from Lemma 3.2 that $\lim_{n\to\infty}|Q_n^{(1)}(\boldsymbol{\eta}) - Q(\boldsymbol{\eta})| = 0$ almost surely and uniformly in $\boldsymbol{\eta}$ on $\mathbb{E}_\delta^0$ when $d\geq 0$. We have thus established Lemma 3.3 in the case where $d \geq 0$, (*cf.* Chen and Deo, 2006, Lemma 2). To establish that Lemma 3.3 also holds on $\mathbb{E}_\delta^0$ when $d < 0$, observe that Lemma 3.1 implies that $Q(\boldsymbol{\eta})$ provides a limit inferior for $Q_n^{(1)}(\boldsymbol{\eta})$ and it therefore only remains for us to establish that $Q(\boldsymbol{\eta})$ also provides a limit superior for $Q_n^{(1)}(\boldsymbol{\eta})$ on $\boldsymbol{\eta} \in \mathbb{E}_\delta^0$ when $d < 0$.

In the latter case $f_1(\boldsymbol{\eta}, \lambda) = |\lambda|^{2|d|} L(\lambda)$ where $L(\lambda)$ is slowly varying and bounded as $\lambda \to 0$ and there exists an $\epsilon \in (0, 2|d|)$ and a $K > 0$, that may depend on ϵ, such that $f_1(\boldsymbol{\eta}, \lambda) = |\lambda|^{2|d|} K |\lambda|^{-\epsilon}$. We therefore have that $f_1(\boldsymbol{\eta}, \lambda) > K|\lambda|^{2|d|}$ when $|\lambda| < 1$ and $h_1(\boldsymbol{\eta}, \lambda) \neq f_1(\boldsymbol{\eta}, \lambda)$ whenever $\lambda < (K^{-1}\delta)^{1/(2|d|-\epsilon)}$, from which it follows that

$$Q_n^{(1)}(\boldsymbol{\eta}) \leq \frac{2\pi}{n} \sum_{j=k_\delta+1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{h_1(\boldsymbol{\eta}, \lambda_j)} + \frac{1}{K} \left(\frac{2\pi}{n} \right)^{1-2|d|} \sum_{j=1}^{k_\delta} I(\lambda_j), \tag{3.45}$$

where $k_\delta = \lfloor (K^{-1}\delta)^{1/(2|d|-\epsilon)} (2\pi/n) \rfloor + 1$. The inequality in (3.45) follows because for all $\lambda_j < (K^{-1}\delta)^{1/(2|d|-\epsilon)} < 2\pi k_\delta / n$ we have

$$\left(\frac{h_1(\boldsymbol{\eta}, \lambda_j)}{f_1(\boldsymbol{\eta}, \lambda_j)} - 1 \right) \leq \left(\frac{\delta}{K} \left(\frac{n}{2\pi} \right)^{2|d|} - 1 \right),$$

and

$$\begin{aligned} Q_n^{(1)}(\boldsymbol{\eta}) &= \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{h_1(\boldsymbol{\eta}, \lambda_j)} + \frac{2\pi}{n} \sum_{j=1}^{k_\delta} I(\lambda_j) \left(\frac{1}{f_1(\boldsymbol{\eta}, \lambda_j)} - \frac{1}{h_1(\boldsymbol{\eta}, \lambda_j)} \right) \\ &\leq \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{h_1(\boldsymbol{\eta}_1, \lambda_j)} + \frac{2\pi}{n} \sum_{j=1}^{k_\delta} \frac{I(\lambda_j)}{h_1(\boldsymbol{\eta}, \lambda_j)} \left(\frac{\delta}{K} \left(\frac{n}{2\pi} \right)^{2|d|} - 1 \right) \\ &= \frac{2\pi}{n} \sum_{j=k_\delta+1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{h_1(\boldsymbol{\eta}, \lambda_j)} + \frac{1}{K} \left(\frac{2\pi}{n} \right)^{1-2|d|} \sum_{j=1}^{k_\delta} I(\lambda_j). \end{aligned}$$

Applying Lemma 3.2 to the first term on the right hand side in (3.45) gives a limit of

$$\frac{\sigma_0^2}{2\pi} \int_{(K^{-1}\delta)^{1/(2|d|-\epsilon)}}^{\pi} \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta}_1, \lambda)} d\lambda.$$

Similarly

$$\lim_{n \to \infty} \frac{2\pi}{n} \sum_{j=1}^{k_\delta} I(\lambda_j) = \frac{\sigma_0^2}{2\pi} \int_0^{(K^{-1}\delta)^{1/(2|d|-\epsilon)}} f_0(\lambda) d\lambda = \frac{\sigma_0^2}{2\pi} f_0(\lambda')(K^{-1}\delta)^{1/(2|d|-\epsilon)},$$

for some $\lambda' \in [0,(K^{-1}\delta)^{1/(2|d|-\epsilon)}]$ by the first mean value theorem for integrals. Setting $\delta = (2\pi)^{2|d|-\epsilon}/n^p$ where $p > 2|d|-\epsilon$, we find that

$$\begin{aligned}\frac{1}{K}\left(\frac{2\pi}{n}\right)^{1-2|d|}\sum_{j=1}^{k_\delta} I(\lambda_j) &\sim \frac{1}{K}\left(\frac{n}{2\pi}\right)^{2|d|}\frac{\sigma_0^2}{2\pi}f_0(\lambda')\frac{2\pi k_\delta}{n}\\ &\sim \frac{1}{K}\left(\frac{2\pi}{n}\right)^{1-2|d|}\frac{\sigma_0^2}{2\pi}f_0(\lambda')\left(\frac{1}{n}\right)^{\frac{p-2|d|+\epsilon}{(2|d|-\epsilon)}},\end{aligned}$$

and hence we can conclude that

$$\limsup_{n\to\infty} Q_n^{(1)}(\boldsymbol{\eta}) \le Q(\boldsymbol{\eta}),$$

uniformly in $\boldsymbol{\eta} \in \mathbb{E}_\delta^0$, as required. ■

Proof of Lemma 3.4. Let $L_1(\boldsymbol{\eta},\lambda) = \lambda^{2d} f_1(\boldsymbol{\eta},\lambda)$ and suppose that $\boldsymbol{\eta} \in \overline{\mathbb{E}}_{\delta 1}^0 \cup \overline{\mathbb{E}}_{\delta 2}^0 \neq \emptyset$. Then

$$\begin{aligned}\liminf_{n\to\infty} Q_n^{(1)}(\boldsymbol{\eta}) &= \liminf_{n\to\infty}\frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)}\\ &= \liminf_{n\to\infty}\frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)\lambda_j^{2d}}{L_1(\boldsymbol{\eta},\lambda_j)}\\ &\ge \liminf_{n\to\infty}\frac{(2\pi)^{-2\delta}}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)}{L_1(\boldsymbol{\eta},\lambda_j)\lambda_j^{1-2(d_0+\delta)}},\end{aligned} \tag{3.46}$$

where the inequality in (3.46) arises because for all $\boldsymbol{\eta} \in \overline{\mathbb{E}}_{\delta 1}^0 \cup \overline{\mathbb{E}}_{\delta 2}^0$ we have $(d_0 - d) > 0.5 - \delta$ and it follows that $\lambda_j^{-2(d_0-d)} \ge (2\pi)^{-(2\delta+1)}\lambda_j^{2\delta-1}$ for all $\lambda_j = 2\pi j/n$, $j = 1,\dots,\lfloor n/2\rfloor$.

Applying Lemma 3.1 and Lemma 3.2 to (3.46) by replacing $f_1(\boldsymbol{\eta},\lambda_j)$ by $L_1(\boldsymbol{\eta},\lambda)\lambda^{1-2(d_0+\delta)}$, and then letting the constant $\nu_f > 0$in the lemmas approach zero, it follows from Fatou's theorem that

$$\lim_{n\to\infty}\left|\frac{(2\pi)^{-2\delta}}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)}{L_1(\boldsymbol{\eta},\lambda_j)\lambda_j^{1-2(d_0+\delta)}} - \frac{1}{(2\pi)^{2\delta+1}}\int_0^\pi \frac{(\sigma_0^2/2\pi)f_0(\lambda)\lambda^{2d_0}}{L_1(\boldsymbol{\eta},\lambda)\lambda^{1-2\delta}}d\lambda\right| = 0,$$

wherein we recognize that $0 \le 1-2(d_0+\delta) \le 2(1-2\delta)$ and that

$$L_1(\boldsymbol{\eta},\lambda) = (\sigma_1^2/2\pi)g_1(\boldsymbol{\beta},\lambda)\,\mathrm{sinc}(\lambda/2)^{-2d},$$

and

$$(\sigma_0^2/2\pi)f_0(\lambda)\lambda^{2d_0} = (\sigma_0^2/2\pi)g_0(\lambda)\,\mathrm{sinc}(\lambda/2)^{-2d_0},$$

where $\mathrm{sinc}(x) = \sin(x)/x$, the cardinal sine function. Since $2/\pi \leq \mathrm{sinc}(\lambda/2) \leq 1$ for $0 \leq \lambda \leq \pi$, it follows from Assumption $(A.3)$ and *Conditions A* that there exists a finite positive constant R such that

$$\begin{aligned}
\frac{1}{(2\pi)^{2\delta+1}}\int_0^\pi \frac{(\sigma_0^2/2\pi)f_0(\lambda)\lambda^{2d_0}}{L_1(\boldsymbol{\eta},\lambda)\lambda^{1-2\delta}}d\lambda &\geq \frac{R}{(2\pi)^{2\delta+1}}.\int_0^\pi \lambda^{2\delta-1}d\lambda \\
&= \frac{R}{(2\pi)^{2\delta+1}}.\frac{\pi^{2\delta}}{2\delta} \\
&\geq \frac{R}{8\pi}.\frac{1}{\delta}. \qquad (3.47)
\end{aligned}$$

The statements in Lemma 3.4 now follow from (3.47), directly in the case of $\boldsymbol{\eta} \in \overline{\mathbb{E}}^0_{\delta 1}$, and for $\boldsymbol{\eta} \in \overline{\mathbb{E}}^0_{\delta 2}$ on setting $\delta < R/(8\pi C)$ and letting $\delta \to 0$ as $C \to \infty$. ■

Proof of Proposition 3.1.

Let $\boldsymbol{\eta}_n$ denote a sequence in $\mathbb{E}^0_\delta$ that converges to $\boldsymbol{\eta}$. For any $\nu_f > 0$ we have

$$\begin{aligned}
\left|\frac{1}{f_1(\boldsymbol{\eta}_n,\lambda)+\nu_f} - \frac{1}{f_1(\boldsymbol{\eta},\lambda)+\nu_f}\right| &= \left|\frac{|f_1(\boldsymbol{\eta}_n,\lambda) - f_1(\boldsymbol{\eta},\lambda)|}{(f_1(\boldsymbol{\eta}_n,\lambda)+\nu_f)(f_1(\boldsymbol{\eta},\lambda)+\nu_f)}\right| \\
&\leq \frac{|f_1(\boldsymbol{\eta}_n,\lambda) - f_1(\boldsymbol{\eta},\lambda)|}{\nu_f^2}.
\end{aligned}$$

Moreover, by assumption $f_1(\boldsymbol{\eta},\lambda)$ is continuous for all $\lambda \neq 0$ and hence uniformly continuous for λ in any closed interval of the form $[\varepsilon,\pi]$, $\varepsilon > 0$. Consequently we can determine a value n' such that for $n \geq n'$ there exists an ε sufficiently small that $|f_1(\boldsymbol{\eta}_n,\lambda) - f_1(\boldsymbol{\eta},\lambda)| < \nu_f^3$ and

$$\left|\frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)}{f_1(\boldsymbol{\eta}_n,\lambda_j)+\nu_f} - \frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)+\nu_f}\right| \leq \frac{2\nu_f\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor} I(\lambda_j). \qquad (3.48)$$

Using Lemma 3.1 in conjunction with (3.48), it follows that

$$\liminf_{n\to\infty} Q_n^{(1)}(\boldsymbol{\eta}_n) \;\geq\; \liminf_{n\to\infty}\frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)}{f_1(\boldsymbol{\eta}_n,\lambda_j)+\nu_f}$$

$$
\begin{aligned}
&\geq \lim_{n\to\infty}\left\{\frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)+\nu_f}-\frac{2\nu_f\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}I(\lambda_j)\right\}\\
&= \frac{\sigma_0^2}{2\pi}\int_0^{\pi}\frac{f_0(\lambda)}{f_1(\boldsymbol{\eta},\lambda)+\nu_f}d\lambda-\nu_f\pi\gamma_0(0)\,,
\end{aligned}
$$

where $\gamma_0(0)$ is the variance of the TDGP. Letting $\nu_f \to 0$ and applying Lebegue's monotone convergence theorem gives

$$
\liminf_{n\to\infty} Q_n^{(1)}(\boldsymbol{\eta}_n) \geq \frac{\sigma_0^2}{2\pi}\int_0^{\pi}\frac{f_0(\lambda)}{f_1(\boldsymbol{\eta},\lambda)}d\lambda = Q(\boldsymbol{\eta})\,.
$$

Since by definition $\boldsymbol{\eta}_1$ minimizes $Q(\boldsymbol{\eta})$ it follows that $Q(\boldsymbol{\eta}_1)$ provides a lower bound to the limit inferior of $Q_n^{(1)}(\boldsymbol{\eta}_n)$ for any sequence in $\mathbb{E}_\delta^0$.

Now let $\boldsymbol{\eta}_n$ denote a sequence in $\overline{\mathbb{E}}_{\delta 1}^0 \cup \overline{\mathbb{E}}_{\delta 2}^0$ that converges to $\boldsymbol{\eta}$. Setting

$$
\delta \ll \min\left\{\frac{\sigma_0^2}{4(2\pi)^2}\frac{C_l}{C_u}\frac{1}{(Q(\boldsymbol{\eta}_1)+q)}, 0.25-0.5(d_0-d_1)\right\} \quad \text{where} \quad q \gg 0\,,
$$

and applying Lemma 3.4 in conjunction with (3.48) implies that

$$
\liminf_{n\to\infty} Q_n^{(1)}(\boldsymbol{\eta}_n) \gg Q(\boldsymbol{\eta}_1)+q\,.
$$

Hence we can conclude that for any sequence $\boldsymbol{\eta}_n \in \overline{\mathbb{E}}_{\delta 1}^0 \cup \overline{\mathbb{E}}_{\delta 2}^0$ the criterion value $Q_n^{(1)}(\boldsymbol{\eta}_n)$ will, for all n sufficiently large, exceed $Q(\boldsymbol{\eta}_1)$, which equals $\lim_{n\to\infty} Q_n^{(1)}(\boldsymbol{\eta}_1)$ by Lemma 3.3.

By definition of $\widehat{\boldsymbol{\eta}}_1^{(1)}$, however, $Q_n^{(1)}(\widehat{\boldsymbol{\eta}}_1^{(1)}) \leq Q_n^{(1)}(\boldsymbol{\eta}_1)$ and it follows from Lemma 3.3 that

$$
\limsup_{n\to\infty} Q_n^{(1)}(\widehat{\boldsymbol{\eta}}_1^{(1)}) \leq \limsup_{n\to\infty} Q_n^{(1)}(\boldsymbol{\eta}_1) = Q(\boldsymbol{\eta}_1)\,.
$$

We can therefore conclude that $|Q_n^{(1)}(\widehat{\boldsymbol{\eta}}_1^{(1)}) - Q(\boldsymbol{\eta}_1)| \to 0$ almost surely and an argument by contradiction then shows that $\widehat{\boldsymbol{\eta}}_1^{(1)} \to \boldsymbol{\eta}_1$ with probability one. ■

Proof of Theorem 3.1. In what follows we assume that the mean is known, and without loss of generality set $\mu = 0$ and suppose that the data is mean corrected.

The Whittle estimator:

Concentrating $Q_n^{(2)}(\sigma^2, \boldsymbol{\eta})$ with respect to σ^2 and setting $n \cdot \lfloor n/2 \rfloor = 0.5$ yields the profile (negative) log-likelihood

$$Q_n^{(2)}(\boldsymbol{\eta}) = \frac{2\pi}{2} \log\left(\frac{\widehat{\sigma}^2(\boldsymbol{\eta})}{2\pi}\right) + \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \log f_1(\boldsymbol{\eta}, \lambda_j) + \pi,$$

where $\widehat{\sigma}^2(\boldsymbol{\eta}) = 2Q_n^{(1)}(\boldsymbol{\eta})$ and $Q_n^{(1)}(\boldsymbol{\eta})$ is as given in (3.7). Now, following Beran (1994, page 116), we have

$$\frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \log f_1(\boldsymbol{\eta}, \lambda_j) = \frac{1}{2} \sum_{r=-\infty}^{\infty} \rho_1(\boldsymbol{\eta}, rn) \rightarrow \frac{1}{2} \int_{-\pi}^{\pi} \log f_1(\boldsymbol{\eta}, \lambda) d\lambda\,,$$

where the Fourier coefficients $\rho_1(\boldsymbol{\eta}, r) = \int_{-\pi}^{\pi} \log f_1(\boldsymbol{\eta}, \lambda) \exp(\imath \lambda r) d\lambda$ form a convergent series and

$$\begin{aligned}
\int_{-\pi}^{\pi} \log f_1(\boldsymbol{\eta}, \lambda) d\lambda &= \int_{-\pi}^{\pi} \log\left(g_1(\boldsymbol{\beta}, \lambda) |2\sin(\lambda/2)|^{-2d}\right) d\lambda \\
&= \int_{-\pi}^{\pi} \log g_1(\boldsymbol{\beta}, \lambda) d\lambda - 2d \int_{-\pi}^{\pi} \log|2\sin(\lambda/2)| d\lambda\,.
\end{aligned}$$

By Assumption $(A.2)$ $\int_{-\pi}^{\pi} \log g_1(\boldsymbol{\beta}, \lambda) d\lambda = 0$, and from standard results for trigonometric integrals Gradshtein and Ryzhik (2007, page 583)

$$\int_{-\pi}^{\pi} \log|2\sin(\lambda/2)| d\lambda = 2\int_{0}^{\pi} \log|2\sin(\lambda/2)| d\lambda = 0\,.$$

Furthermore, since $\log f_1(\boldsymbol{\eta}, \lambda)$ is integrable, and continuously differentiable for all $\lambda \neq 0$ by Assumption A.3, $\rho_1(\boldsymbol{\eta}, n) = o(1/n)$, which implies that

$$\frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \log f_1(\boldsymbol{\eta}, \lambda_j) = \sum_{r=1}^{\infty} \rho_1(\boldsymbol{\eta}, rn) = O(n^{-1} \log n)\,.$$

Hence it follows that

$$\left| Q_n^{(2)}(\boldsymbol{\eta}) - \pi \log Q_n^{(1)}(\boldsymbol{\eta}) - \pi(\log \pi + 1) \right| = O(n^{-1} \log n)\,, \tag{3.49}$$

almost surely and uniformly in $\boldsymbol{\eta}$. From this we can deduce that

$$\lim_{n\to\infty}\left|Q_n^{(2)}(\widehat{\boldsymbol{\eta}}_1^{(2)})-\pi\log Q_n^{(1)}(\widehat{\boldsymbol{\eta}}_1^{(1)})-\pi(\log\pi+1)\right|=0\quad a.s.,$$

where $\widehat{\boldsymbol{\eta}}_1^{(1)}$ is the value of $\boldsymbol{\eta}$ that minimizes the profile log-likelihood, *having first deleted the term* $2\pi\sum_{j=1}^{\lfloor n/2\rfloor}\log f_1(\boldsymbol{\eta},\lambda_j)/n$, namely $\widehat{\boldsymbol{\eta}}_1^{(1)}=\arg\min_{\boldsymbol{\eta}} Q_n^{(1)}(\boldsymbol{\eta})$. We are thereby lead directly to the conclusion that $\widehat{\boldsymbol{\eta}}_1^{(2)}$ and $\widehat{\boldsymbol{\eta}}_1^{(1)}$ converge, *i.e.* $\lim_{n\to\infty}\|\widehat{\boldsymbol{\eta}}_1^{(2)}-\widehat{\boldsymbol{\eta}}_1^{(1)}\|=0$.

The TML estimator:

Using the argument employed by Hannan (1973, pages 134-135) in the proof of his Lemma 4, following the detailed steps given by Brockwell and Davis (1991, §10.8.2, pages 380-382) in their proof of their Proposition 10.8.3, shows that

$$\lim_{n\to\infty}\left|\frac{1}{n}\mathbf{Y}^\top\boldsymbol{\Sigma}_\eta^{-1}\mathbf{Y}-\frac{4\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)}\right|=0\quad a.s.,\tag{3.50}$$

and the convergence is uniform in $\boldsymbol{\eta}$ on $\mathbb{E}_\delta^0$. From a theorem due to Grenander and Szego (1958, Chapter 5) we know that

$$\frac{1}{n}\log|\boldsymbol{\Sigma}_\eta|=\frac{1}{2\pi}\int_{-\pi}^{\pi}\log f_1(\boldsymbol{\eta},\lambda)d\lambda+O(n^{-1}),\tag{3.51}$$

for the second term in (3.11). That the convergence in (3.51) is uniform in $\boldsymbol{\eta}$ is not stated in Grenander and Szego, although it follows from the uniformity of the order relations used in their proof. Their proof depends on approximating $f_1(\boldsymbol{\eta},\lambda)$ by trigonometric polynomials, and since $f_1(\boldsymbol{\eta},\lambda)$ is a continuous function of $\boldsymbol{\eta}$ and λ for all $\lambda\neq 0$ by Assumption A.3 the Stone-Weierstrass Theorem implies that $f_1(\boldsymbol{\eta},\lambda)$ can be so approximated uniformly. It follows that

$$\lim_{n\to\to\infty}\left|Q_n^{(3)}(\sigma^2,\boldsymbol{\eta})-\log\sigma^2-\frac{2Q_n^{(1)}(\boldsymbol{\eta})}{\sigma^2}\right|=0$$

almost surely, and the convergence is uniform in $\boldsymbol{\eta}$ on $\mathbb{E}_\delta^0$.

The almost sure limit of the criterion function $Q_n^{(3)}(\sigma^2, \boldsymbol{\eta})$ is therefore

$$\mathcal{Q}^{(3)}(\sigma^2, Q(\boldsymbol{\eta})) = \log \sigma^2 + \frac{2Q(\boldsymbol{\eta})}{\sigma^2},$$

uniformly in $\boldsymbol{\eta}$ on $\mathbb{E}_\delta^0$ by Lemma 3.3, whereas $Q_n^{(3)}(\sigma^2, \boldsymbol{\eta})$ is either arbitrarily large for δ sufficiently small or divergent on $\overline{\mathbb{E}}_{\delta 1}^0 \cup \overline{\mathbb{E}}_{\delta 2}^0$ by Lemma 3.4. Concentrating $\mathcal{Q}^{(3)}(\sigma^2, Q(\boldsymbol{\eta}))$ with respect to σ^2 we find that the minimum of the asymptotic criterion function is given by $\log(2Q(\boldsymbol{\eta}_1)) + 1$. Once again $\boldsymbol{\eta}_1 = \arg\min_\eta Q(\boldsymbol{\eta})$ is the pseudo-true parameter for the estimator under mis-specification and we can conclude that $\lim_{n\to\infty} \widehat{\boldsymbol{\eta}}_1^{(3)} = \boldsymbol{\eta}_1$ and $\lim_{n\to\infty} \|\widehat{\boldsymbol{\eta}}_1^{(3)} - \widehat{\boldsymbol{\eta}}_1^{(1)}\| = 0$.

The CSS estimator:

Let $\mathbf{T}_\eta$ and $\mathbf{H}_\eta$ denote the $n \times n$ upper triangular Toeplitz matrix with non-zero elements $\tau_{|i-j|}(\boldsymbol{\eta})$, $i, j = 1, \ldots, n$, and the $n \times \infty$ reverse Hankel matrix with typical element $\tau_{n-i+j}(\boldsymbol{\eta})$, $i = 1, \ldots, n$, $j = 1, \ldots, \infty$, respectively. Let $\mathbf{A}_\eta = [a_{s-r}(\eta)]$ where

$$a_{s-r}(\eta) = \int_{-\pi}^{\pi} \frac{1}{f_1(\boldsymbol{\eta},\lambda)} \exp(i(s-l)\lambda) d\lambda\,, \quad r, s = 1, \ldots, n\,. \tag{3.52}$$

Then from (3.52) we can deduce that $\mathbf{A}_\eta = \mathbf{T}_\eta \mathbf{T}_\eta^\top + \mathbf{H}_\eta \mathbf{H}_\eta^\top$ and from (3.13) and (3.14) it follows that $Q_n^{(4)}(\boldsymbol{\eta}) = \frac{1}{n} \mathbf{Y}^\top \mathbf{T}_\eta \mathbf{T}_\eta^\top \mathbf{Y}$. Replacing $\boldsymbol{\Sigma}_\eta^{-1}$ by $\mathbf{A}_\eta$ in (3.50) and adapting the argument used to establish (3.50) accordingly, in a manner similar to the proof of Lemma 3.1, shows that

$$\lim_{n\to\infty} \left| \frac{1}{n} \mathbf{Y}^\top \mathbf{A}_\eta \mathbf{Y} - \frac{4\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)} \right| = 0, \quad a.s.\,, \tag{3.53}$$

and that the convergence is uniform in $\boldsymbol{\eta}$ on $\mathbb{E}_\delta^0$. It is also shown below that $\frac{1}{n} \mathbf{Y}^\top \mathbf{H}_\eta \mathbf{H}_\eta^\top \mathbf{Y} = o(1)$ when $|d| < 0.5$, $|d_0| < 0.5$ and $d_0 - d < 0.5$.

We can therefore conclude that $\left| Q_n^{(4)}(\boldsymbol{\eta}) - 2Q_n^{(1)}(\boldsymbol{\eta}) \right|$ converges to zero almost surely when $\boldsymbol{\eta} \in \mathbb{E}_\delta^0$, and hence that the limiting value of the criterion function $Q_n^{(4)}(\boldsymbol{\eta})$ is $2Q(\boldsymbol{\eta})$ by Lemma

3.3. When $\boldsymbol{\eta} \in \overline{\mathbb{E}}^0_{\delta 1} \cup \overline{\mathbb{E}}^0_{\delta 2}$, expression (3.53) and Lemma 3.4, together with the equality $Q_n^{(4)}(\boldsymbol{\eta}) = \frac{1}{n}\mathbf{Y}^\top \mathbf{A}_\eta \mathbf{Y} - \frac{1}{n}\mathbf{Y}^\top \mathbf{H}_\eta \mathbf{H}_\eta^\top \mathbf{Y}$, imply that $\liminf_{n\to\infty} Q_n^{(4)}(\boldsymbol{\eta}) \geq \lim_{n\to\infty} \frac{1}{n}\mathbf{Y}^\top \mathbf{A}_\eta \mathbf{Y}$ and the CSS criterion function is either arbitrarily large for δ sufficiently small or divergent. That the pseudo-true parameter for the CSS estimator under mis-specification is $\boldsymbol{\eta}_1 = \arg\min_{\boldsymbol{\eta}} Q(\boldsymbol{\eta})$ and $\lim \widehat{\boldsymbol{\eta}}_1^{(4)} = \boldsymbol{\eta}_1$ and $\lim_{n\to\infty} \|\widehat{\boldsymbol{\eta}}_1^{(4)} - \widehat{\boldsymbol{\eta}}_1^{(1)}\| = 0$ follows directly.

It remains for us to establish that $\frac{1}{n}\mathbf{Y}^\top \mathbf{H}_\eta \mathbf{H}_\eta^\top \mathbf{Y} = o(1)$ in regions of the parameter space where $d_0 - d < 0.5$. Suppressing the dependence on the parameter $\boldsymbol{\eta}$ for notational simplicity, set $\mathbf{M} = \mathbf{H}\mathbf{H}^\top$. Then $\mathbf{M} = [m_{ij}]_{i,j=1,\ldots,n}$ where $m_{ij} = \sum_{u=0}^{\infty} \tau_{u+n-i}\tau_{u+n-j}$, and

$$E_0[\mathbf{Y}^\top \mathbf{M}\mathbf{Y}] = \mathrm{tr}\,(\mathbf{M}\boldsymbol{\Sigma}_0) = \sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}\gamma_0(j-i),$$

where $\gamma_0(\tau)$, $\tau = 0, \pm 1, \pm 2, \ldots$, denotes the autocovariance function of the TDGP. Since $|\tau_k| \sim k^{-(1+d)}\mathcal{C}_\tau$, $\mathcal{C}_\tau < \infty$, the series $\sum_{k=0}^{\infty} |\tau_k|^2 \sim \mathcal{C}_\tau^2 \zeta(2(d+1))$ for all $d > -0.5$, where $\zeta(\cdot)$ denotes the Riemann zeta function, from which we can deduce that $|m_{ij}| \sim \{(n-i+1)(n-j+1)\}^{-(1+d)}\mathcal{C}'_m$ for some $\mathcal{C}'_m < \infty$. Hence on setting $r = n-i+1$ and $s = n-j+1$ we have that

$$0 \leq \sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}\gamma_0(j-i) \sim \mathcal{C}_m n^{-2(d+1)} \sum_{r=1}^{n}\sum_{s=1}^{n} |\gamma_0(r-s)|\,, \tag{3.54}$$

where $\mathcal{C}_m < \infty$. But $|\gamma_0(\tau)| \leq \mathcal{C}_\varrho \gamma_0(0)|\tau|^{2d_0-1}$, $\mathcal{C}_\varrho < \infty$, for all $\tau \neq 0$, and

$$\begin{aligned}
n^{-2(d+1)} \sum_{r=1}^{n}\sum_{s=1}^{n} |\gamma_0(r-s)| &\leq n^{-2(d+1)}\gamma_0(0)(n + 2\mathcal{C}_\varrho \sum_{k=1}^{n-1}(n-k)k^{2d_0-1}) \\
&\leq n^{-(2d+1)}\gamma_0(0)(1 + 2\mathcal{C}_\varrho \sum_{k=1}^{n-1} k^{2d_0-1}) \\
&\sim \frac{\gamma_0(0)}{n^{(2d+1)}} \times \begin{cases} 1 + 2\mathcal{C}_\varrho\zeta(1-2d_0)\,, & d_0 < 0; \\ 1 + 2\mathcal{C}_\varrho \log n\,, & d_0 = 0; \\ 1 + 2\mathcal{C}_\varrho n^{2d_0}/2d_0\,, & d_0 > 0. \end{cases}
\end{aligned}$$

It follows that for all d where $|d| < 0.5$

$$E_0[\mathbf{Y}^\top \mathbf{M}\mathbf{Y}] \leq \frac{\mathcal{C}_m\gamma_0(0)}{n^{1-2(d_0-d)}} \times \begin{cases} 1 + 2\mathcal{C}_\varrho\zeta(1-2d_0)/n^{2d_0}\,, & d_0 < 0; \\ 1 + 2\mathcal{C}_\varrho \log n\,, & d_0 = 0; \\ 1 + \mathcal{C}_\varrho/d_0\,, & d_0 > 0; \end{cases}$$

We can therefore conclude that

$$Pr\left(n^{-1}\mathbf{Y}^{\top}\mathbf{M}\mathbf{Y} > \epsilon\right) = \begin{cases} O(n^{-2(d+1)}), & 0.5 < d_0 < 0; \\ O(\log n / n^{2(d+1)}), & d_0 = 0; \\ O(n^{2(d_0-d)-2}), & 0 < d_0 < 0.5; \end{cases} \tag{3.55}$$

for all $\epsilon > 0$ by Markov's inequality. Since ϵ is arbitrary it follows that when $|d| < 0.5$ and $|d_0| < 0.5$ the almost sure limit of $n^{-1}\mathbf{Y}^{\top}\mathbf{M}\mathbf{Y}$ is zero whenever $d_0 - d < 0.5$, by the Borell-Cantelli lemma, giving the desired result. ■

Proof of Theorem 3.2.

First note that

$$Q_N(\boldsymbol{\eta}) = \left\{\frac{\sigma_0^2\Gamma(1-2(d_0-d))}{2\Gamma^2(1-(d_0-d))}\right\} K_N(\boldsymbol{\eta}), \tag{3.56}$$

by the same argument that gives (3.19). Now let $\Delta C_N(z) = \sum_{j=N+1}^{\infty} c_j z^j = C(z) - C_N(z)$. Then

$$\begin{aligned} |C(e^{i\lambda})|^2 &= |C_N(e^{i\lambda})|^2 + C_N(e^{i\lambda})\Delta C_N(e^{-i\lambda}) \\ &\quad + \Delta C_N(e^{i\lambda})C_N(e^{-i\lambda}) + |\Delta C_N(e^{i\lambda})|^2, \end{aligned}$$

and the remainder term can be decomposed as $R_N = R_{1N} + R_{2N}$ where

$$R_{1N} = \left(\frac{\sigma_0^2}{2\pi}\right)\int_0^{\pi} |\Delta C_N(e^{i\lambda})|^2 |2\sin(\lambda/2)|^{-2(d_0-d)} d\lambda, \tag{3.57}$$

and

$$R_{2N} = \left(\frac{\sigma_0^2}{2\pi}\right)\int_{-\pi}^{\pi} \Delta C_N(e^{i\lambda})C_N(e^{-i\lambda}) |2\sin(\lambda/2)|^{-2(d_0-d)} d\lambda\,. \tag{3.58}$$

The first integral in (3.57) equals

$$\left\{\frac{\sigma_0^2\Gamma(1-2(d_0-d))}{2\Gamma^2(1-(d_0-d))}\right\}\left(\sum_{j=N+1}^{\infty} c_j^2 + 2\sum_{k=N+1}^{\infty}\sum_{j=k+1}^{\infty} c_j c_k \rho(j-k)\right).$$

Because $B(z) \neq 0$, $|z| \leq 1$, it follows that $|c_j| < \mathcal{C}\zeta^j$, $j = 1, 2, \ldots$, for some $\mathcal{C} < \infty$ and $\zeta \in (0,1)$, and hence that

$$\sum_{j=N+1}^{\infty} c_j^2 < \zeta^{2(N+1)} \frac{\mathcal{C}^2}{(1-\zeta^2)}\,.$$

Furthermore, since $|d_0 - d| < 0.5$ Sterling's approximation can be used to show that $|\rho(h)| < \mathcal{C}'^{2(d_0-d)-1}$, $h = 1, 2, \ldots$, for some $\mathcal{C}' < \infty$. This implies that

$$\begin{aligned}\left|\sum_{k=N+1}^{\infty}\sum_{j=k+1}^{\infty} c_j c_k \rho(j-k)\right| &< \sum_{r=0}^{\infty}\sum_{s=r+1}^{\infty} \mathcal{C}^2\mathcal{C}'^{2(N+1)}\zeta^r\zeta^s(s-r)^{2(d_0-d)-1} \\ &< \zeta^{2(N+1)}\frac{\mathcal{C}^2\mathcal{C}'}{(1-\zeta)^2}\,.\end{aligned}$$

Thus we can conclude that $R_{1N} < const.\,\zeta^{2(N+1)}$ where $0 < \zeta < 1$. Applying the Cauchy-Schwarz inequality to the second integral in (3.58) enables us to bound $|R_{2N}|$ by $2(\sigma_0/\sigma)\sqrt{I_N \cdot R_{1N}}$. It therefore follows from the preceding analysis that $|R_{2N}| < const.\,\zeta^{(N+1)}$. Since $|R_N| \leq R_{1N} + |R_{2N}|$ and $(N+1)/\exp(-(N+1)\log\zeta) \to 0$ as $N \to \infty$ it follows that $R_N = o(N^{-1})$, as stated. The gradient vector of $Q(\boldsymbol{\eta})$ with respect to $\boldsymbol{\eta}$ is

$$\frac{\partial Q(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \left(\frac{\sigma_0^2}{2\pi}\right)\int_{-\pi}^{\pi}\frac{C(e^{i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}}\frac{\partial}{\partial \boldsymbol{\eta}}\{\frac{C(e^{-i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}}\}d\lambda,$$

and substituting $C(z) = C_N(z) + \Delta C_N(z)$ gives $\partial Q(\boldsymbol{\psi})/\partial\eta_j = \partial Q_N(\boldsymbol{\eta})/\partial\eta_j + R_{3N} + R_{4N}$ for the typical j'th element where

$$R_{3N} = \left(\frac{\sigma_0^2}{2\pi}\right)\int_{-\pi}^{\pi}\frac{C_N(e^{i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}}\frac{\partial}{\partial \eta_j}\{\frac{\Delta C_N(e^{-i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}}\}d\lambda,$$

and

$$R_{4N} = \left(\frac{\sigma_0^2}{2\pi}\right)\int_{-\pi}^{\pi}\frac{\Delta C_N(e^{i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}}\frac{\partial}{\partial \eta_j}\{\frac{C(e^{-i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}}\}d\lambda\,.$$

The Cauchy-Schwarz inequality now yields the inequalities

$$|R_{3N}|^2 \leq R_{1N}\left(\frac{\sigma_0^2}{2\pi}\right)\int_{-\pi}^{\pi}\frac{|C_N(e^{i\lambda})|^2}{|2\sin(\lambda/2)|^{2(d_0-d)}}\left|\frac{\partial}{\partial \eta_j}\{\log\frac{\Delta C_N(e^{-i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}}\}\right|^2 d\lambda,$$

and

$$|R_{4N}|^2 \leq R_{1N}\left(\frac{\sigma_0^2}{2\pi}\right)\int_{-\pi}^{\pi}\left|\frac{\partial}{\partial \eta_j}\{\frac{C(e^{-i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}}\}\right|^2 d\lambda\,,$$

from which we can infer that $|R_{3N} + R_{4N}| \leq const.\, \zeta^{(N+1)} = o(N^{-1})$, thus completing the proof. ■

Proof of Theorem 3.3.

The distributions exhibited in the three cases presented in Theorem 3.3 correspond to those given in Theorems 1, 3 and 2 of Chen and Deo (2006), and in the following lemmas we state the properties necessary to generalize the applicability of these distributions and establish their validity under the current scenario and assumptions. Although the distributions are non-standard, the proof proceeds standardly via the use of the mean value theorem and convergence in probability of a Hessian in a neighbourhood of $\boldsymbol{\eta}_1$, plus the application to the criterion differential function of an appropriate central limit theorem. ■

Lemma 3.5 *Let*

$$\frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} y_t \exp(-i\lambda t) = \xi(\lambda) = \xi_c(\lambda) - \imath\xi_s(\lambda),$$

and set $\mathbf{X}^\top = (\xi_c(\lambda_1), \xi_s(\lambda_1), \ldots, \xi_c(\lambda_{\lfloor n/2 \rfloor}), \xi_s(\lambda_{\lfloor n/2 \rfloor})\mathbf{F}_0^{-1/2}$ *where*

$$\mathbf{F}_0 = diag(f_0(\lambda_1), f_0(\lambda_1), \ldots, f_0(\lambda_{\lfloor n/2 \rfloor}), f_0(\lambda_{\lfloor n/2 \rfloor}))\,.$$

Assume that Conditions A hold. Then under Assumption $A.1'$ the vector $\mathbf{X}^\top$ converges in distribution to a Gaussian random variable with zero mean and variance-covariance matrix $\mathbf{\Omega} = \frac{1}{2}(\mathbf{I} + \mathbf{\Delta})$, $\mathbf{X}^\top \xrightarrow{D} \boldsymbol{\xi} \sim N(\mathbf{0}, \mathbf{\Omega})$, where $\mathbf{\Delta} = [\triangle_{rc}]$, $\triangle_{rc} = O(j^{-d_0}k^{d_0-1}\log k)$ for $r = 2j-1$ or $r = 2j$, and $c = 2k-1$ or $2k$, $1 \leq j \leq k \leq \lfloor n/2 \rfloor$.

Proof of Lemma 3.5.

Assumption $(A.1')$ implies that Assumption $(A.1)$ of Lahiri (2003) holds. Since *Conditions A* imply that Assumption $(A.3)$ of Lahiri (2003) also holds, the asymptotic normality of $\mathbf{X}^T$ follows from Theorem 2.1 of Lahiri (2003). The stated covariance structure follows from Lemmas

1 and 4 of Moulines and Soulier (1999) in which the moment properties of $\xi_c(\lambda_j)$ and $\xi_s(\lambda_j)$ are derived supposing that *exact* Gaussianity holds for the sine and cosine transforms for all n, with bounds that are uniform with respect to n for each $j = 1, \ldots, \lfloor n/2 \rfloor$. See also Corollary 5.2 of Lahiri (2003) and the discussion in Lahiri (2003, page 624).

Since the limiting joint distribution of the sine and cosine transforms is Gaussian, and the sine and cosine transforms are uniformly integrable, the form of the asymptotic distribution and covariance properties of the corresponding periodogram ordinates are determined by the limit law of $\xi_c(\lambda_j)$ and $\xi_s(\lambda_j)$, $j = 1, \ldots, \lfloor n/2 \rfloor$. ∎

Corollary 3.1 *Assume that the conditions of Lemma 3.5 hold, and for each $j = 1, \ldots \lfloor n/2 \rfloor$ set $Z_j = I(\lambda_j)/f_0(\lambda_j) = |\xi(\lambda_j)|^2/f_0(\lambda_j)$ and let $\rho_j = Cov_0[\xi_c(\lambda_j)\xi_s(\lambda_j)]/f_0(\lambda_j)$. Then $Z_j - \rho_j \xi_c(\lambda_j)\xi_s(\lambda_j)/f_0(\lambda_j)$ converges in distribution to $\frac{1}{2}\chi^2(2)(1 + \triangle_{2j2j})(1 - \rho_j^2)$ where $\chi^2(2)$ denotes a Chi-squared random variable with two degrees of freedom. Furthermore, $E_0[Z_j] = 1 + O(\log j/j)$, $Var_0[Z_j] = 1 + O(\log j/j)$ and $Cov_0[Z_j Z_k] = O(j^{-2|d_0|}k^{2|d_0|-2}log^2 k)$ for $1 \leq j < k \leq \lfloor n/2 \rfloor$.*

Proof of Corollary 3.1.

For $j = 1, \ldots \lfloor n/2 \rfloor$ set

$$U_j = \frac{\xi_c(\lambda_j) - \xi_s(\lambda_j)]}{\sqrt{f_0(\lambda_j)(1 + \triangle_{2j2j})(1 - \rho_j)}} \quad \text{and} \quad V_j = \frac{\xi_c(\lambda_j) + \xi_s(\lambda_j)]}{\sqrt{f_0(\lambda_j)(1 + \triangle_{2j2j})(1 + \rho_j)}}.$$

Then the Continuous Mapping Theorem implies that

$$\frac{Z_j - \rho_j \xi_c(\lambda_j)\xi_s(\lambda_j)/f_0(\lambda_j)}{(1 + \triangle_{2j2j})(1 - \rho_j^2)} = U_j^2 + V_j^2 \xrightarrow{D} \tfrac{1}{2}\chi^2(2),$$

since by Lemma 3.5 $\mathbf{X}^T \xrightarrow{D} \boldsymbol{\xi} \sim N(\mathbf{0}, \boldsymbol{\Omega})$. Let $\mathbf{A}$ and $\mathbf{B}$ be any $\lfloor n/2 \rfloor \times \lfloor n/2 \rfloor$ symmetric selection matrices. Then $E[\boldsymbol{\xi}^\top \mathbf{A} \boldsymbol{\xi}] = \mathrm{tr}\boldsymbol{\Omega}\mathbf{A}$ and $E[(\boldsymbol{\xi}^\top \mathbf{A}\boldsymbol{\xi})(\boldsymbol{\xi}^\top \mathbf{B}\boldsymbol{\xi})] = \mathrm{tr}\boldsymbol{\Omega}\mathbf{A}\mathrm{tr}\boldsymbol{\Omega}\mathbf{B} + \mathrm{tr}\boldsymbol{\Omega}\mathbf{A}\boldsymbol{\Omega}\mathbf{B}$, from which the stated moments can be derived via appropriate choice of $\mathbf{A}$ and $\mathbf{B}$. Note, in particular, that $\rho_j = Cov_0[\xi_c(\lambda_j)\xi_s(\lambda_j)]/f_0(\lambda_j) = \frac{1}{2}\triangle_{(2j-1)2j} = O(\log j/j)$ and $Cov[\xi_j^2\xi_k^2] = (E[\xi_j\xi_k])^2 = \frac{1}{4}\triangle_{2j2k}^2 = O(j^{-2|d_0|}k^{2|d_0|-2}log^2 k)$ for $1 \leq j < k \leq \lfloor n/2 \rfloor$. ■

The remaining steps in the proof of Theorem 3.3 are based on Taylor expansions of the gradient vector (or score function) of the criterion functions. For the FML estimator we have

$$\mathbf{0} = \frac{\partial Q_n^{(1)}(\boldsymbol{\eta}_1)}{\partial \boldsymbol{\eta}} + \frac{\partial^2 Q_n^{(1)}(\overline{\boldsymbol{\eta}}_1)}{\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}'}(\widehat{\boldsymbol{\eta}}_1 - \boldsymbol{\eta}_1)$$

where

$$\begin{aligned}\frac{\partial Q_n^{(1)}(\boldsymbol{\eta}_1)}{\partial \boldsymbol{\eta}} &= -\frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2 \rfloor}\frac{I(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)^2}\frac{\partial f_1(\boldsymbol{\eta},\lambda_j)}{\partial \boldsymbol{\eta}} = -\frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2 \rfloor}\frac{I(\lambda_j)}{f_0(\lambda_j)}\mathbf{w}(\boldsymbol{\eta},\lambda_j) \\ \mathbf{w}(\boldsymbol{\eta},\lambda_j) &= \frac{f_0(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)}\frac{\partial \log f_1(\boldsymbol{\eta},\lambda_j)}{\partial \boldsymbol{\eta}},\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2 Q_n^{(1)}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}'} &= \frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2 \rfloor}\frac{I(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)}\mathbf{H}(\boldsymbol{\eta},\lambda_j), \\ \mathbf{H}(\boldsymbol{\eta},\lambda_j) &= 2\frac{\partial \log(f_1(\boldsymbol{\eta},\lambda_j))}{\partial \boldsymbol{\eta}}\frac{\partial \log(f_1(\boldsymbol{\eta},\lambda_j))}{\partial \boldsymbol{\eta}'} - \frac{1}{f_1(\boldsymbol{\eta},\lambda_j)}\frac{\partial^2 f_1(\boldsymbol{\eta},\lambda_j)}{\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}'},\end{aligned}$$

and the components of $\overline{\boldsymbol{\eta}}_1$ lie on the line segment between $\widehat{\boldsymbol{\eta}}_1$ and $\boldsymbol{\eta}_1$. Existence of the Taylor expansion is justified by convexity and Assumptions $(A.3)$ and $(A.5)$.

Lemma 3.6 *Let $dQ_n^{(1)}(\boldsymbol{\eta};\mathbf{t})$, where $\mathbf{t} = (t_1,\ldots,t_{l+1})^\top$, denote the differential of $Q_n^{(1)}(\boldsymbol{\eta})$. Then under the assumptions of Theorem 3.3*

$$\frac{n}{2\pi}\left(\sum_{j=1}^{\lfloor n/2 \rfloor}\left(\mathbf{t}^\top\mathbf{w}(\boldsymbol{\eta}_1,\lambda_j)\right)^2\right)^{-\frac{1}{2}}\left(dQ_n^{(1)}(\boldsymbol{\eta}_1;\mathbf{t}) - E[dQ_n^{(1)}(\boldsymbol{\eta}_1;\mathbf{t})]\right) \xrightarrow{D} Z \sim N(0,1),$$

for all $\mathbf{t} \in \mathbb{R}^{l+1}$, $0 < \|\mathbf{t}\| < \infty$.

Proof. By Assumption A.3 the differential of $Q_n^{(1)}(\boldsymbol{\eta})$ exists and is given by $\partial Q_n^{(1)}(\boldsymbol{\eta}_1)/\partial \boldsymbol{\eta}^\top\mathbf{t}$, from which it follows that

$$dQ_n^{(1)}(\boldsymbol{\eta}_1;\mathbf{t}) - E[dQ_n^{(1)}(\boldsymbol{\eta}_1;\mathbf{t})] = -\frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2 \rfloor}(Z_j - E[Z_j])\mathbf{w}(\boldsymbol{\eta}_1,\lambda_j)^\top\mathbf{t}.$$

Theorem 2 of Moulines and Soulier (1999) provides a generalization of central limit theorems for triangular arrays of martingale differences and weakly dependent sequences to similarly weighted sums of correlated variables. Replacing Moulines and Soulier's η_{nj} by $Z_j - E[Z_j]$ and their $b_{n,j}$ by $\mathbf{w}(\boldsymbol{\eta}, \lambda_j)^\top \mathbf{t}$, recognizing from Corollary 3.1 that $Z_j - E[Z_j]$, $j = 1, \ldots, \lfloor n/2 \rfloor$, share the same moment structure and order of correlation as Moulines and Soulier's η_{nj}, the proof of the lemma follows Moulines and Soulier's proof of their Theorem 2 presented in Moulines and Soulier (1999, Appendix B). Conditions (i) and (ii) of Theorem 2 of Moulines and Soulier (1999) are satisfied because $C_1 \lambda_j^{-2d^*} \log \lambda_j \leq \|\mathbf{w}(\boldsymbol{\eta}, \lambda_j)\| \leq C_2 \lambda_j^{-2d^*} \log \lambda_j$ for some constants C_1 and C_2 (see Chen and Deo, 2006, expression (21) page. 276) and

$$\lim_{n\to\infty} \sup_{j=1,\ldots,\lfloor n/2 \rfloor} \left(\sum_{j=1}^{\lfloor n/2 \rfloor} \left(\mathbf{t}^\top \mathbf{w}(\boldsymbol{\eta}_1, \lambda_j) \right)^2 \right)^{-1} (\mathbf{w}(\boldsymbol{\eta}_1, \lambda_j)^\top \mathbf{t})^2 = 0 \, . \quad \blacksquare$$

The following lemma parallels Lemma 3 of Chen and Deo (2006) and is derived in a similar fashion. The lemma and its proof are presented here for completeness. ■

Lemma 3.7 *Let $\mathbb{E}_c$ denote a compact convex subset of $\mathbb{E}_\delta^0$ and denote the second order differential of the FML criterion function by $d^2 Q_n^{(1)}(\boldsymbol{\eta}; \mathbf{t}) = \mathbf{t}^\top \left(\partial^2 Q_n^{(1)}(\boldsymbol{\eta}) / \partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top \right) \mathbf{t}$. Then for all $\mathbf{t}$, $\|\mathbf{t}\| < \infty$,*

$$plim_{n\to\infty} \sup_{\boldsymbol{\eta}\in\mathbb{E}_c} \left| d^2 Q_n^{(1)}(\boldsymbol{\eta}; \mathbf{t}) - d^2 Q(\boldsymbol{\eta}; \mathbf{t}) \right| = 0 \, .$$

under Assumptions $(A.1')$ and $(A.2) - (A.5)$.

Proof of Lemma 3.7. By definition of the second order differential we have

$$\begin{aligned} E_0 \left[d^2 Q_n^{(1)}(\boldsymbol{\eta}; \mathbf{t}) \right] &= E_0 \left[\frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{f_1(\boldsymbol{\eta}, \lambda_j)} \mathbf{t}^\top \mathbf{H}(\boldsymbol{\eta}, \lambda_j) \mathbf{t} \right] \\ &= \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{f_0(\lambda_j)}{f_1(\boldsymbol{\eta}, \lambda_j)} \mathbf{t}^\top \mathbf{H}(\boldsymbol{\eta}, \lambda_j) \mathbf{t} \\ &\quad + \left(\frac{E_0[I(\lambda_j)]}{f_0(\lambda_j)} - 1 \right) \frac{f_0(\lambda_j)}{f_1(\boldsymbol{\eta}, \lambda_j)} \mathbf{t}^\top \mathbf{H}(\boldsymbol{\eta}, \lambda_j) \mathbf{t} \, , \end{aligned}$$

where $E_0[I(\lambda_j)]/f_0(\lambda_j) - 1 = O(\log j/j)$, by Corollary 3.1, and $\mathbf{t}^\top \mathbf{H}(\boldsymbol{\eta}, \lambda_j)\mathbf{t} = O(\log^2 \lambda_j)$ since $\sup_{\boldsymbol{\eta}} \partial \log f_1(\boldsymbol{\eta}, \lambda_j)/\partial \boldsymbol{\eta}$ is of order $O(\log \lambda_j)$ by Assumptions $(A.2)$ and $(A.3)$ and $\|\mathbf{t}\| < \infty$. Thus we can conclude that

$$\begin{aligned}
\frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2 \rfloor}\left(\frac{E_0[I(\lambda_j)]}{f_0(\lambda_j)} - 1\right)\frac{f_0(\lambda_j)}{f_1(\boldsymbol{\eta}, \lambda_j)}\mathbf{t}^\top \mathbf{H}(\boldsymbol{\eta}, \lambda_j)\mathbf{t} &= O\left(\frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2 \rfloor}\frac{\log j}{j}\lambda_j^{-2d^*}\log^2 \lambda_j\right) \\
&= O\left(n^{2d^*-1}\sum_{j=1}^{\lfloor n/2 \rfloor}\frac{\log j}{j} j^{-2d^*}\log^2 (j/n)\right) \\
&= \begin{cases} O(n^{2d^*-1}\log^2 n), & 0 < d^* < 0.5\,; \\ O(n^{-1}\log^4 n), & -1.0 < d^* \le 0\,, \end{cases}
\end{aligned}$$

and hence that $E_0\left[d^2 Q_n^{(1)}(\boldsymbol{\eta}; \mathbf{t})\right] \to \mathbf{t}^\top \frac{\partial^2 Q(\boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'}\mathbf{t} = d^2 Q(\boldsymbol{\eta}; \mathbf{t})$. Similarly, setting $h(\boldsymbol{\eta}; \mathbf{t}, \lambda_j, \lambda_k) = \mathbf{t}^\top \mathbf{H}(\boldsymbol{\eta}, \lambda_j)\mathbf{t} \cdot \mathbf{t}^\top \mathbf{H}(\boldsymbol{\eta}, \lambda_k)\mathbf{t}$ and invoking Corollary 3.1 once again we have

$$\begin{aligned}
Var_0\left[d^2 Q_n^{(1)}(\boldsymbol{\eta}; \mathbf{t})\right] &= \left(\frac{2\pi}{n}\right)^2 \sum_{j=1}^{\lfloor n/2 \rfloor}\sum_{k=1}^{\lfloor n/2 \rfloor}\frac{f_0(\lambda_j)}{f_1(\boldsymbol{\eta}, \lambda_j)}\frac{f_0(\lambda_k)}{f_1(\boldsymbol{\eta}, \lambda_k)} h(\boldsymbol{\eta}; \mathbf{t}, \lambda_j, \lambda_k) Cov_0\left[\frac{I(\lambda_j)}{f_0(\lambda_j)}\frac{I(\lambda_k)}{f_0(\lambda_k)}\right] \\
&= O\left(\frac{1}{n^2}\sum_{j=1}^{\lfloor n/2 \rfloor}\sum_{k \ge j}^{\lfloor n/2 \rfloor}\lambda_j^{-2d^*}\lambda_k^{-2d^*}\log^2 \lambda_j \log^2 \lambda_k j^{-2|d_0|}k^{2|d_0|-2}\log^2 k\right) \\
&= O\left(\frac{1}{n^2}\sum_{j=1}^{\lfloor n/2 \rfloor}\sum_{k \ge j}^{\lfloor n/2 \rfloor}\lambda_j^{-2d^*}\lambda_k^{-2d^*}\log^2 \lambda_j \log^2 \lambda_k j^{-2|d_0|}k^{2|d_0|-2}\log^2 k\right) \\
&= \begin{cases} O(n^{4d^*-2}\log^4 n), & d^* + |d_0| > 0.5 \quad 0 < d^* < 0.5\,; \\ O(n^{-(1+2(|d_0|-d^*))}\log^5 n), & d^* + |d_0| \le 0.5 \quad 0 < d^* < 0.5\,; \\ O(n^{-(1+2|d_0|)}\log^5 n), & d^* + |d_0| \le 0.5 \quad -1 < d^* \le 0\,. \end{cases}
\end{aligned}$$

It therefore follows from Markov's inequality that $d^2 Q_n^{(1)}(\boldsymbol{\eta}; \mathbf{t})$ converges in probability to $\mathrm{d}^2 Q(\boldsymbol{\eta}; \mathbf{t})$.

Now, by the Mean Value Theorem, for any $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ in $\mathbb{E}_c$

$$\left|d^2 Q_n^{(1)}(\boldsymbol{\eta}_1; \mathbf{t}) - d^2 Q_n^{(1)}(\boldsymbol{\eta}_2; \mathbf{t})\right| \le \left\|\frac{\partial\left\{d^2 Q_n^{(1)}(\overline{\boldsymbol{\eta}}; \mathbf{t})\right\}}{\partial \boldsymbol{\eta}}\right\| \cdot \|\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2\|,$$

for some $\overline{\boldsymbol{\eta}}$ between $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$. Moreover,

$$E_0\left[\frac{\partial\left\{d^2 Q_n^{(1)}(\boldsymbol{\eta}; \mathbf{t})\right\}}{\partial \boldsymbol{\eta}}\right] = \frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2 \rfloor}\left\{\frac{f_0(\lambda_j)}{f_1(\boldsymbol{\eta}, \lambda_j)} + \left(\frac{E_0[I(\lambda_j)]}{f_0(\lambda_j)} - 1\right)\frac{f_0(\lambda_j)}{f_1(\boldsymbol{\eta}, \lambda_j)}\right\} \cdot \mathbf{k}(\boldsymbol{\eta}; \mathbf{t}, \lambda_j)$$

$$= \frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{f_0(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)}\mathbf{k}(\boldsymbol{\eta};\mathbf{t},\lambda_j)+\mathbf{r}_n, \tag{3.59}$$

where

$$\mathbf{k}(\boldsymbol{\eta};\mathbf{t},\lambda_j)=\frac{\partial\mathbf{t}^\top\mathbf{H}(\boldsymbol{\eta},\lambda_j)\mathbf{t}}{\partial\boldsymbol{\eta}}-\mathbf{t}^\top\mathbf{H}(\boldsymbol{\eta},\lambda_j)\mathbf{t}\,\frac{\partial\log f_1(\boldsymbol{\eta},\lambda_j)}{\partial\boldsymbol{\eta}}=O(\log^3\lambda_j),$$

and the remainder

$$\begin{aligned}\mathbf{r}_n &= \frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\left(\frac{E_0[I(\lambda_j)]}{f_0(\lambda_j)}-1\right)\frac{f_0(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)}\mathbf{k}(\boldsymbol{\eta};\mathbf{t},\lambda_j)\\ &= O\left(\frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{\log j}{j}\lambda_j^{-2d^*}\log^3\lambda_j\right)\\ &= \begin{cases} O(n^{2d^*-1}\log^3 n), & 0<d^*<0.5\,;\\ O(n^{-1}\log^5 n), & -1<d^*\le 0\,,\end{cases}\end{aligned}$$

From Assumption $(A.3)$ and $(A.5)$ it follows that the components of the first term on the right hand side of (3.59) converge to finite constants, and hence that

$$\left|d^2Q_n^{(1)}(\boldsymbol{\eta}_1;\mathbf{t})-d^2Q_n^{(1)}(\boldsymbol{\eta}_2;\mathbf{t})\right|\le C_n\|\boldsymbol{\eta}_1-\boldsymbol{\eta}_2\|,$$

where

$$C_n=\sup_{\boldsymbol{\eta}\in\mathbb{E}_c}\left\|\frac{\partial\left\{d^2Q_n^{(1)}(\boldsymbol{\eta};\mathbf{t})\right\}}{\partial\boldsymbol{\eta}}\right\|=O_p(1),$$

since $\sup_n E_0\left[\partial\left\{d^2Q_n^{(1)}(\boldsymbol{\eta};\mathbf{t})\right\}/\partial\boldsymbol{\eta}\right]<\infty$ for all $\boldsymbol{\eta}\in\mathbb{E}_c$. We can therefore conclude that $d^2Q_n^{(1)}(\boldsymbol{\eta}_1;\mathbf{t})$ is stochastically equicontinuous, and hence that

$$\operatorname*{plim}_{n\to\infty}\sup_{\boldsymbol{\eta}\in\mathbb{E}_c}\left|d^2Q_n^{(1)}(\boldsymbol{\eta};\mathbf{t})-d^2Q(\boldsymbol{\eta};\mathbf{t})\right|=0\,,$$

for all $\mathbf{t}$, $\|\mathbf{t}\|<\infty$, as required. ■

That the FML estimator possesses the asymptotic distributions as specified in Theorem 3.3 now follows by replacing Lemma 5 of Chen and Deo (2006) by Lemma and Corollary 3.5, Lemmas 8 and 9 by Lemma 3.6, and Lemma 3 of Chen and Deo (2006) by Lemma 3.7.

Having made these replacements we then find that the convergence rates and asymptotic approximations given in Chen and Deo's Lemma 4 and for their Cases 1, 2 and 3 in their lemmas 6, 7, 10, 11 and 12 remain valid, thus establishing Theorem 3.3 for the FML estimator.

For the Whittle estimator we have, via definition of the differential and application of the chain rule, that

$$\begin{aligned}\left|dQ_n^{(2)}(\boldsymbol{\eta};\mathbf{t}) - \frac{dQ_n^{(1)}(\boldsymbol{\eta};\mathbf{t})}{Q_n^{(1)}(\boldsymbol{\eta})}\right| &\leq \left|\nabla Q_n^{(2)}(\boldsymbol{\eta};\mathbf{t}) - dQ_n^{(2)}(\boldsymbol{\eta};\mathbf{t})\right| \\ &\quad + \left|\nabla Q_n^{(2)}(\boldsymbol{\eta};\mathbf{t}) - \nabla \log Q_n^{(1)}(\boldsymbol{\eta};\mathbf{t})\right| \\ &\quad + \left|\nabla \log Q_n^{(1)}(\boldsymbol{\eta};\mathbf{t}) - \frac{dQ_n^{(1)}(\boldsymbol{\eta};\mathbf{t})}{Q_n^{(1)}(\boldsymbol{\eta})}\right| \\ &\leq 2\epsilon\|\mathbf{t}\| + |\nabla Q_n^{(2)}(\boldsymbol{\eta};\mathbf{t}) - \nabla \log Q_n^{(1)}(\boldsymbol{\eta};\mathbf{t})|,\end{aligned}$$

where

$$\nabla \log Q_n^{(1)}(\boldsymbol{\eta};\mathbf{t}) = \log Q_n^{(1)}(\boldsymbol{\eta}+\mathbf{t}) - \log Q_n^{(1)}(\boldsymbol{\eta}) \quad \textit{and} \quad \nabla Q_n^{(2)}(\boldsymbol{\eta};\mathbf{t}) = Q_n^{(2)}(\boldsymbol{\eta}+\mathbf{t}) - Q_n^{(2)}(\boldsymbol{\eta})$$

and $\epsilon \to 0$ as $\|\mathbf{t}\| \to 0$. Setting $\|\mathbf{t}\| = O(n^{-1}\log n)$, noting that (3.49) implies that the difference in differences $|\nabla Q_n^{(2)}(\boldsymbol{\eta};\mathbf{t}) - \nabla \log Q_n^{(1)}(\boldsymbol{\eta};\mathbf{t})| = O(n^{-1}\log n)$, we find that

$$\left|dQ_n^{(2)}(\boldsymbol{\eta};\mathbf{t}) - \frac{dQ_n^{(1)}(\boldsymbol{\eta};\mathbf{t})}{Q_n^{(1)}(\boldsymbol{\eta})}\right| \leq O(n^{-1}\log n)\,. \tag{3.60}$$

Equation (3.60) leads, in turn, to the conclusion that

$$\left|d^2Q_n^{(2)}(\boldsymbol{\eta};\mathbf{t}) - \frac{d^2Q_n^{(1)}(\boldsymbol{\eta};\mathbf{t})}{Q_n^{(1)}(\boldsymbol{\eta})}\right| \leq \left\{\frac{dQ_n^{(1)}(\boldsymbol{\eta};\mathbf{t})}{Q_n^{(1)}(\boldsymbol{\eta})}\right\}^2 + O(n^{-1}\log n). \tag{3.61}$$

But by Lemma 4 of Chen and Deo (2006)

$$\begin{aligned} E\left[dQ_n^{(1)}(\boldsymbol{\eta}_1;\mathbf{t})\right] &= -\frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor} E\left[\frac{I(\lambda_j)}{f_0(\lambda_j)}\right]\mathbf{w}(\boldsymbol{\eta}_1,\lambda_j)^\top\mathbf{t} \\ &= \begin{cases} O(n^{2d^*-1}\log n), & 0 < d^* < 0.5\,; \\ O(n^{-1}\log^3 n), & -1.0 < d^* \leq 0\,. \end{cases} \end{aligned} \tag{3.62}$$

In addition,

$$\begin{aligned}
Var_0\left[dQ_n^{(1)}(\boldsymbol{\eta};\mathbf{t})\right] &= O\left(\frac{1}{n^2}\sum_{j=1}^{\lfloor n/2\rfloor}\sum_{k\geq j}^{\lfloor n/2\rfloor}\lambda_j^{-2d^*}\lambda_k^{-2d^*}\log\lambda_j\log\lambda_k j^{-2|d_0|}k^{2|d_0|-2}\log^2 k\right)\\
&= O\left(n^{4d^*-2}\sum_{j=1}^{\lfloor n/2\rfloor} j^{-2(d^*+|d_0|)}\log(j/n)\sum_{k=1}^{\lfloor n/2\rfloor}k^{-2(d^*-|d_0|)-2}\log^2 k\log(k/n)\right)\\
&= \left\{\begin{array}{lll} O(n^{4d^*-2}\log^2 n), & d^*+|d_0|>0.5 & 0<d^*<0.5\,;\\ O(n^{-(1-2(d^*-|d_0|))}\log^3 n), & d^*+|d_0|\leq 0.5 & 0<d^*<0.5\,;\\ O(n^{-(1+2|d_0|)}\log^3 n), & d^*+|d_0|\leq 0.5 & -1.0<d^*\leq 0\,. \end{array}\right.
\end{aligned} \tag{3.63}$$

The asymptotic equivalence of the FML and Whittle estimators now follows since: by Lemma 3.3 $Q_n^{(1)}(\boldsymbol{\eta}_1)$ converges almost surely to $Q(\boldsymbol{\eta}_1)\geq\sigma_0^2>0$; equations (3.61), (3.62) and (3.63) imply that $|d^2Q_n^{(2)}(\boldsymbol{\eta}_1;\mathbf{t})-d^2Q_n^{(1)}(\boldsymbol{\eta}_1;\mathbf{t})/Q_n^{(1)}(\boldsymbol{\eta}_1)|=o_p(1)$; and equation (3.60) implies that

$$\begin{aligned}
\frac{n}{2\pi}\left(\sum_{j=1}^{\lfloor n/2\rfloor}\left(\mathbf{t}^\top\mathbf{w}(\boldsymbol{\eta}_1,\lambda_j)\right)^2\right)^{-\frac{1}{2}}&\left|dQ_n^{(2)}(\boldsymbol{\eta}_1;\mathbf{t})-\frac{dQ_n^{(1)}(\boldsymbol{\eta}_1;\mathbf{t})}{Q_n^{(1)}(\boldsymbol{\eta}_1)}\right|\\
&= \left\{\begin{array}{ll} O(n^{-2(d_0-d_1)}), & 0.25<d_0-d_1<0.5\,;\\ O((n\log n)^{-\frac{1}{2}}), & -1.0<d_0-d_1\leq 0.25\,, \end{array}\right.
\end{aligned}$$

since

$$\begin{aligned}
\sum_{j=1}^{\lfloor n/2\rfloor}\left\{\mathbf{t}^\top\mathbf{w}(\boldsymbol{\eta},\lambda_j)\right\}^2 &= O\left(\sum_{j=1}^{\lfloor n/2\rfloor}\lambda_j^{-4d^*}\log^2\lambda_j\right)\\
&= \left\{\begin{array}{ll} O(n^{4d^*}\log^2 n), & 0.25<d^*<0.5\,;\\ O(n\log^3 n), & -1.0<d^*\leq 0.25\,. \end{array}\right.
\end{aligned}$$

This establishes that Lemma 3.6 also holds with $\mathrm{d}Q_n^{(1)}(\boldsymbol{\eta}_1;\mathbf{t})$ replaced by $Q(\boldsymbol{\eta}_1)\mathrm{d}Q_n^{(2)}(\boldsymbol{\eta}_1;\mathbf{t})$.

For the TML estimator we begin by noting that

$$\left|\frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\log f_1(\boldsymbol{\eta},\lambda_j)-\frac{\pi}{n}\log|\boldsymbol{\Sigma}_\eta|\right|=O(n^{-1}\log n)\,,$$

and concentrating $Q_n^{(2)}(\sigma^2,\boldsymbol{\eta})$ and $Q_n^{(3)}(\sigma^2,\boldsymbol{\eta})$ with respect to σ^2 yields the inequality

$$\left|Q_n^{(2)}(\boldsymbol{\eta})-\pi Q_n^{(3)}(\boldsymbol{\eta})\right|\leq O(n^{-1}\log n)+\left|\log 2Q_n^{(1)}(\boldsymbol{\eta})-\log(2\pi/n)\mathbf{Y}^\top\boldsymbol{\Sigma}_\eta^{-1}\mathbf{Y}\right|\,. \tag{3.64}$$

If we let $\mathbf{U}$ denote the $n \times n$ unitary matrix with entries $n^{-\frac{1}{2}} \exp(\imath 2\pi(r-1)(c-1)/n)$ in row r and column c, $r, c = 1, \ldots, n$, then the off diagonal entries in $\mathbf{U}\boldsymbol{\Sigma}_\eta\mathbf{U}^*$ are of order $O(n^{-1})$, and the diagonal entries are

$$\sum_{s=-(n-1)}^{n-1} \left(1 - \frac{|s|}{n}\right) \frac{\gamma_1(s)}{\sigma^2} \exp(\imath 2\pi(j-1)s/n) \quad j = 1, \ldots, n\,.$$

Since $f_1(\boldsymbol{\eta}, \lambda)$ is absolutely integrable on $[-\pi, \pi]$, and by Assumptions 3 and 5 $f_1(\boldsymbol{\eta}, \lambda)$ is continuously differentiable for all $\lambda \neq 0$, from Fejer's Theorem it follows that $\mathbf{U}\boldsymbol{\Sigma}_\eta\mathbf{U}^* - \mathbf{F}_1 = O(n^{-1})$ where $\mathbf{F}_1$ equals

$$\begin{cases} \mathrm{diag}(Csf_1, f_1(\boldsymbol{\eta}, \lambda_1) \ldots, f_1(\boldsymbol{\eta}, \lambda_{\lfloor n/2 \rfloor}), f_1(\boldsymbol{\eta}, \lambda_{\lfloor n/2 \rfloor}), \ldots, f_1(\boldsymbol{\eta}, \lambda_1)), & \text{for } n \text{ odd;} \\ \mathrm{diag}(Csf_1, f_1(\boldsymbol{\eta}, \lambda_1) \ldots, f_1(\boldsymbol{\eta}, \lambda_{(n-2)/2}), f_1(\boldsymbol{\eta}, \lambda_{\lfloor n/2 \rfloor}), f_1(\boldsymbol{\eta}, \lambda_{(n-2)/2}), \ldots, f_1(\boldsymbol{\eta}, \lambda_1)), & \text{for } n \text{ even,} \end{cases}$$

and the Cèsaro sum

$$Csf_1 = \sum_{s=-(n-1)}^{n-1} \left(1 - \frac{|s|}{n}\right) \frac{\gamma_1(s)}{\sigma^2} = \begin{cases} O(n^{2d} \log n), & 0 < d < 0.5 \\ O(1), & -0.5 < d \leq 0\,. \end{cases}$$

Conditions A and Assumption A.3 imply that $\boldsymbol{\Sigma}_\eta$ and $\mathbf{F}_1$ are positive definite and it therefore follows, upon application of the Rayleigh-Ritz theorem, that

$$\begin{aligned} \frac{2\pi}{n} \left| \mathbf{Y}^\top \boldsymbol{\Sigma}_\eta^{-1} \mathbf{Y} - \mathbf{Y}^\top \mathbf{U} \mathbf{F}_1^{-1} \mathbf{U}^* \mathbf{Y} \right| &= \left| \frac{2\pi}{n} \mathbf{Y}^\top \boldsymbol{\Sigma}_\eta^{-1} \mathbf{Y} - 2Q_n^{(1)}(\boldsymbol{\eta}) \right| \\ &= \frac{1}{n} |\mathbf{Y}^\top \mathbf{R}_\eta \mathbf{Y}| \\ &\leq \frac{1}{n} \max_{i=1,\ldots,n} \{|\mu_i(\mathbf{R}_\eta)|\} \, \|\mathbf{Y}\|^2 , \end{aligned}$$

where $\mu_i(\mathbf{R}_\eta)$, $i = 1, \ldots, n$, are the eigenvalues of the residual $\mathbf{R}_\eta = \boldsymbol{\Sigma}_\eta^{-1} - \mathbf{U}\mathbf{F}_1^{-1}\mathbf{U}^* = O(n^{-1})$. Evaluating the characteristic polynomial of $\mathbf{R}_\eta$ via the leading principle minors, or using the Faddeev-Leverrier method, then indicates that $|\mu_i(\mathbf{R}_\eta)|^n \leq |\mu_i(\mathbf{R}_\eta)|^{n-1} O(n^{-1})$ and the spectral radius of $\mathbf{R}_\eta$ is $O(n^{-1})$.

We can therefore use the method leading to (3.60) and (3.61) to deduce from the inequality in (3.64) that the first and second differentials satisfy

$$|dQ_n^{(2)}(\boldsymbol{\eta}_1;\mathbf{t}) - \pi dQ_n^{(3)}(\boldsymbol{\eta}_1;\mathbf{t})| = O(n^{-1}\log n),$$

and

$$|d^2Q_n^{(2)}(\boldsymbol{\eta}_1;\mathbf{t}) - \pi d^2Q_n^{(3)}(\boldsymbol{\eta}_1;\mathbf{t})| = o_p(1).$$

It therefore follows that the Whittle estimator and the TML estimator converge in distribution as

$$\frac{n}{2\pi}\left(\sum_{j=1}^{\lfloor n/2\rfloor}\left(\mathbf{t}^\top\mathbf{w}(\boldsymbol{\eta}_1,\lambda_j)\right)^2\right)^{-\frac{1}{2}}\left|dQ_n^{(2)}(\boldsymbol{\eta}_1;\mathbf{t}) - \pi dQ_n^{(3)}(\boldsymbol{\eta}_1;\mathbf{t})\right|$$
$$= \begin{cases} O(n^{-2(d_0-d_1)}), & 0.25 < d_0 - d_1 < 0.5\,; \\ O((n\log n)^{-\frac{1}{2}}), & -1.0 < d_0 - d_1 \leq 0.25\,. \end{cases}.$$

For the CSS estimator we have $Q_n^{(4)}(\boldsymbol{\eta}_1) = \left\{\mathbf{Y}^\top\mathbf{A}_\eta\mathbf{Y} - \mathbf{Y}^\top\mathbf{M}_\eta\mathbf{Y}\right\}/n$. Replacing $\boldsymbol{\Sigma}_\eta$ by $\mathbf{A}_\eta$ and adapting the argument used previously shows that $\mathbf{U}\mathbf{A}_\eta\mathbf{U}^* = 2\pi\mathbf{F}_1^{-1} + O(n^{-1})$ and hence, using (3.55), that

$$\left|Q_n^{(4)}(\boldsymbol{\eta}_1) - 2Q_n^{(1)}(\boldsymbol{\eta}_1)\right| \leq O(n^{-1}) + o_p(n^{-\frac{1}{2}})\,.$$

Apart from notational changes, the remaining steps in showing that the CSS and FML estimators converge in distribution are the same as those used in establishing the equivalence of the FML, Whittle and TML estimators, and are therefore omitted.

The preceding derivations imply that Lemma 3.6 also holds with $dQ_n^{(1)}(\boldsymbol{\eta}_1;\mathbf{t})$ replaced by $Q(\boldsymbol{\eta}_1)dQ_n^{(2)}(\boldsymbol{\eta}_1;\mathbf{t})$, $\pi Q(\boldsymbol{\eta}_1)dQ_n^{(3)}(\boldsymbol{\eta}_1;\mathbf{t})$ and $(1/2)dQ_n^{(4)}(\boldsymbol{\eta}_1;\mathbf{t})$. As with the FML estimator, we then find that the convergence rates and asymptotic approximations given in lemmas 4, 6, 7, 10, 11 and 12 of Chen and Deo (2006) remain valid, thus establishing Theorem 3.3 for the Whittle, TML and CSS estimators, and hence confirming that the four estimators $\widehat{\boldsymbol{\eta}}_1^{(1)}$, $\widehat{\boldsymbol{\eta}}_1^{(2)}$, $\widehat{\boldsymbol{\eta}}_1^{(3)}$ and $\widehat{\boldsymbol{\eta}}_1^{(4)}$ are asymptotically equivalent.

3.B Appendix: Evaluation of bias-correction term

For the FML estimator we have

$$\begin{aligned} E_0\left(\frac{\partial Q_n^{(1)}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right) &= \frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor} E_0(I(\lambda_j))\frac{\partial f_1(\boldsymbol{\eta},\lambda_j)^{-1}}{\partial \boldsymbol{\eta}} \\ &= \frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\left(\sum_{|k|<n}\left(1-\frac{|k|}{n}\right)\gamma_0(k)\exp(ik\lambda_j)\right)\frac{\partial f_1(\boldsymbol{\eta},\lambda_j)^{-1}}{\partial \boldsymbol{\eta}}, \end{aligned}$$

where $\gamma_0(k)$ denotes the autocovariance at lag k of the TDGP (see, for example, Brockwell and Davis, 1991, Proposition 10.3.1). Similarly, for the Whittle estimator we have

$$\begin{aligned} E_0\left(\frac{\partial Q_n^{(2)}(\sigma^2,\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right) &= \frac{4}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{\partial \log f_1(\boldsymbol{\eta}_1,\lambda_j)}{\partial \boldsymbol{\eta}} \\ &\quad +\frac{8\pi}{\sigma^2 n}\sum_{j=1}^{\lfloor n/2\rfloor}\left(\sum_{|k|<n}\left(1-\frac{|k|}{n}\right)\gamma_0(k)\exp(ik\lambda_j)\right)\frac{\partial f_1(\boldsymbol{\eta},\lambda_j)^{-1}}{\partial \boldsymbol{\eta}}. \end{aligned}$$

Differentiating the TML criterion function with respect to $\boldsymbol{\eta}$ gives

$$\frac{\partial Q_n^{(3)}(\sigma^2,\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \frac{1}{n}\mathrm{tr}\boldsymbol{\Sigma}_\eta^{-1}\frac{\partial \boldsymbol{\Sigma}_\eta}{\partial \boldsymbol{\eta}} + \frac{1}{n\sigma^2}\mathbf{Y}^T\frac{\partial \boldsymbol{\Sigma}_\eta^{-1}}{\partial \boldsymbol{\eta}}\mathbf{Y},$$

which has expectation

$$E_0\left(\frac{\partial Q_n^{(3)}(\sigma^2,\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right) = \frac{1}{n}\mathrm{tr}\boldsymbol{\Sigma}_\eta^{-1}\frac{\partial \boldsymbol{\Sigma}_\eta}{\partial \boldsymbol{\eta}} - \frac{1}{n\sigma^2}\mathrm{tr}\boldsymbol{\Sigma}_\eta^{-1}\frac{\partial \boldsymbol{\Sigma}_\eta}{\partial \boldsymbol{\eta}}\boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Sigma}_0,$$

where $\boldsymbol{\Sigma}_0 = [\gamma_0(|i-j|)]$ and $\sigma^2\boldsymbol{\Sigma}_\eta = [\gamma_1(|i-j|)]$, $i,j = 1,2,...,n$. The criterion function for the CSS estimator in (3.13) can be re-written as

$$Q_n^{(4)}(\boldsymbol{\eta}) = \frac{1}{n}\sum_{t=1}^{n}\left(\sum_{i=0}^{t-1}\tau_i y_{t-i}\right)^2 = \frac{1}{n}\sum_{t=1}^{n}\sum_{i=0}^{t-1}\sum_{j=0}^{t-1}\tau_i\tau_j y_{t-i}y_{t-j},$$

where τ_i is as defined in (3.15). The gradient of $Q_n^{(4)}(\boldsymbol{\eta})$, recalling that $\tau_i = \tau_i(\boldsymbol{\eta})$, is thus given by

$$\frac{\partial Q_n^{(4)}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \frac{1}{n}\sum_{t=1}^{n}\sum_{i=0}^{t-1}\sum_{j=0}^{t-1}\left(\tau_i\frac{\partial \tau_j}{\partial \boldsymbol{\eta}} + \tau_j\frac{\partial \tau_i}{\partial \boldsymbol{\eta}}\right)y_{t-i}y_{t-j},$$

and the expected value of the gradient is

$$E_0\left(\frac{\partial Q_n^{(4)}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right)=\frac{1}{n}\sum_{t=1}^{n}\sum_{i=0}^{t-1}\sum_{j=0}^{t-1}\left(\tau_i\frac{\partial \tau_j}{\partial \boldsymbol{\eta}}+\tau_j\frac{\partial \tau_i}{\partial \boldsymbol{\eta}}\right)\gamma_0(i-j)\,.$$

3.C Appendix: Additional Graphical results

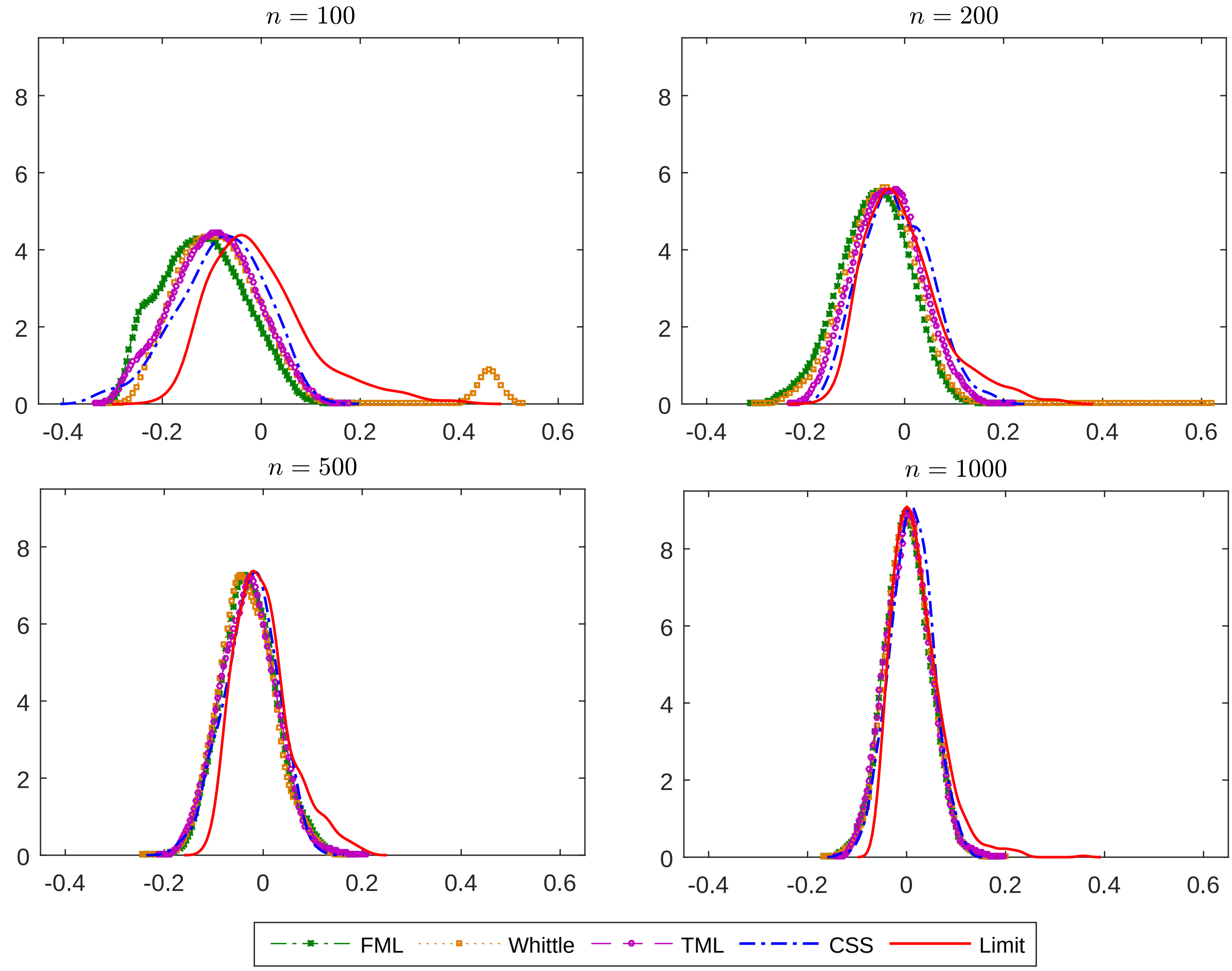

Figure 3.6: Kernel density of $\frac{n^{1-2d^*}}{\log n}\left(\widehat{d}_1-d_1-\mu_n\right)$ for an $ARFIMA(0,d_0,1)$ TDGP with $d_0=-0.2$ and $\theta_0=-0.7$, and an $ARFIMA(0,d,1)$ MisM; $d^*>0.25$.

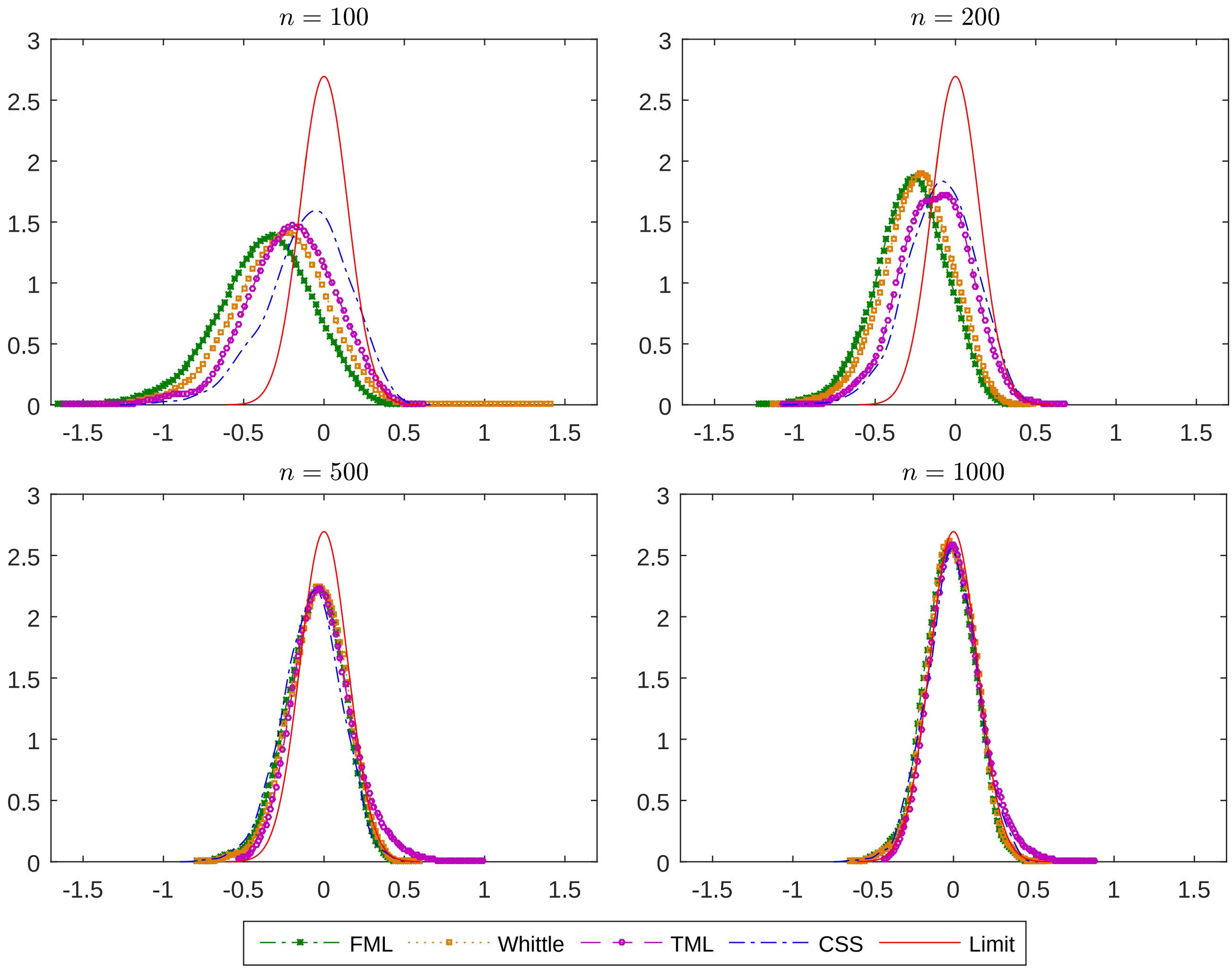

Figure 3.7: Kernel density of $n^{1/2}[\overline{\Lambda}_{dd}]^{-1/2}\left(\widehat{d}_1 - d_1\right)$ for an $ARFIMA(0, d_0, 1)$ TDGP with $d_0 = -0.2$ and $\theta_0 = -0.444978$, and an $ARFIMA(0, d, 1)$ MisM; $d^* = 0.25$.

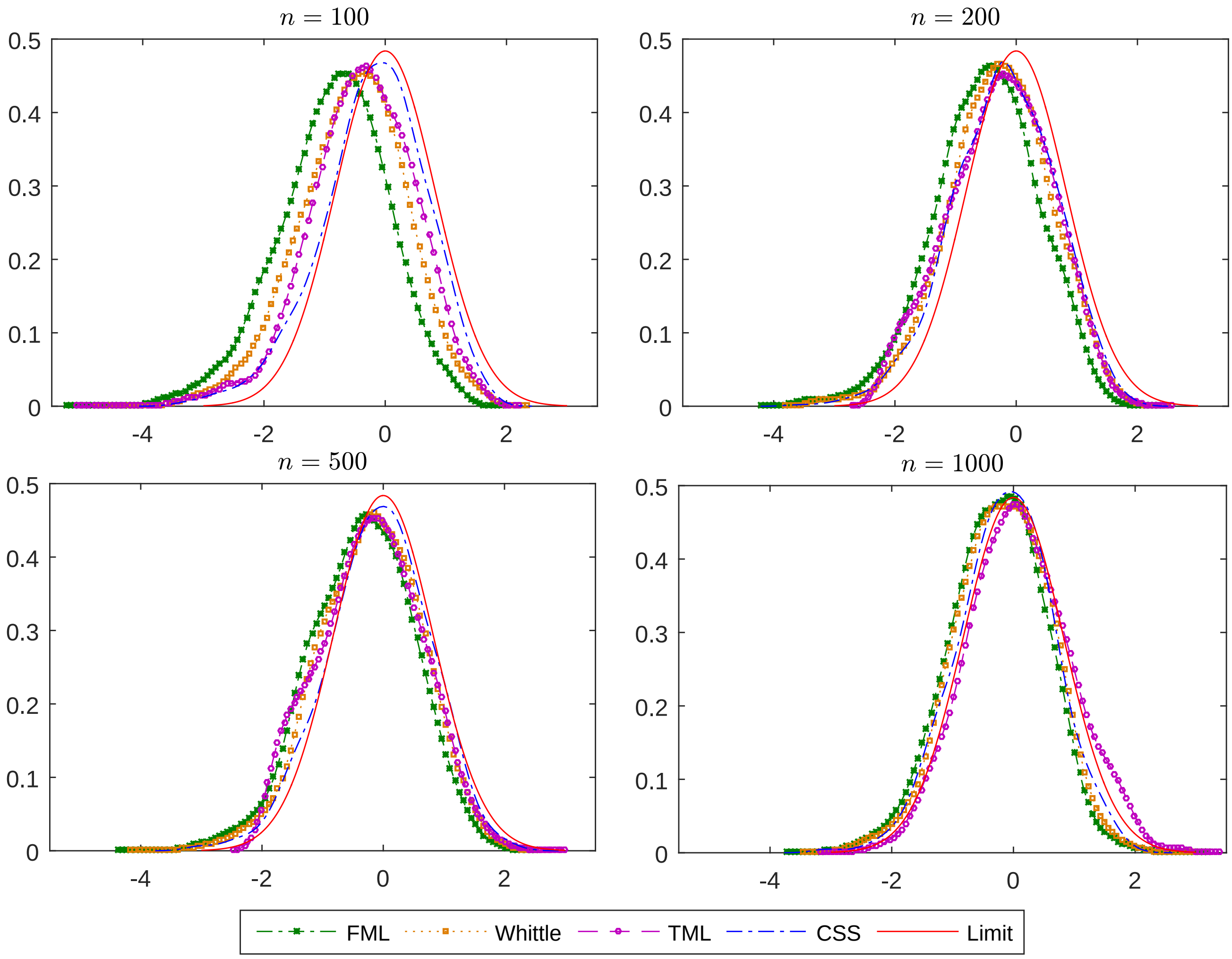

Figure 3.8: Kernel density of $\sqrt{n}\left(\widehat{d}_1 - d_1\right)$ for an $ARFIMA(0, d_0, 1)$ TDGP with $d_0 = -0.2$ and $\theta_0 = -0.3$, and an $ARFIMA(0, d, 1)$ MisM; $d^* < 0.25$.

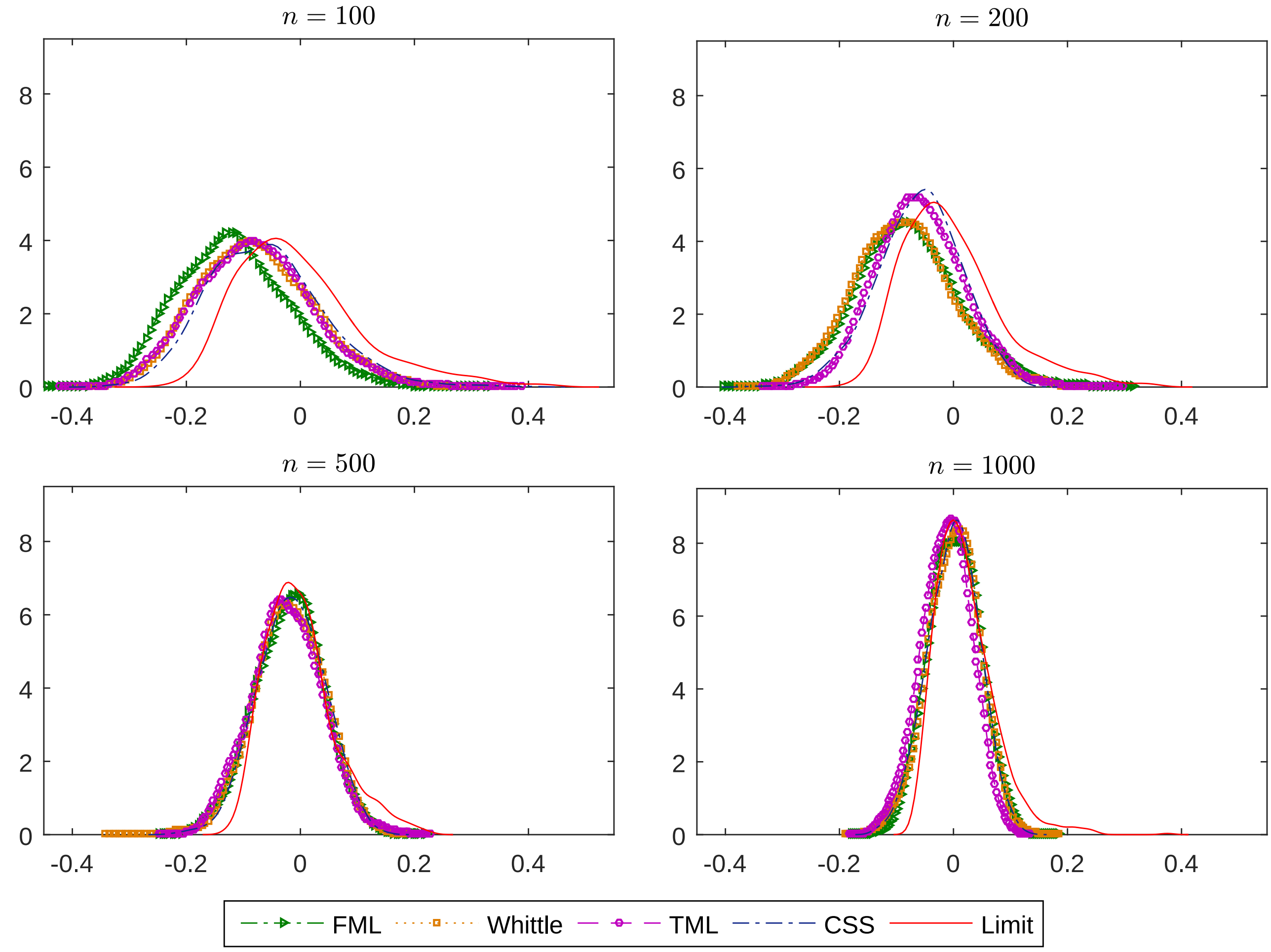

Figure 3.9: Kernel density of $\frac{n^{1-2d^*}}{\log n}\left(\widehat{d}_1 - d_1 - \mu_n\right)$ for an $ARFIMA(0, d_0, 1)$ TDGP with $d_0 = 0.4$ and $\theta_0 = -0.7$, and an $ARFIMA(0, d, 1)$ MisM; $d^* > 0.25$.

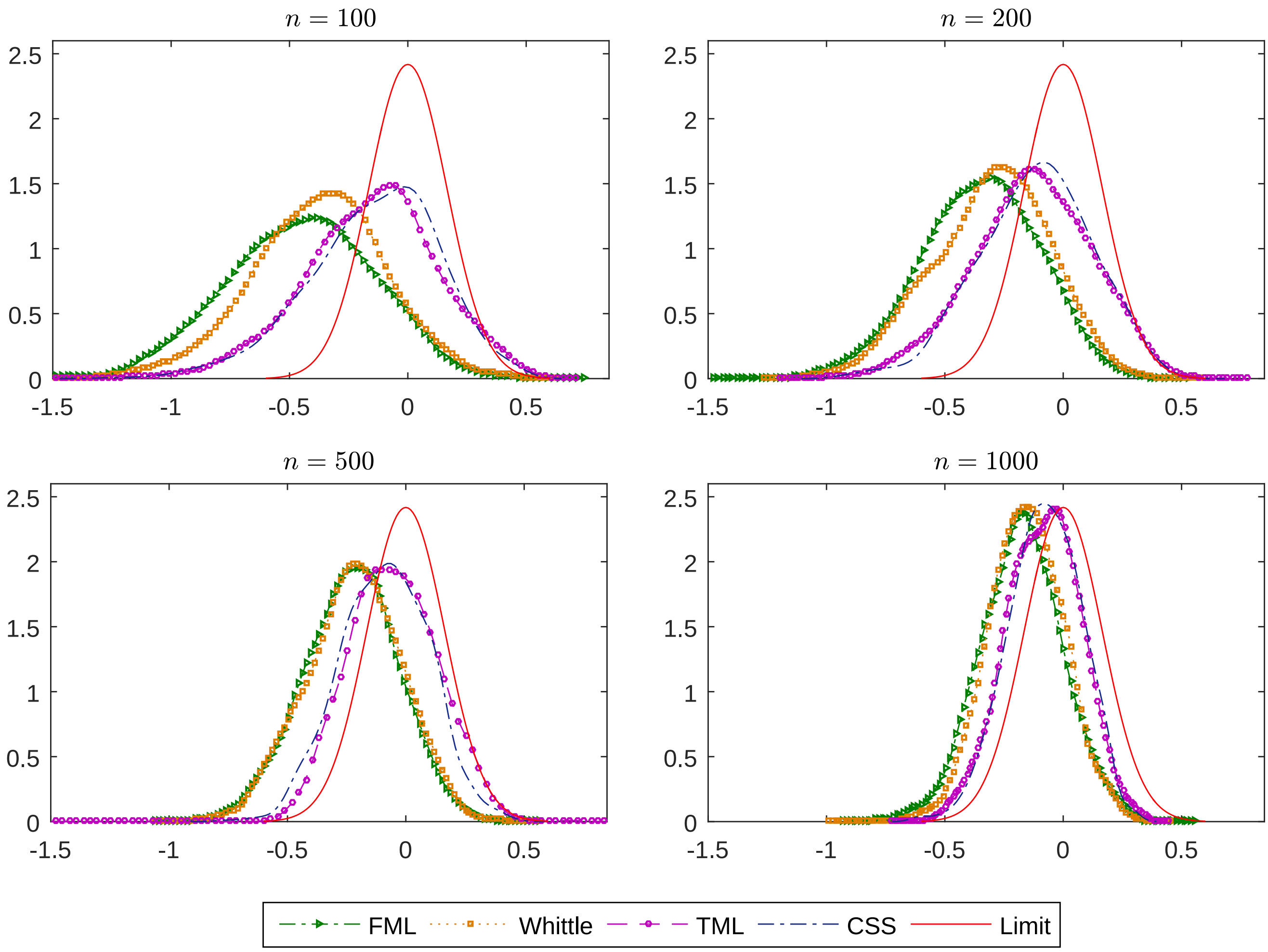

Figure 3.10: Kernel density of $n^{1/2}[\overline{\Lambda}_{dd}]^{-1/2}\left(\widehat{d}_1 - d_1\right)$ for an $ARFIMA(0, d_0, 1)$ TDGP with d_0= 0.4 and θ_0= -0.444978, and an $ARFIMA(0, d, 1)$ MisM; d^*= 0.25.

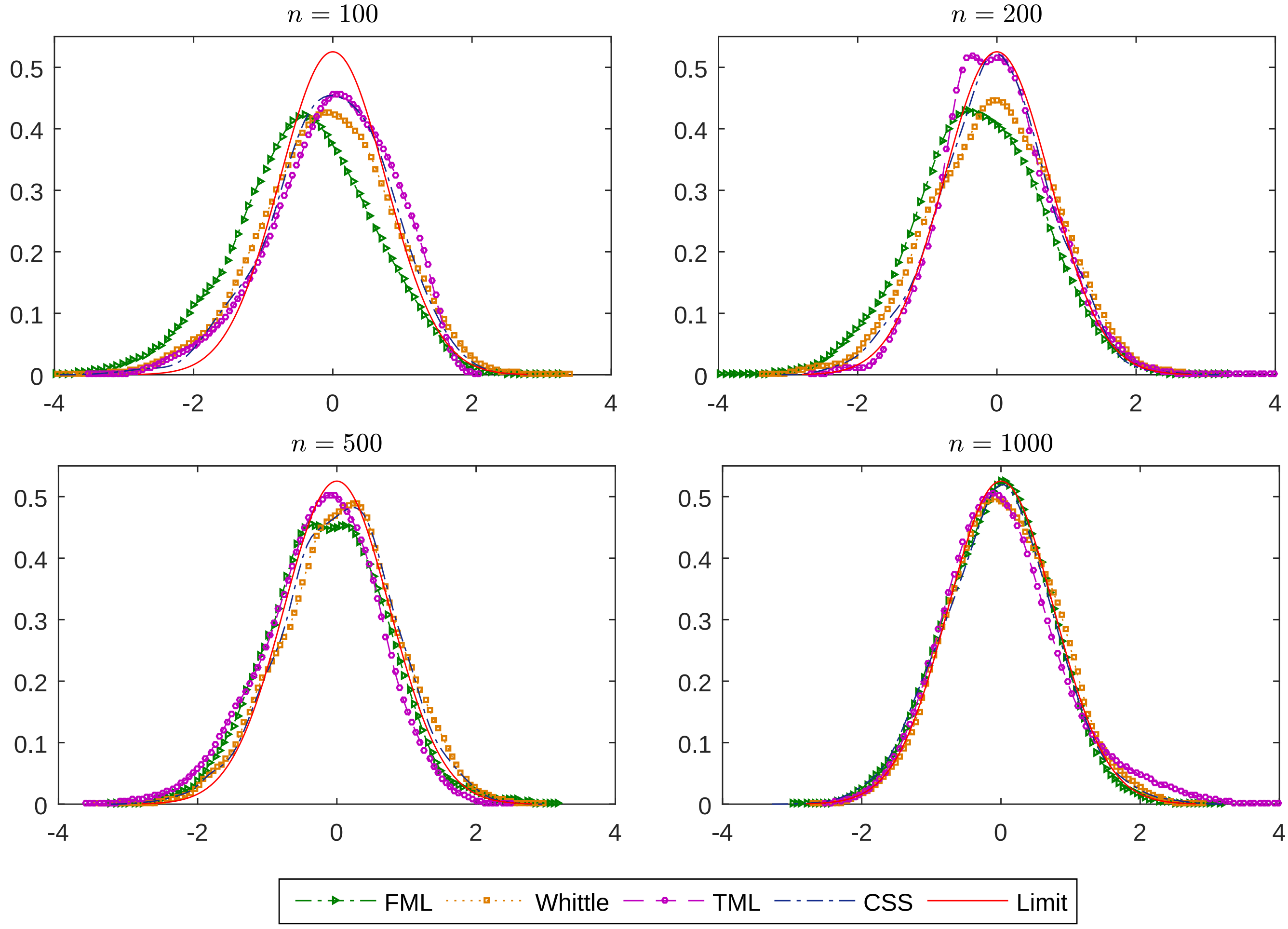

Figure 3.11: Kernel density of $\sqrt{n}\left(\widehat{d}_1 - d_1\right)$ for an $ARFIMA(0, d_0, 1)$ TDGP with $d_0 = 0.4$ and $\theta_0 = -0.3$, and an $ARFIMA(0, d, 1)$ MisM; $d^* < 0.25$.

Chapter 4

Mean correction in mis-specified fractionally integrated models

4.1 Introduction

The class of *fractionally integrated autoregressive moving average* [ARFIMA] models, as introduced by Granger and Joyeux (1980) and Hosking (1981) have been extremely useful for modelling long memory processes, particularly due to their tractable likelihoods and the ease of forecasting. However, the likelihoods presuppose that the structure of the true data generating process [TDGP] is correctly specified, apart from the values of a finite number of parameters that are to be estimated (see Giraitis and Surgailis, 1990, Beran, 1995 and Lieberman *et al.*, 2012, among others). Significant contributions to the issue of mis-specification in the context of long memory models have been made by Yajima (1993), Chen and Deo (2006) and Martin *et al.* (2018) (hereinafter referred to as Chapter 3 of this thesis). We contribute to this literature by producing a suite of theoretical developments on consequences of mis-specifying the short memory dynamics in ARFIMA models, under the assumption of an unknown mean, with attention given to five parametric estimation techniques: (i) Frequency domain maximum likelihood [FML], (ii) discrete Whittle [DWH], (iii) Exact Whittle [EWH], (iv) time domain maximum likelihood [TML], and (v) conditional sum of squares [CSS].

The aforementioned studies demonstrate that under mis-specification of the short memory dynamics, when the mean is *known*, the above parametric estimators converge to the same pseudo-true value. They are also shown to share the same limiting distribution whose form depends on the deviation between the true value of the fractional differencing parameter, d_0, and its pseudo-true value, d_1, $d^* = (d_0 - d_1)$. Further, the Monte-Carlo experiments presented in Chapter 3 shows that, in finite samples, the CSS estimator performs the best overall, when the mean is known. When the mean estimated by the sample mean, the DWH estimator is shown to outperform the others.

This chapter is interested in addressing two main questions: (i) How does the limiting behaviour of the parametric estimators of the dynamic parameters differ in the two cases of known and unknown process mean?, and, (ii) Can the effect of mean parameter estimation be significant enough to change the ranking of the parametric estimators, in terms of their finite sample performance, in comparison to the case of the known mean? From a practical point of view, it is restrictive to impose the requirement that the process mean is known. Hence, these questions are not only of theoretical interest but also of practical importance. For example, if the sample mean is used, then the limiting behaviour of any estimator of the dynamic parameters may be influenced by the slower rate of convergence (than the usual $\sqrt{n}$) that depends on d_0 (see Hosking, 1996), given that the estimator of d_0 itself has a slower rate of convergence under certain forms of mis-specification. The indications are that the limiting properties of the estimators of the dynamic parameters established in Chapter 3 can be extended to the case where the mean is unknown. However, the technical details are challenging. Since FML and DWH are invariant to mean, in this chapter we establish new asymptotic results only for the EWH, TML and CSS estimators, in the case of an estimated mean.

The two methods that are most commonly used for estimating the process mean in long memory models are the sample mean and the best linear unbiased estimator [BLUE]. Although the sample mean estimator is invariant to the specification of the model, the BLUE is influenced by the specifications of the model. Hence, we begin by deriving the asymptotic properties of the BLUE under incorrect specification of the model. We show that BLUE is consistent for the true mean even under mis-specification. We also derive the limiting distributions under the correct specification. In this case, our results coincide with the classical limit theory for estimating the mean, for example Adenstedt (1974).

We establish that the parametric estimators (FML, DWH, EWH, TML and CSS) of the parameters in the fitted model converge to the same pseudo-true value under mis-specification of the short memory dynamics, regardless of whether the process mean is known or unknown. We also show that the limiting distributions of the above five parametric estimators are identical. Thereby, we establish that the parametric estimators are asymptotically equivalent even when the process mean is unknown. Our simulation results show that the finite sample ranking of the five parametric estimators of the fractional differencing parameter – in terms of bias and root mean squared error [RMSE] – differs in the two cases of known and unknown mean. However, when it comes to estimation of the fractional differencing parameter, the *choice* of mean estimator does not have a significant influence on the ranking. Amongst the five parametric estimators, the DWH estimator displays the best overall finite sample performance, with the same qualitative results holding under both Gaussian and standardized Chi-squared errors.

The remainder of this chapter is organized as follows. In Section 4.2 we provide the assumptions and notation required to build the theoretical results. Section 4.3 contains several important results on mean estimators. The asymptotic properties of the three estimators – time domain maximum likelihood, conditional sum of squares and exact Whittle – are provided in Section 4.4. Section 4.5 establishes the asymptotic results corresponding to the unknown mean case. Section 4.6 contains the results of a Monte-Carlo study to evaluate finite sample performance. The proofs of all the results stated in the chapter are provided in the Appendix. Throughout the chapter, we use the notation "$\rightarrow^{P}$" to denote convergence in probability, and "$\rightarrow^{D}$" stands for convergence in distribution.[1]

4.2 Preliminaries: Assumptions and notation

Suppose that $\{y_t\}$ is generated from a stationary TDGP that is mean reverting with spectral density given by

$$\frac{\sigma_0^2}{2\pi} f_0(\lambda) = \frac{\sigma_0^2}{2\pi} \left|1 - \exp(-\imath\lambda)\right|^{-2d_0} g_0(\lambda), \ -\pi < \lambda < \pi, \tag{4.1}$$

where $d_0 \in (-0.5, 0.5)$, $\sigma_0^2 > 0$, and $g_0(.)$ is a spectral density continuous on the interval $[-\pi, \pi]$, bounded above and bounded away from zero with continuous second derivatives. Associated with (4.1) is an MA representation of $\{y_t\}$ expressed as follows,

$$y_t = \mu_0 + \textstyle\sum_{j=0}^{\infty} b_{0,j}\varepsilon_{t-j}, \ t \in \mathbb{Z}, \tag{4.2}$$

where μ_0 is the true process mean, $\{\varepsilon_t\}$ is a zero-mean white noise sequence with finite variance $\sigma_0^2 > 0$. Further, $\left\{b_{0,j}\right\}$ is a sequence of constants satisfying $b_{0,0} = 1$ and $\sum_{j=0}^{\infty} b_{0,j}^2 < \infty$ such that $(1-z)^{d_0} b_0(z) = c_0(z) = \sum_{j=0}^{\infty} c_{0,j} z^j$, and $|c_0(z)| > 0$, $|z| \leq 1$. Hence (4.1) can be expressed as $f_0(\lambda) = \left|b_0(\exp(\imath\lambda))\right|^2$, where $\lambda \in [-\lambda, \lambda]$. We suppose that $c_0(\exp(\imath\lambda))$ is differentiable in λ for $\lambda \neq 0$ and $\partial c_0(\exp(\imath\lambda)) / \partial\lambda = O\left(\left|\lambda^{-1}\right|\right)$ as $\lambda \rightarrow 0$. For the innovation process $\{\varepsilon_t\}$, $t \in \mathbb{Z}$, in (4.2), we assume that:

[1]As noted in Chapter 1, this chapter has been written as a draft for a self-contained article for journal submission. Hence, there is a certain amount of repetition of material presented in other chapters, and re-definition of notation.

$(A.1)$ For all $t \in \mathbb{Z}$, $E_0\left(\varepsilon_t|\mathcal{F}_{t-1}\right) = 0$ and $E_0\left(\varepsilon_t^2|\mathcal{F}_{t-1}\right) = \sigma_0^2$, a.s. where $\mathcal{F}_{t-1}$ is the sigma-field of events generated by ε_s, $s < t$. Hereinafter, the zero subscript denotes that the moments are defined with respect to the TDGP.

The estimated model has a spectral density given by

$$\frac{\sigma^2}{2\pi} f_1\left(\boldsymbol{\eta}, \lambda\right) = \frac{\sigma^2}{2\pi} \left|1 - \exp\left(-\imath\lambda\right)\right|^{-2d} g_1\left(\boldsymbol{\beta}, \lambda\right), \; -\pi < \lambda < \pi, \tag{4.3}$$

where $\boldsymbol{\eta} = \left(d, \boldsymbol{\beta}^\top\right)^\top$, with d as mentioned earlier and $\boldsymbol{\beta} \in \mathbb{B}$, with $\mathbb{B}$ an l-dimensional compact convex set in $\mathbb{R}^l$. Further, $g_1\left(\boldsymbol{\beta}, \lambda\right)$ is a real valued symmetric function such that $g_1\left(\boldsymbol{\beta}, \lambda\right) \neq g_0\left(\lambda\right)$ for all $\boldsymbol{\beta} \in \mathbb{B}$. We assume that $g_1\left(\boldsymbol{\beta}, \lambda\right)$ is continuous on the interval $[-\pi, \pi]$, bounded above and bounded away from zero with continuous second derivatives. We will also impose the following assumptions to prove the asymptotic results:

$(A.1a)$ $E_0\left(|\varepsilon_t|^{4+\tau}\right) < \infty$ for some $p \in (0, \infty)$ with $E_0\left(\varepsilon_t^3|\mathcal{F}_{t-1}\right) < \infty$ and $E_0\left(\varepsilon_t^4|\mathcal{F}_{t-1}\right) < \infty$, for all $t \in \mathbb{Z}$.

$(A.2)$ For all $\boldsymbol{\beta} \in \mathbb{B}$, $\int_{-\pi}^{\pi} \log g_1\left(\boldsymbol{\beta}, \lambda\right) d\lambda = 0$, and $\boldsymbol{\beta} \neq \boldsymbol{\beta}'$ implies that $g_1\left(\boldsymbol{\beta}, \lambda\right) \neq g_1\left(\boldsymbol{\beta}', \lambda\right)$ on a set of positive Lebesgue measure.

$(A.3)$ The function $g_1(\boldsymbol{\beta}, \lambda)$ is differentiable with respect to λ, with derivative $\partial g_1(\boldsymbol{\beta}, \lambda)/\partial\lambda$ continuous at all $(\boldsymbol{\beta}, \lambda)$, $\lambda \neq 0$, and $|\partial g_1(\boldsymbol{\beta}, \lambda)/\partial\lambda| = O(|\lambda|^{-1})$ as $\lambda \to 0$. Furthermore,

$$\inf_{\boldsymbol{\beta}} \inf_{\lambda} g_1\left(\boldsymbol{\beta}, \lambda\right) > 0 \text{ and } \sup_{\boldsymbol{\beta}} \sup_{\lambda} g_1\left(\boldsymbol{\beta}, \lambda\right) < \infty.$$

$(A.4)$ The function $g_1\left(\boldsymbol{\beta}, \lambda\right)$ is thrice differentiable with continuous third derivatives, such that,

$$\sup_{\lambda} \sup_{\boldsymbol{\beta}} \left|\frac{\partial g_1\left(\boldsymbol{\beta}, \lambda\right)}{\partial \beta_i}\right| \;\; < \;\; \infty, \; 1 \leq i \leq l,$$

$$\sup_{\lambda}\sup_{\beta}\left|\frac{\partial^2 g_1(\boldsymbol{\beta},\lambda)}{\partial\beta_i\partial\beta_j}\right| < \infty, \ \sup_{\lambda}\sup_{\beta}\left|\frac{\partial^2 g_1(\boldsymbol{\beta},\lambda)}{\partial\beta_i\partial\lambda}\right| < \infty,\ 1\le i\le l,\ 1\le i,j\le l,$$
$$\sup_{\lambda}\sup_{\beta}\left|\frac{\partial^3 g_1(\boldsymbol{\beta},\lambda)}{\partial\beta_i\partial\beta_j\partial\beta_k}\right| < \infty,\ 1\le i,j,k\le l.$$

Consider a user-assigned compact subset of the parameter space $(0, 0.5) \times \mathbb{B}$, where

$$\mathbb{E}_\delta = \mathbb{D}_\delta \times \mathbb{B} \quad \text{where} \quad \mathbb{D}_\delta = \{d : |d| \ \leq 0.5 - \delta\}\,, \text{ for some } 0 < \delta \ll 0.5\,. \tag{4.4}$$

$(A.5)$ There exists a unique pseudo-true parameter vector $\boldsymbol{\eta}_1 = \left(d_1, \boldsymbol{\beta}_1^\top\right)^\top$ belonging to the subset $\mathbb{E}_\delta^0$ that satisfies $\boldsymbol{\eta}_1 = \arg\min_{\boldsymbol{\eta}} Q(\boldsymbol{\eta})$, where

$$Q(\boldsymbol{\eta}) = \frac{\sigma_0^2}{2\pi}\int_0^\pi \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta},\lambda)}d\lambda, \tag{4.5}$$

and $\mathbb{E}_\delta^0$ is as defined in Chapter 3, page 50, that is, $\mathbb{E}_\delta^0 = \mathbb{D}_\delta^0 \times \mathbb{B}$ where $\mathbb{D}_\delta^0 = \mathbb{D}_\delta \cap \{d : -(1-2\delta) \leq (d_0 - d) \leq 0.5 - \delta\}$.

The assumptions $(A.1) - (A.5)$ and $(A.1a)$ are as stated in Chapter 3. Assumption $(A.1)$ is a weaker assumption compared to the classical requirement of Gaussianity. The differentiability conditions given in $(A.4)$ allow one to derive limit of the Fisher information matrix and to obtain the limiting distribution of the estimator under the incorrectly specified model. Assumption $(A.5)$ ensures that the pseudo-true parameters are well defined in $\mathbb{E}_\delta$, on which the convergence and distributional results are established. For more details on nature of all assumptions, see Chapter 3.

We shall now define an ARFIMA(p, d, q) model:

$$\phi(L)(1-L)^d\{y_t - \mu\} = \theta(L)\varepsilon_t, \tag{4.6}$$

where, $d \in (-0.5, 0.5)$, $\mu = E(y_t)$, L is the lag operator such that, $L^k x_t = x_{t-k}$ for $k \geqslant 0$ and $\phi(L) = 1 + \phi_1 L + ... + \phi_p L^p$ and $\theta(L) = 1 + \theta_1 L + ... + \theta_q L^q$ are the autoregressive and moving average operators respectively. Therefore, $\boldsymbol{\eta} = \left(d, \boldsymbol{\beta}^\top\right)^\top$ and $\boldsymbol{\beta} = \left(\phi_1, \phi_2, ..., \phi_p, \theta_1, \theta_2, ..., \theta_q\right)^\top$.

The realizations are generated from an ARFIMA(p_0, d_0, q_0) model, with the true spectral density, defined in (4.1), denoted by setting $g_0(\lambda) = |\theta_0(\exp(\imath\lambda))|^2 \big/ |\phi_0(\exp(\imath\lambda))|^2$. The realizations are modelled by an ARFIMA(p, d, q) model for any of the p and q such that $\{p \neq p_0 \cup q \neq q_0\} \setminus \{p_0 \leq p \cap q_0 \leq q\}$, with the spectral density defined in (4.3), setting $g_1(\boldsymbol{\beta}, \lambda) = |\theta(\exp(\imath\lambda))|^2 \big/ |\phi(\exp(\imath\lambda))|^2$. This model is referred to as the mis-specified model (abbreviated by 'MisM'). To comply with the assumptions stated above, the roots of the characteristic equations, $\theta_0(z)$, $\phi_0(z)$, $\theta(z)$ and $\phi(z)$ lie outside the unit circle so that the TDGP is a stationary process and the estimated model is also restricted to the stationary region.

4.3 Mean estimation and some useful results

In this section we introduce the two main estimators of the mean available in the class of fractionally integrated models; namely, the sample mean and the BLUE. A brief review of their statistical properties is provided when the model is correctly specified. As mentioned in Section 4.1, BLUE is influenced by the specification of the model. We then establish some interesting asymptotic properties of the BLUE, assuming that the model is mis-specified.

4.3.1 Sample mean

Let $\mathbf{y} = (y_1, y_2, ..., y_n)^\top$ be a random sample of n observations, with a spectral density as given in (4.1). The sample mean is given by $\widehat{\mu}_{SM} = \sum_{i=1}^n y_i \big/ n$. From (4.2), the sample mean is an unbiased estimator for μ. Hosking (1996) states that the sample mean is consistent for μ_0, but with a slower rate of convergence than $\sqrt{n}$, when $d_0 \in (-0.5, 0.5)$. Further, the limit distribution of $\widehat{\mu}_{SM}$ is given by

$$n^{1/2-d_0}\left(\widehat{\mu}_{SM} - \mu_0\right) \rightarrow^{D} N\left(0, \nu^2\right), \;\; with \;\; \nu^2 = \frac{\sigma_0^2 g_0\left(0\right)\Gamma\left(1-2d_0\right)}{\left(1+2d_0\right)\Gamma\left(1+d_0\right)\Gamma\left(1-d_0\right)}. \tag{4.7}$$

The asymptotic variance of the sample mean is thus seen to depend only on the behaviour of the spectrum at the origin. As described in Hosking (1996), the above theoretical results are valid under weaker conditions than Gaussianity of the time series.

4.3.2 BLUE for mean

An alternative unbiased estimator for mean is the BLUE introduced by Adenstedt (1974). The BLUE for mean is simply the weighted average of the sample observations. The weights are determined by the autocovariance matrix associated the fitted model for the DGP. Hence, the fitted model plays a key role here. In this section, we analyze the properties of BLUE under the following cases; (1) the fitted model and the model corresponding to the data generating process are the same, and, (2) the fitted model is incorrectly specified in terms of the short memory dynamics.

Case 1: Correct specification of the model

Denote by $\widehat{\mu}_{BLU,0}$ the BLUE and define the estimator by,

$$\widehat{\mu}_{BLU,0} = \frac{\mathbf{1}^{\top}\mathbf{\Sigma}_0^{-1}\mathbf{y}}{\mathbf{1}^{\top}\mathbf{\Sigma}_0^{-1}\mathbf{1}}, \tag{4.8}$$

where $\sigma_0^2\mathbf{\Sigma}_0 := \left[\gamma_{0,i-j}\left(\boldsymbol{\eta}\right)\right], i, j = 1, \ldots, n$, with $\gamma_{0,k}\left(\boldsymbol{\eta}\right)$ being the autocovariance at lag k of the series, $\{y_t : t = 1, \ldots, n\}$, and $\mathbf{1}$ is the column vector of n ones. The variance of BLUE is given by $\sigma_0^2\left(\mathbf{1}^{\top}\mathbf{\Sigma}_0^{-1}\mathbf{1}\right)^{-1}$ and for large n,

$$Var\left(\widehat{\mu}_{BLU,0}\right) \approx n^{2d_0-1}\frac{\sigma_0^2 g_0\left(0\right)\Gamma\left(1-2d_0\right)}{B\left(1-d_0, 1-d_0\right)},$$

from Theorem 5.2 of Adenstedt. Here $B(.)$ is the beta function. Here too, the asymptotic variance of the estimator depends only on the behaviour of the spectrum at the origin.

Provided that the data generating process is Gaussian, Samarov and Taqqu (1988) showed that in the ARFIMA setting BLUE is often better than the sample mean and that the asymptotic efficiency of $\widehat{\mu}_{SM}$ relative to $\widehat{\mu}_{BLU,0}$ is given by,

$$\breve{K} = \lim_{n \to \infty} \frac{Var\left(\widehat{\mu}_{BLU,0}\right)}{Var\left(\widehat{\mu}_{SM}\right)} = \frac{\pi d_0 \left(1 + 2d_0\right)}{B\left(1 - d_0, 1 - d_0\right) \sin\left(\pi d_0\right)}. \tag{4.9}$$

The interesting fact is that the ratio depends only on d_0 and the asymptotic efficiency is determined only by the behaviour of the spectral density at the origin, not its complete information. Figure 4.1 is the plot of asymptotic efficiency of the sample mean against the BLUE.[2] According to the plot, the asymptotic efficiency is less than one. This implies that the BLUE is uniformly (asymptotically) more efficient than the sample mean throughout the range of stationary values for the fractional differencing parameter. In particular, when $-0.5 < d_0 < 0$, BLUE is strictly preferred over the sample mean as the efficiency (of $\widehat{\mu}_{SM}$) is decreasing at an exponential rate as $d_0 \to -0.5$. If $d_0 = 0$, the limit of the asymptotic efficiency is one, and this implies that no loss in efficiency is incurred by the sample mean in large samples (see Grenander, 1954 for this result for the short memory case). When $0 < d_0 < 0.5$, the asymptotic efficiency tends to decrease as the differencing parameter moves away from zero and after 0.3 the efficiency increases and moves towards one.

The BLUE is obviously a feasible estimator *only* when the autocovariances of the TDGP is known. Hence, BLUE is not feasible in practice. One immediate remedy is to plug in the empirical counterpart of the autocovariances, the sample autocovariances. In this case,

[2] The plot given here is reproduction of Figure $2(a)$ of Samarov and Taqqu (1988). We reproduce the plot only in the stationary region.

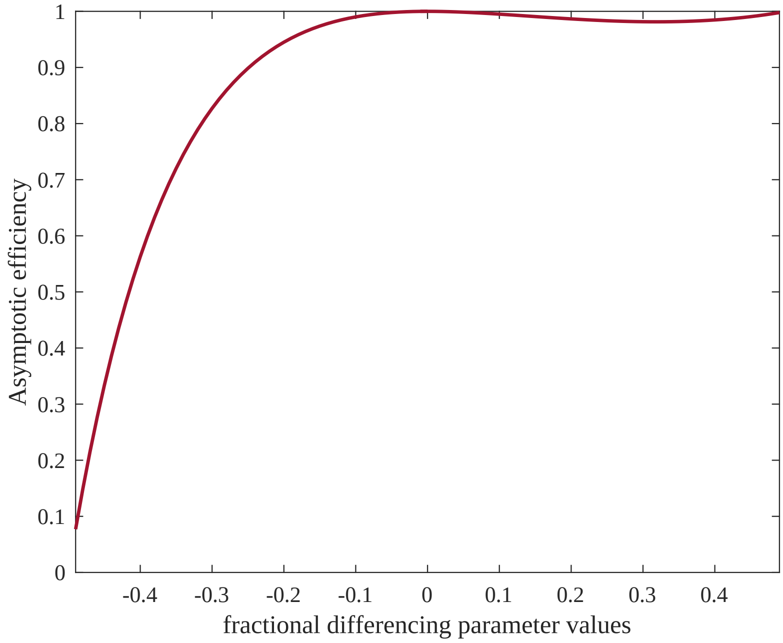

Figure 4.1: The plot displays the asymptotic efficiency of the sample mean against the best linear unbiased estimator, for different values of the fractional differencing parameter in a correctly specified Gaussian ARFIMA(p_0, d_0, q_0) model.

the efficiency of BLUE over the sample mean may not be achieved. Moreover, deriving the asymptotic results of this feasible BLUE is complicated. We do not aim to establish the large sample properties of the feasible estimator under the correct specification of the model.

Theorem 4.1 *Suppose that the TDGP of $\{y_t\}$ is as prescribed in equations (4.1) and (4.2).*

(i) *If $-0.5 < d_0 < 0.5$, then $n^{1/2-d_0}\left(\widehat{\mu}_{BLU,0} - \mu_0\right) \to^D N\left(0, \check{K}\nu^2\right)$, where $\check{K}$ is as defined in (4.9) and ν^2 is as defined in (4.7).*

(ii) *If $d_0 = -0.5$, then $n\left(\log n\right)^{-1/2}\left(\widehat{\mu}_{BLU,0} - \mu_0\right) \to^D N\left(0, 2\pi^{-1}\sigma_0^2 g_0\left(0\right)\check{K}\right)$.*

Theorem 4.1 provides the asymptotic distribution of the BLUE under the assumption that the correct model is identified for the given data generating process. The asymptotic distribution result is similar to that of the sample mean except that a multiplicative constant appears in the asymptotic variance of the BLUE. If $d_0 = 0$, then $\breve{K} = 1$ and in that case, the result coincides with the standard result for a short memory process (*cf*. Brockwell and Davis, 1991, Theorem 7.1.2).

Case 2: Mis-specification of the short memory dynamics in the model

Now define the BLUE under incorrect specification of the short memory dynamics associated with the TDGP as follows:

$$\widehat{\mu}_{BLU} = \frac{\mathbf{1}^\top \boldsymbol{\Sigma}_\eta^{-1} \mathbf{y}}{\mathbf{1}^\top \boldsymbol{\Sigma}_\eta^{-1} \mathbf{1}}, \tag{4.10}$$

where

$$\sigma^2 \boldsymbol{\Sigma}_\eta := \left[\gamma_{i-j} \left(\boldsymbol{\eta} \right) \right], \; i, j = 1, \ldots, n, \tag{4.11}$$

with $\gamma_k(\boldsymbol{\eta})$ being the autocovariance at lag k associated with the incorrectly specified model such that $\sigma^2 \boldsymbol{\Sigma} \neq \sigma_0^2 \boldsymbol{\Sigma}_0$.

Theorem 4.2 *Suppose that the TDGP of $\{y_t\}$ is as prescribed in equations (4.1) and (4.2), and that the MisM is specified as in (4.3). Assume that Assumptions $(A.1) - (A.4)$ are satisfied. Then $\widehat{\mu}_{BLU} = \mu_0 + o_p\left(n^{1-2d_0}\right)$.*

Theorem 4.2 states that even when a wrong model is chosen, the linear unbiased estimator is still consistent for the true mean. However, despite the impact of the estimator due to the choice of incorrect model being asymptotically negligible, it will be reflected in finite sample performance of both the BLUE and any estimator of $\boldsymbol{\eta}$ that depends on it.

4.4 Estimation of the dynamic parameters under mis-specification of the short memory dynamics

In this section, we define five estimation methods namely, FML, DWH, EWH, TML and CSS, for estimating the dynamic parameters under incorrect specification of the short memory dynamics in the model. In Section 4.4.1, we define the frequency domain estimation methods (FML, DWH and EWH), and then in Section 4.4.2 the time domain estimation methods (TML and CSS) are introduced.

Index by $i = 1, 2, 3, 4$ and 5 respectively, the FML, DWH, EWH, TML and CSS estimation method. Let $\widehat{\boldsymbol{\eta}}_1^{(i)}$ denote, respectively, the FML, DWH, EWH, TML and CSS estimator of the parameter vector $\boldsymbol{\eta} = \left(d, \boldsymbol{\beta}^\top\right)^\top$ of the MisM. Denote by $Q_n^{(i)}(\boldsymbol{\eta})$ the objective function of the i^{th} estimation method.

4.4.1 Frequency domain estimation

The three frequency domain estimators – FML, DWH and EWH – are defined in the following sections.

Approximate frequency domain maximum likelihood estimation

The FML objective function $Q_n^{(1)}(\boldsymbol{\eta})$ is defined by Chen and Deo (2006) as follows:

$$Q_n^{(1)}(\boldsymbol{\eta}) = \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j, \mu)}{f_1(\boldsymbol{\eta}, \lambda_j)}, \tag{4.12}$$

where $I(\lambda, \mu)$ is the periodogram defined as

$$I(\lambda, \mu) = |D(\lambda, \mu)|^2; \quad D(\lambda, \mu) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} (y_t - \mu) \exp(-\imath \lambda t), \tag{4.13}$$

and $D(\lambda, \mu)$ is the discrete Fourier transform (DFT) of the realizations, $\mathbf{y}$, being measured at Fourier frequencies, $\lambda_j = 2\pi j/n$; $(j = 1, ..., \lfloor n/2 \rfloor)$, where $\lfloor x \rfloor$ is the largest integer not greater than x. Here $\imath = \sqrt{-1}$ is the imaginary unit. The minimizer of $Q_n^{(1)}(\boldsymbol{\eta})$ is labelled as the FML estimator.

Suppose that Assumptions $(A.1)-(A.3)$ and $(A.5)$ hold, Proposition 1 of Chapter 3 states that $\lim_{n\to\infty} Q_n^{(1)}(\widehat{\boldsymbol{\eta}}_1^{(1)}) = Q(\boldsymbol{\eta})$ and $\widehat{\boldsymbol{\eta}}_1^{(1)} \to \boldsymbol{\eta}_1$ almost surely, where $Q(\boldsymbol{\eta})$ is as defined in (4.5).

Assumption $(A.5)$ is the most important assumption for the convergence either in probability or distribution of an estimator of $\boldsymbol{\eta}$ to hold under mis-specification when the mean is unknown. The minimum value of the limiting criterion function, $Q(\boldsymbol{\eta})$, defines the pseudo-true parameter value. For the class of stationary ARFIMA (mis-specified) models $Q(\boldsymbol{\eta})$ exists (see, Lemma 3.1 of Chapter 3).

Discrete version of the exact Whittle

An approximation to the EWH is the DWH estimation method and this method has been used in Beran (1994). The DWH log-likelihood function is given by,

$$Q_n^{(2)}(\boldsymbol{\eta}, \sigma^2) = \frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor} \log \frac{\sigma^2}{2\pi} f_1(\boldsymbol{\eta}, \lambda_j) + \frac{(2\pi)^2}{\sigma^2 n}\sum_{j=1}^{\lfloor n/2\rfloor} \frac{I(\lambda_j, \mu)}{f_1(\boldsymbol{\eta}, \lambda_j)}. \tag{4.14}$$

Remark 4.1 *The FML and DWH objective functions are invariant to mean (cf. Remark 2.1) and therefore both $Q_n^{(1)}(\boldsymbol{\eta})$ and $Q_n^{(2)}(\boldsymbol{\eta}, \sigma^2)$ are not expressed as functions of μ, even though the periodogram is defined as a function of μ (and λ_j).*

Unlike the FML estimator, DWH involves estimating $(\sigma^2, \boldsymbol{\eta})$. By concentrating out σ^2 in (4.14), the DWH estimator is obtained. The optimizer of σ^2 is evaluated as follows:

$$\frac{\partial}{\partial \sigma^2} Q_n^{(2)}(\boldsymbol{\eta}, \sigma^2) = \pi \frac{1}{\sigma^2} - \frac{(2\pi)^2}{(\sigma^2)^2 n}\sum_{j=1}^{\lfloor n/2\rfloor} \frac{I(\lambda_j, \mu)}{f_1(\boldsymbol{\eta}, \lambda_j)},$$

and this gives that

$$\widehat{\sigma}^2 = \frac{4\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I\left(\lambda_j, \mu\right)}{f_1\left(\boldsymbol{\eta}, \lambda_j\right)} = 2Q_n^{(1)}\left(\boldsymbol{\eta}, \mu_0\right).$$

Concentrating out σ^2 in (4.14) gives the associated profile function given by,

$$\begin{aligned} Q_n^{(2)}\left(\boldsymbol{\eta}\right) &= \pi \log \widehat{\sigma}^2 + \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \log f_1\left(\boldsymbol{\eta}, \lambda_j\right) - \pi \log \pi + \frac{(2\pi)^2}{\widehat{\sigma}^2 n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I\left(\lambda_j, \mu\right)}{f_1\left(\boldsymbol{\eta}, \lambda_j\right)} \\ &= \pi \log \widehat{\sigma}^2 + \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \log f_1\left(\boldsymbol{\eta}, \lambda_j\right) - \pi \log \pi + \frac{(2\pi)^2}{\widehat{\sigma}^2 n} \frac{n\widehat{\sigma}^2}{4\pi} \\ &= \pi \log \left[\frac{4}{n}\right] + \pi + \pi \log \left[\sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I\left(\lambda_j, \mu\right)}{f_1\left(\boldsymbol{\eta}, \lambda_j\right)}\right] + \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \log f_1\left(\boldsymbol{\eta}, \lambda_j\right). \end{aligned} \tag{4.15}$$

Minimizing the (4.15) with respect to $\boldsymbol{\eta}$ produces the DWH estimator, $\widehat{\boldsymbol{\eta}}_1^{(2)}$. Chapter 3 shows that $\widehat{\boldsymbol{\eta}}_1^{(2)} \rightarrow^{a.s} \boldsymbol{\eta}_1$.

For the EWH, TML and CSS estimation methods, the mean is also jointly estimated with the dynamic parameters. Hence, we impose the following additional assumption on the choice of the mean estimator:

$(A.6)$ For $\delta > 0$, $\widehat{\mu}_n = \mu_0 + o_p\left(n^{-1/2+d_0+\delta}\right)$.

This assumption has been adopted by Dahlhaus (1989) and Lieberman *et al.* (2012). The sample mean and the BLUE defined in (4.10) satisfy this assumption whenever $d_0 \in (-0.5, 0.5)$.

Exact Whittle estimation

This is an approximation to the time domain Gaussian maximum likelihood estimation procedure and has been exploited by Fox and Taqqu (1986) and Dahlhaus (1989) among others, by adopting the technique of Whittle (1952). The Exact Whittle estimation objective function is given by,

$$Q_n^{(3)}\left(\boldsymbol{\eta},\mu,\sigma^2\right)=\int_{-\pi}^{\pi}\left\{\log\left(\frac{\sigma^2}{2\pi}f_1\left(\boldsymbol{\eta},\lambda\right)\right)+\frac{2\pi I\left(\lambda,\mu\right)}{\sigma^2 f_1\left(\boldsymbol{\eta},\lambda\right)}\right\}d\lambda. \tag{4.16}$$

The EWH estimator of $\boldsymbol{\eta}$ is simply the minimizer of (4.16). The above objective function under exact Whittle estimation can be further simplified under Assumption $(A.2)$ as follows.

Consider the first component of the integral in (4.16):

$$\begin{aligned}\int_{-\pi}^{\pi}\log\left(\frac{\sigma^2}{2\pi}f_1\left(\boldsymbol{\eta},\lambda\right)\right)d\lambda &= \int_{-\pi}^{\pi}\log\left(\frac{\sigma^2}{2\pi}g_1(\boldsymbol{\beta},\lambda)\left|1-\exp\left(-\imath\lambda\right)\right|^{-2d}\right)d\lambda\\ &= 2\pi\log\left(\frac{\sigma^2}{2\pi}\right)+\int_{-\pi}^{\pi}\log g_1(\boldsymbol{\beta},\lambda)d\lambda\\ &\quad -2d\int_{-\pi}^{\pi}\log\left|1-\exp\left(-\imath\lambda\right)\right|d\lambda. \end{aligned}\tag{4.17}$$

In the above expression, the second component on the right hand side is exactly zero by Assumption $(A.2)$. The third component of (4.17) is deduced as follows using the standard result for $\int_0^{\pi}\log(\sin bx)\frac{dx}{1-2a\cos x+a^2}$ from Gradshtein and Ryzhik (2007, page 583):

$$\begin{aligned}2d\int_{-\pi}^{\pi}\log\left|1-\exp\left(-\imath\lambda\right)\right|d\lambda &= 2d\int_0^{\pi}\log\left(4\sin^2\left(\lambda/2\right)\right)d\lambda\\ &= 2d\int_0^{\pi}\log 4d\lambda+4d\int_0^{\pi}\log\left(\sin\left(\lambda/2\right)\right)d\lambda\\ &= 4\pi d\log 2+4d\pi\log\left(1/2\right)\\ &= 0.\end{aligned}$$

Therefore

$$\int_{-\pi}^{\pi}\log\left(\frac{\sigma^2}{2\pi}f_1\left(\boldsymbol{\eta},\lambda\right)\right)d\lambda=2\pi\log\left(\frac{\sigma^2}{2\pi}\right). \tag{4.18}$$

This leads (4.16) to become

$$Q_n^{(3)}\left(\boldsymbol{\eta},\mu,\sigma^2\right)=2\pi\log\left(\frac{\sigma^2}{2\pi}\right)+\int_{-\pi}^{\pi}\frac{2\pi I\left(\lambda,\mu\right)}{\sigma^2 f_1\left(\boldsymbol{\eta},\lambda\right)}d\lambda. \tag{4.19}$$

Unlike the FML and DWH objective functions, that of EWH is not invariant to the mean. Hence, we need to estimate both μ and σ^2 prior to estimating $\boldsymbol{\eta}$. Let us firstly find the estimator of σ^2 is as follows.

$$
\begin{aligned}
\frac{\partial}{\partial\sigma^2} Q_n^{(3)}\left(\boldsymbol{\eta},\mu,\sigma^2\right) &= 2\pi\frac{1}{\sigma^2} - \frac{2\pi}{\left(\sigma^2\right)^2}\int_{-\pi}^{\pi}\frac{I\left(\lambda,\mu\right)}{f_1\left(\boldsymbol{\eta},\lambda\right)}d\lambda. \\
\widehat{\sigma}^2 &= \int_{-\pi}^{\pi}\frac{I\left(\lambda\right)}{f_1\left(\boldsymbol{\eta},\lambda\right)}d\lambda. \qquad (4.20)
\end{aligned}
$$

Concentrating out σ^2 with (4.20) the EWH objective function is given by,

$$
Q_n^{(3)}\left(\boldsymbol{\eta},\mu\right) = 2\pi\log\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{I\left(\lambda,\mu\right)}{f_1\left(\boldsymbol{\eta},\lambda\right)}d\lambda\right) + 2\pi. \qquad (4.21)
$$

Then, the EWH estimator of $\boldsymbol{\eta}$, denoted by, $\widehat{\boldsymbol{\eta}}_1^{(3)}$, is obtained by minimizing the following function after replacing μ in (4.21) with some consistent estimate of μ_0 that satisfies Assumption $(A.6)$:

$$
Q_n^{(3)}\left(\boldsymbol{\eta},\widehat{\mu}\right) = 2\pi\log\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\widetilde{I}\left(\lambda,\widehat{\mu}\right)}{f_1\left(\boldsymbol{\eta},\lambda\right)}d\lambda\right) + 2\pi, \qquad (4.22)
$$

where,

$$
\tilde{I}\left(\lambda,\widehat{\mu}\right) = \left|\frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n}\left(y_t-\widehat{\mu}\right)\exp\left(-\imath\lambda t\right)\right|^2. \qquad (4.23)
$$

Lemma 4.1 *Let the TDGP be as prescribed in equations (4.1) and (4.2) and that the MisM is specified as in (4.3). Suppose Assumptions* $(A.1)-(A.3)$ *and* $(A.6)$ *hold. Then with probability* 1,

$$
\lim_{n\to\infty}\int_{-\pi}^{\pi}\frac{\widetilde{I}\left(\lambda,\widehat{\mu}\right)}{f_1\left(\boldsymbol{\eta},\lambda\right)} = \frac{\sigma_0^2}{2\pi}\int_{-\pi}^{\pi}\frac{f_0\left(\lambda\right)}{f_1\left(\boldsymbol{\eta},\lambda\right)}d\lambda,
$$

uniformly in $\boldsymbol{\eta}$ *on* $\mathbb{E}_\delta^0$.

4.4.2 Time domain estimation

In this section, the two time domain estimators – TML and CSS – are defined.

Maximum likelihood estimation

The Gaussian log-likelihood function for the TML estimator is

$$L\left(\boldsymbol{\eta},\mu,\sigma^2|\mathbf{y}\right)=-\frac{n}{2}\log\left(2\pi\sigma^2\right)-\frac{1}{2}\log|\boldsymbol{\Sigma}_{\eta}|-\frac{1}{2\sigma^2}\left(\mathbf{y}-\mu\mathbf{1}\right)^{\top}\boldsymbol{\Sigma}_{\eta}^{-1}\left(\mathbf{y}-\mu\mathbf{1}\right),\tag{4.24}$$

where $\sigma^2\Sigma_\eta$ is the variance covariance matrix of $\mathbf{y}$ derived from the mis-specified model as defined in (4.11). Maximizing (4.24) is equivalent to minimizing the criterion function

$$Q_n^{(4)}(\boldsymbol{\eta},\mu,\sigma^2)=\frac{n}{2}\log\left(\sigma^2\right)+\frac{1}{2}\log|\boldsymbol{\Sigma}_{\eta}|+\frac{1}{2\sigma^2}\left(\mathbf{y}-\mu\mathbf{1}\right)^{\top}\boldsymbol{\Sigma}_{\eta}^{-1}\left(\mathbf{y}-\mu\mathbf{1}\right).\tag{4.25}$$

Here, the parameters, μ, σ^2 and $\boldsymbol{\eta}$ are estimated by minimizing (4.25). Since the objective function involves estimating location, scale and dynamic parameters, we estimate $\boldsymbol{\eta}$ by concentrating out the location and scale parameters.

The first derivatives of $Q_n^{(4)}(\mu,\sigma^2,\boldsymbol{\eta})$ with respect to μ and σ^2 are

$$\begin{aligned}\frac{\partial}{\partial\mu}Q_n^{(4)}(\boldsymbol{\eta},\mu,\sigma^2) &= \frac{\partial}{\partial\mu}\left(\mathbf{y}-\mu\mathbf{1}\right)^{\top}\boldsymbol{\Sigma}_{\eta}^{-1}\left(\mathbf{y}-\mu\mathbf{1}\right)\\ &= \frac{\partial}{\partial\mu}\left[\mathbf{y}^{\top}\boldsymbol{\Sigma}_{\eta}^{-1}\mathbf{y}-2\mu\mathbf{1}^{\top}\boldsymbol{\Sigma}_{\eta}^{-1}\mathbf{y}+\mu^2\mathbf{1}^{\top}\boldsymbol{\Sigma}_{\eta}^{-1}\mathbf{1}\right]\\ &= -2\left(\mathbf{1}^{\top}\boldsymbol{\Sigma}_{\eta}^{-1}\mathbf{y}\right)+2\mu\mathbf{1}^{\top}\boldsymbol{\Sigma}_{\eta}^{-1}\mathbf{1},\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial\sigma^2}Q_n^{(4)}(\boldsymbol{\eta},\mu,\sigma^2) &= \frac{\partial}{\partial\sigma^2}\left[-\frac{n}{2}\log\left(\sigma^2\right)+\frac{1}{2}\log|\boldsymbol{\Sigma}_{\eta}^{-1}|+\frac{\sigma^2}{2}\left(\mathbf{y}-\mu\mathbf{1}\right)^{\top}\boldsymbol{\Sigma}_{\eta}^{-1}\left(\mathbf{y}-\mu\mathbf{1}\right)\right]\\ &= -\frac{n}{2}\frac{1}{\sigma^2}+\frac{1}{2}\left(\mathbf{y}-\mu\mathbf{1}\right)^{\top}\boldsymbol{\Sigma}_{\eta}^{-1}\left(\mathbf{y}-\mu\mathbf{1}\right).\end{aligned}$$

Solving the FOCs with respect to μ and σ^2 gives,

$$\tilde{\mu}=\left(\mathbf{1}^{\top}\boldsymbol{\Sigma}_{\eta}^{-1}\mathbf{1}\right)^{-1}\mathbf{1}^{\top}\boldsymbol{\Sigma}_{\eta}^{-1}\mathbf{y}=\widehat{\mu}_{BLU}\quad \textit{and}\quad \tilde{\sigma}^2=\frac{1}{n}\left(\mathbf{y}-\tilde{\mu}\mathbf{1}\right)^{\top}\boldsymbol{\Sigma}_{\eta}^{-1}\left(\mathbf{y}-\tilde{\mu}\mathbf{1}\right).\tag{4.26}$$

Under the assumption of a Gaussian process, the MLE of μ is $\widehat{\mu}_{BLU}$, and the finite sample properties of $\widehat{\mu}_{BLU}$ depend on whether the assumed model is correctly specified or not. In particular, if the model is correctly specified, then $\tilde{\mu}$ will be efficient relative to (say) the sample mean estimator of μ. However, it is appropriate to substitute into the modified profile log-likelihood (MPL) function, any estimator for μ, $\widehat{\mu}$, that satisfies assumption $(A.6)$, including the sample mean, yielding:

$$Q_n^{(4)}(\boldsymbol{\eta},\widehat{\mu}) = -\frac{n}{2}\log n + \frac{n}{2}\log\left((\mathbf{y}-\widehat{\mu}\mathbf{1})^\top \boldsymbol{\Sigma}_{\eta}^{-1}(\mathbf{y}-\widehat{\mu}\mathbf{1})\right) + \frac{1}{2}\log|\boldsymbol{\Sigma}_{\eta}| + \frac{n}{2}. \tag{4.27}$$

Then $\boldsymbol{\eta}$ can be estimated by maximizing (4.27). Denote the TML estimator here by $\widehat{\boldsymbol{\eta}}_1^{(4)}$.

Conditional sum of squares estimation

CSS estimation does not provide an estimator for σ^2 explicitly, as we do in maximum likelihood estimation. Instead, we estimate $\boldsymbol{\eta}$ by minimizing the sum of squares of residuals conditioning on $y_{t-s} = 0$, $\forall t \leqslant s$. Denote the CSS estimator by $\widehat{\boldsymbol{\eta}}_1^{(5)}$, with the CSS objective function given by

$$Q_n^{(5)}(\boldsymbol{\eta},\widehat{\mu}) = \frac{1}{n}\sum_{t=1}^{n} e_t^2, \tag{4.28}$$

where

$$e_t = \sum_{i=0}^{t-1} \tau_i(\boldsymbol{\eta})\{y_{t-i} - \widehat{\mu}\}, \tag{4.29}$$

with $\widehat{\mu}$ being replaced by a consistent estimator of the mean satisfying Assumption $(A.6)$. Here, $\tau_j(\boldsymbol{\eta})$ is such that $\tau_0(\boldsymbol{\eta}) = 1$ and

$$\tau_i(\boldsymbol{\eta}) = \sum_{s=0}^{i} \frac{k(i-s)\Gamma(i-d)}{\Gamma(i+1)\Gamma(-d)};\ \forall i > 0, \tag{4.30}$$

where $k(i-s)$ are the coefficients of the lag operators of the stationary and invertible ARMA process such that,

$$\frac{\phi(z)}{\theta(z)} = \sum_{i=0}^{\infty} k(i) z^i, \tag{4.31}$$

implying $\sum_{j=0}^{\infty} |k(j)| < \infty$. Using Sterling's approximation we can say that,

$$\tau_i(\boldsymbol{\eta}) = \frac{\kappa(1)}{\Gamma(-d)} i^{-d-1} \text{ as } i \to \infty \text{ for } |d| < 0.5. \tag{4.32}$$

Hence the CSS estimator is obtained as follows.

$$\widehat{\boldsymbol{\eta}}_1^{(5)} = \arg\min_{\boldsymbol{\eta}} Q_n^{(5)}(\boldsymbol{\eta}).$$

Remark 4.2 *When the process mean is unknown and is estimated by an estimator that is independent of $\boldsymbol{\eta}$ (for example, sample mean), the estimation of $\boldsymbol{\eta}$ is a two step procedure. In the first step, we demean the data using the estimator of mean and then in the second step, we estimate $\boldsymbol{\eta}$ as usual with the demeaned data. Suppose the estimator of mean is itself a function of $\boldsymbol{\eta}$ (for example, BLUE), then $\boldsymbol{\eta}$ is estimated in one step without explicitly estimating mean.*

4.5 Convergence properties of the parametric estimators under mis-specification

Suppose that $\{y_t\}$ is a long range dependent process with known mean. In Chapter 3 it is established that on subsets $\mathbb{E}_\delta$, $Q(\boldsymbol{\eta})$ exists and $\lim_{n\to\infty} Q_n^{(1)}(\boldsymbol{\eta}) = Q(\boldsymbol{\eta})$, where $Q(\boldsymbol{\eta})$ is defined as in Assumption $(A.5)$. Further, Theorem 3.1 in that chapter states that the convergence properties of FML, DWH, TML and CSS are the same and they converge to a common pseudo-true value. In Theorem 3.3, it is shown that all these four estimators are asymptotically equivalent.

Here we establish the limiting properties of the EWH, TML and CSS estimators in the case of an unknown mean. Then we relate the convergence properties of these three estimators to those of FML and DWH, in order to investigate whether all five estimators are asymptotically equivalent even when the mean is unknown.

Theorem 4.3 *Suppose that the TDGP of $\{y_t\}$ is as prescribed in equations (4.1) and (4.2), and that the MisM is specified as in (4.3). Assume that Assumptions $(A.1)-(A.3)$ and $(A.5)-(A.6)$ are satisfied. Then, $\lim_{n\to\infty}\left\|\widehat{\boldsymbol{\eta}}_1^{(i)}-\widehat{\boldsymbol{\eta}}_1^{(j)}\right\|=0$ almost surely for all $i,j=1,2,3,4,5$, where the common limiting value of $\widehat{\boldsymbol{\eta}}_1^{(i)}$, $i=1,2,3,4,5$, is $\boldsymbol{\eta}_1=\arg\min Q(\boldsymbol{\eta})$.*

Theorem 4.3 is an extension of Theorem 3.1 of Chapter 3 to the case where the process mean is unknown. The above theorem states that when the process mean is unknown, all five parametric estimators converge to a common pseudo-true parameter value under common mis-specification. This result holds when either the sample mean or the BLUE is used to estimate μ. Intuitively, this means that mean being estimated by either the sample mean or the BLUE does not alter the pseudo-true parameter value.

Theorem 4.4 *Suppose that the TDGP of $\{y_t\}$ is as prescribed in equations (4.1) and (4.2), and that the MisM is specified as in (4.3), and assume that Assumptions $(A.1)-(A.6)$ and $(A.1a)$ hold. Let*

$$\mathbf{B}=-\frac{\sigma_0^2}{\pi}\int_{-\pi}^{\pi}\frac{f_0(\lambda)}{f_1^3(\boldsymbol{\eta}_1,\lambda)}\frac{\partial f_1(\boldsymbol{\eta}_1,\lambda)}{\partial\boldsymbol{\eta}}\frac{\partial f_1(\boldsymbol{\eta}_1,\lambda)}{\partial\boldsymbol{\eta}^T}d\lambda+\frac{\sigma_0^2}{2\pi}\int_{-\pi}^{\pi}\frac{f_0(\lambda)}{f_1^2(\boldsymbol{\eta}_1,\lambda)}\frac{\partial^2 f_1(\boldsymbol{\eta}_1,\lambda)}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T}d\lambda\,, \tag{4.33}$$

and set $\boldsymbol{\mu}_n=\mathbf{B}^{-1}E_0\left(\frac{\partial Q_n(\boldsymbol{\eta}_1)}{\partial\boldsymbol{\eta}}\right)$ where $Q_n(\cdot)$ denotes the objective function that defines $\widehat{\boldsymbol{\eta}}_1$. Let $\widehat{\boldsymbol{\eta}}_1$ denote the estimator obtained by minimizing $Q_n(\boldsymbol{\eta})$ over the compact set $\mathbb{E}_\delta$ where $\boldsymbol{\eta}_1\in\mathbb{E}_\delta$ and assume that $\boldsymbol{\eta}_1\ni\partial\mathbb{E}_\delta$ where $\partial\mathbb{E}_\delta$ is the boundary of the set $\mathbb{E}_\delta$. Then the limiting distribution of any one of the FML, DWH, EWH, TML or CSS estimators is as delineated in Cases 4.1, 4.2 and 4.3:

Case 4.1 *When $d^*=d_0-d_1>0.25$,*

$$\frac{n^{1-2d^*}}{\log n}\left(\widehat{\boldsymbol{\eta}}_1-\boldsymbol{\eta}_1-\boldsymbol{\mu}_n\right)\to^D\mathbf{B}^{-1}\left[\sum_{j=1}^{\infty}W_j,0,...,0\right]^{\top}, \tag{4.34}$$

where $\sum_{j=1}^{\infty} W_j$ *is the mean square limit of the random sequence* $\sum_{j=1}^{s} W_j$ *as* $s \to \infty$ *wherein*

$$W_j = \frac{(2\pi)^{1-2d^*} g_0(\boldsymbol{\eta}_0, 0)}{j^{2d^*} g_1(\boldsymbol{\eta}_1, 0)} \left[U_j^2 + V_j^2 - E_0 \left(U_j^2 + V_j^2 \right) \right],$$

where, $\{U_j, V_k\}$ *are a sequence of random normal variables with zero mean and the covariance structure of* $\{U_j, V_k\}$ *denoted by* $Cov_0(.)$ *is as follows,*

$$\begin{aligned} Cov_0\left(U_j, V_k\right) &= \iint\limits_{[0,1]^2} \{\sin(2\pi jx)\sin(2\pi ky) + \sin(2\pi kx)\sin(2\pi jy)\} \, |x-y|^{2d_0-1} \, dxdy, \\ Cov_0\left(U_j, U_k\right) &= Cov_0\left(U_j, V_k\right) = Cov_0\left(V_j, V_k\right), \forall j, k \in \mathbb{N}. \end{aligned} \tag{4.35}$$

Case 4.2 *When* $d^* = d_0 - d_1 = 0.25$,

$$n^{1/2}\overline{\Lambda}^{-1/2} \left(\widehat{\boldsymbol{\eta}}_1 - \boldsymbol{\eta}_1\right) \to^D \mathbf{B}^{-1} \left(Z, 0, ..., 0\right)^\top, \tag{4.36}$$

where Z *is a standard normal random variable and*

$$\overline{\Lambda} = \frac{1}{n} \sum_{j=1}^{n/2} \left(\frac{f_0(\lambda_j)}{f_1(\boldsymbol{\eta}_1, \lambda_j)} \frac{\partial \log f_1(\boldsymbol{\eta}_1, \lambda_j)}{\partial d} \right)^2. \tag{4.37}$$

Case 4.3 *When* $d^* = d_0 - d_1 < 0.25$,

$$\sqrt{n} \left(\widehat{\boldsymbol{\eta}}_1 - \boldsymbol{\eta}_1\right) \to^D N(0, \Xi), \tag{4.38}$$

where $\Xi = \mathbf{B}^{-1}\mathbf{\Lambda}\mathbf{B}^{-1}$, *and*

$$\mathbf{\Lambda} = 2\pi \int_0^{\pi} \left(\frac{f_0(\lambda)}{f_1(\boldsymbol{\eta}_1, \lambda)} \right)^2 \left(\frac{\partial \log f_1(\boldsymbol{\eta}_1, \lambda)}{\partial \boldsymbol{\eta}} \right) \left(\frac{\partial \log f_1(\boldsymbol{\eta}_1, \lambda)}{\partial \boldsymbol{\eta}} \right)^\top d\lambda. \tag{4.39}$$

Theorem 4.4 is an extension of Theorem 3.3 of Chapter 3 showing that all five parametric estimators share the same limiting distribution even when the process mean is unknown. This implies that estimating the mean does not change the form of limiting distribution of the parametric estimators of the dynamic parameters. Again, this result holds when either the sample mean or the BLUE is used to estimate μ. When the difference between the true and the pseudo-true value of d, $d^* \,(= d_0 - d_1) > 0.25$, the parametric estimator still converges to a non-Gaussian distribution which is a function of an infinite sum of non-Gaussian random variables. Further, the slower rate of convergence $n^{1-2d^*} \big/ \log n$, under the known mean case, is not affected by the slower rate of convergence of the mean estimator. Although asymptotic normality is preserved when $d^* = 0.25$, the rate of convergence is of order $\left(n / \log^3 n\right)^{1/2}$, as is the case in the known mean case. When $d^* < 0.25$, we continue to achieve asymptotic normality with the usual $\sqrt{n}-$ rate of convergence.

To establish that the limiting distribution is the same for all five parametric estimators, we investigate the first-order Taylor expansion of $\partial Q_n^{(i)}(\boldsymbol{\eta}_1, \mu_0) \big/ \partial \boldsymbol{\eta}$ about$\partial Q_n^{(i)}(\widehat{\boldsymbol{\eta}}_1, \widehat{\mu}) \big/ \partial \boldsymbol{\eta} = 0$, (only) for $i = 3$, 4 and 5, as the limiting distribution of the FML and DWH estimators are invariant to mean. This leads to

$$
\begin{aligned}
R_n\left(\widehat{\boldsymbol{\eta}}_1^{(i)} - \boldsymbol{\eta}_1\right) \;=\; & -\left(\left[\partial^2 Q_n^{(i)}(\overline{\boldsymbol{\eta}}, \widehat{\mu})\right]^{-1} - \left[\partial^2 Q_n^{(i)}(\overline{\boldsymbol{\eta}}, \mu_0)\right]^{-1}\right) \times \left[R_n \times \partial Q_n^{(i)}(\boldsymbol{\eta}_1, \mu_0)\right] \\
& - \left[\partial^2 Q_n^{(i)}(\overline{\boldsymbol{\eta}}, \mu_0)\right]^{-1} \times \left[R_n \times \partial Q_n^{(i)}(\boldsymbol{\eta}_1, \mu_0)\right], \qquad (4.40)
\end{aligned}
$$

where R_n is the rate of convergence applicable in the three different cases outlined in the theorem. In the above expansion, the second component has the limiting distribution that is defined in theorem. If the first component in (4.40) converges to zero in probability, then it follows that the limiting distribution of the estimator (EWH or TML or CSS) under the unknown mean case is the same as that under the known mean case. Indeed we show that

$$
\operatorname*{plim}_{n\to\infty}\left(\left[\partial^2 Q_n^{(i)}(\overline{\boldsymbol{\eta}}, \widehat{\mu})\right]^{-1} - \left[\partial^2 Q_n^{(i)}(\overline{\boldsymbol{\eta}}, \mu_0)\right]^{-1}\right) \times \left[R_n \times \partial Q_n^{(i)}(\boldsymbol{\eta}_1, \mu_0)\right] = 0. \qquad (4.41)
$$

The idea behind condition (4.41) is to explore the limiting behaviour of the distance of between the second order derivative of the objective function evaluated at a given point measured with mean estimator, $(\overline{\boldsymbol{\eta}}, \widehat{\mu})$ and another the point measured at the true mean, $(\overline{\boldsymbol{\eta}}, \mu_0)$. This implicitly verifies whether the fact that the mean being treated as another parameter to be estimated makes any change in the asymptotic behaviour of $\partial Q_n^{(i)}(.)$.

4.6 Simulation results

In this section, a Monte-Carlo study has been carried out to explore the finite sample performance of the five parametric estimators of the pseudo-true value of the long memory parameter, d_1, under specific types of mis-specification. The five parametric estimators focused on here are: $\widehat{d}_1^{(1)}$ (FML), $\widehat{d}_1^{(2)}$ (DWH), $\widehat{d}_1^{(3)}$ (EWH), $\widehat{d}_1^{(4)}$ (TML) and $\widehat{d}_1^{(5)}$ (CSS), as are defined in Sections 4.4.1 and 4.4.2. For each of the estimators, Monte-Carlo results are obtained for one specific value corresponding to the cases, $d^* > 0.25$, $d^* = 0.25$, and $d^* < 0.25$, as given in Theorem 4.4. We firstly document the form of the finite sample distributions for each of the standardized versions of these estimators and compare with corresponding asymptotic distribution associated with each of the three cases based on d^*. We then tabulate the performance based on bias and RMSE. For our Monte-Carlo study, MATLAB 2015*b*, version 8.6.0.267246 is used.

Here we consider a particular form of mis-specification, which will be defined later. For each form of mis-specification, we generate $100,000$ artificial time series from a zero mean $ARFIMA(p_0, d_0, q_0)$ process with sample size, $n = 100$, 200, 500 and 1000 as follows. For a given sample size, we generate a vector of *i.i.d.* random variables, denoted by $\mathbf{e}$, obtained from two distributions, (i) standard normal, and, (ii) standardized Chi-square with 4 degrees of freedom. We then compute the analytic autocovariance matrix associated with the DGP, $\boldsymbol{\Sigma}_0$, together with its Cholesky decomposition, $\mathbf{C}$, $\boldsymbol{\Sigma}_0 = \mathbf{C}\mathbf{C}^\top$. The simulated series is then constructed as: $\mathbf{y} = \mu_0 \mathbf{1} + \mathbf{Ce}$, where μ_0 is the true mean of the process. It has been suggested in the literature (see, Granger and Joyeux, 1980) that, in order for the long memory ARFIMA series $\{y_t\}$ not to be affected by "the initial values", it is advisable to simulate a longer series and drop the first subset of values. We find that such a practice is unnecessary if $\{y_t\}$ is simulated using our method. The type of mis-specification considered in our simulation study is as follows:

Example 4.1 *An ARFIMA$(0, d_0, 1)$ TDGP, with parameter values $d_0 = \{-0.25, 0.2, 0.45\}$ and $\theta_0 = \{-0.7, -0.444978, -0.3\}$; and the MisM is ARFIMA$(0, d, 0)$. The values of θ_0 considered here correspond to the cases, $d^* > 0.25$, $d^* = 0.25$, and $d^* < 0.25$, respectively.*

In Subsection 4.6.1 we explore the forms of the finite sample distributions of the standardized versions of the five estimators under the first type of mis-specification described above, and for $d_0 = 0.2$ only. Both time domain estimators are obtained under the assumption of an unknown mean and estimated with both the sample mean and the BLUE. We omit the graphical results of the estimators for the values of $d_0 = -0.25$ and 0.45 as the finite sample distributions are qualitatively the same as those reported here.

We then proceed to report the bias and RMSE of the five estimators (in terms of estimating the pseudo-true value d_1) for all three values of d_0. We report the bias and RMSE estimates for the sample sizes $n = 100$, 500 and 1000 only due to space limitation. The results corresponding to the $n = 200$ case do support the entire interpretation of the simulation study. The FML and DWH estimators are invariant to the process mean and hence are obtained by allowing the mean to be known. However, TML, CSS and EWH are not so. The bias and RMSE results for the latter estimators are obtained under the following scenarios; (i) process mean is known, and, (ii) process mean is unknown and estimated with both the sample mean and the BLUE. The results for the known mean case are presented in Section 4.6.2, and those for the unknown mean case in Section 4.6.3. The following standard formulae of bias and RMSE are applied to all five estimators. In the formulae, we use $\widehat{d}_1$ to denote the generic estimator. When the necessity of denoting a particular parametric estimator arises, we do re-introduce the superscript notation. These formulae of bias and variances are evaluated under the TDGP and hence we make this explicit with appropriate subscript notation '0'. The estimate of the expectation of $\widehat{d}_1$ is given by,

$$\widehat{E}_0\left(\widehat{d}_1\right) = \frac{1}{R}\sum_{r=1}^{R}\widehat{d}_{1,r}.$$

Thus, the estimators of bias, sampling variance, and RMSE are as follows,

$$\begin{aligned}
\widehat{Bias}_0\left(\widehat{d}_1\right) &= \widehat{E}_0\left(\widehat{d}_1\right) - d_1 = \frac{1}{R}\sum_{r=1}^{R}\widehat{d}_r - d_1, \\
\widehat{Var}_0\left(\widehat{d}_1\right) &= \widehat{E}_0\left(\widehat{d}_1\right)^2 - \left(\widehat{E}_0\left(\widehat{d}_1\right)\right)^2 = \frac{1}{R}\sum_{r=1}^{R}\left(\widehat{d}_{1,r}\right)^2 - \left(\frac{1}{R}\sum_{r=1}^{R}\widehat{d}_{1,r}\right)^2, \\
\widehat{MSE}_0\left(\widehat{d}_1\right) &= \widehat{Bias}_0^2 + \widehat{Var}_0\left(\widehat{d}_1\right),
\end{aligned}$$

and the RMSE is the square root of the MSE value.

4.6.1 Finite sample distributional results

We consider the three cases of Theorem 4.4. Again, for notational simplicity, denote by $\widehat{d}_1$ any of the parametric estimators obtained under mis-specification and denote by $Q_n\left(.\right)$ the relevant objective function. The form of mis-specification we discuss here has only one parameter to be estimated. Therefore, the matrices and vectors associated with Theorem 4.4 reduce to scalars.

In the figures, the 'Limit' curve denotes the limiting distribution of the relevant statistic.

The 'FML', 'DWH', 'EWH', 'TML' and 'CSS' curves corresponds to the sampling distributions of the standardized statistics obtained under respective estimation methods with the known mean of zero. We use 'EWH_{SM}' and 'EWH_{BL}' to denote the sampling distribution of the statistic obtained under EWH estimation method with the process mean being estimated by sample mean and BLUE, respectively. Similar notation is used for the TML and CSS estimators when the mean is estimated. The first column of the figures represents the kernel densities of the FML, DWH, EWH, TML and CSS estimators obtained under the known mean case. The second column displays the kernel densities corresponding to the EWH, TML and CSS estimators obtained by estimating the mean with both the sample mean and the BLUE, plus the FML and DWH densities.

Case 1: $d^* > 0.25$

The limiting distribution of $\widehat{d_1}$ in this case is as follows,

$$\frac{n^{1-2d^*}}{\log n}\left(\widehat{d_1} - d_1 - \mu_n\right) \to^D b^{-1}\sum_{j=1}^{\infty} W_j, \tag{4.42}$$

where, μ_n is as mentioned earlier and can be evaluated for each estimation method in finite samples using the formula provided in Appendix 3.B of Chapter 3. Also, b is the first element in the matrix of $\mathbf{B}$ which is evaluated as follows,

$$\begin{aligned} b &= -2\int_{-\pi}^{\pi} \frac{f_0(\lambda)}{f_1^3(d_1,\lambda)}\left(\frac{\partial f_1(d_1,\lambda)}{\partial d}\right)^2 d\lambda + \int_{-\pi}^{\pi}\frac{f_0(\lambda)}{f_1^2(d_1,\lambda)}\left(\frac{\partial^2 f_1(d_1,\lambda)}{\partial d^2}\right)d\lambda \\ &= -2\int_{0}^{\pi}(1+\theta_0^2+2\theta_0\cos(\lambda))(2\sin(\lambda/2))^{-2d^*}(\log(2\sin(\lambda/2)))^2 d\lambda, \end{aligned} \tag{4.43}$$

and $W_j = \frac{(2\pi)^{1-2d^*}(1+\theta_0^2)}{j^{2d^*}}\left[U_j^2 + V_j^2 - E_0(U_j^2+V_j^2)\right]$, with U_j and V_j as mentioned in Theorem 4.4. The limiting distribution associated with the case $d^* > 0.25$ has a non-standard distribution that is a series of non-Gaussian random variables. Further, the distribution does not possess a closed-form representation. To this end, we shall obtain the limiting distribution by numerically truncating the series of the random variable W_j at some value s, as suggested in Section 3.5.2 in Chapter 3. The basic idea behind the methodology is to find the appropriate choice of the truncation point of the series, (say, $\sum_{j=1}^{s} W_j$, $1 \leq s \leq \lfloor n/2 \rfloor$) such that the difference between the variance of $\sum_{j=1}^{s} W_j$ and the estimated variance of the FML estimator (this stands as a benchmark) is minimized. Then, the 'Limit' curve is obtained using this value of s.

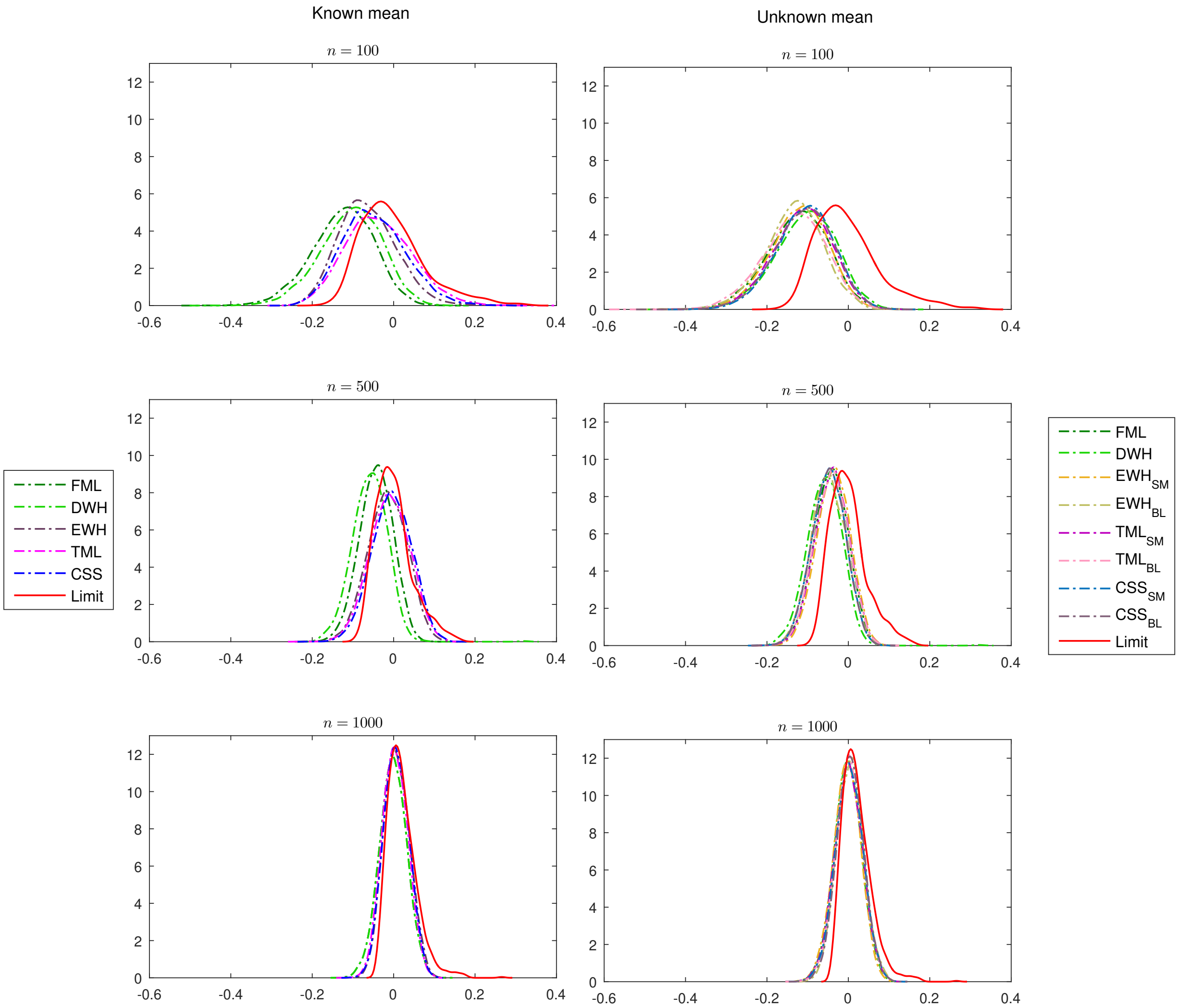

Figure 4.2: Kernel density of $\frac{n^{1-2d^*}}{\log n}(\widehat{d}_1 - d_1 - \mu_n)$ for an ARFIMA$(0, d_0, 1)$ TDGP with $d_0 = 0.2$ and $\theta_0 = -0.7$ and an ARFIMA$(0, d, 0)$ MisM, $d^* > 0.25$.

Figure 4.2 provides the kernel density estimate of $\frac{n^{1-2d^*}}{\log n}(\widehat{d}_1 - d_1 - \mu_n)$ under the five estimation methods considering both mean being known (presented in the first column of the figure) and unknown (presented in the second column of the figure), for a specification as labeled above each plot, plus the limiting distribution given in (4.42) for the given s. The particular parameter values employed in the specification of the TGDP are $d_0 = 0.2$ and $\theta_0 = -0.7$ with $d^* = 0.3723$ in this case. Each panel of Figure 4.2 indicates that $\frac{n^{1-2d^*}}{\log n}(\widehat{d}_1 - d_1 - \mu_n)$ is centered to the left of zero for all sample sizes, for all estimation methods, with the CSS estimation method under the known mean case being the closest to the limiting distribution in the known mean case and DWH being the closest in the unknown mean case. When the process mean is known, we observe that two distinct clustering occur amongst the five estimation methods for all the sample sizes. One cluster is formed with the EWH, TML and CSS estimation methods.[3] The other is formed with the FML and DWH methods. However, we do not observe such a clustering when the mean is unknown and estimated with either the sample mean or the BLUE. Moreover, the distance between the distributions of the different types of estimators is less substantial in the estimated mean case, at least visually. Nevertheless, as the sample size increases the point of central location of $\frac{n^{1-2d^*}}{\log n}(\widehat{d}_1 - d_1 - \mu_n)$ approaches zero and all distributions of the standardized statistics go close to matching the limiting distribution, regardless of whether the mean is known or not.

Case 2: $d_0 - d_1 = 0.25$

The simplified version of the form of the limiting distribution is as follows:

$$n^{1/2}\left[\overline{\Lambda}_{dd}\right]^{-1/2}\left(\widehat{d}_1 - d_1\right) \to^D N(0, b^{-2}), \tag{4.44}$$

[3] This feature is observed in Chapter 3 for the known mean case, for four parametric estimation methods, FML, DWH, TML and CSS. Here, we have considered an additional estimation method, exact Whittle.

where $\overline{\Lambda}_{dd}$ is the reduced form of the component defined in (3.27) for Example 4.1 given by

$$\overline{\Lambda}_{dd} = \frac{1}{n}\sum_{j=1}^{n/2}(1+\theta_0^2+2\theta_0\cos(\lambda_j))^2(2\sin(\lambda_j/2))^{-4d^*}\left(2\log(2\sin(\lambda_j/2))\right)^2,$$

and b is as in (4.43) with $\theta_0 = -0.444978$. Once again the value of $d_0 = 0.2$ is adopted for the TDGP.

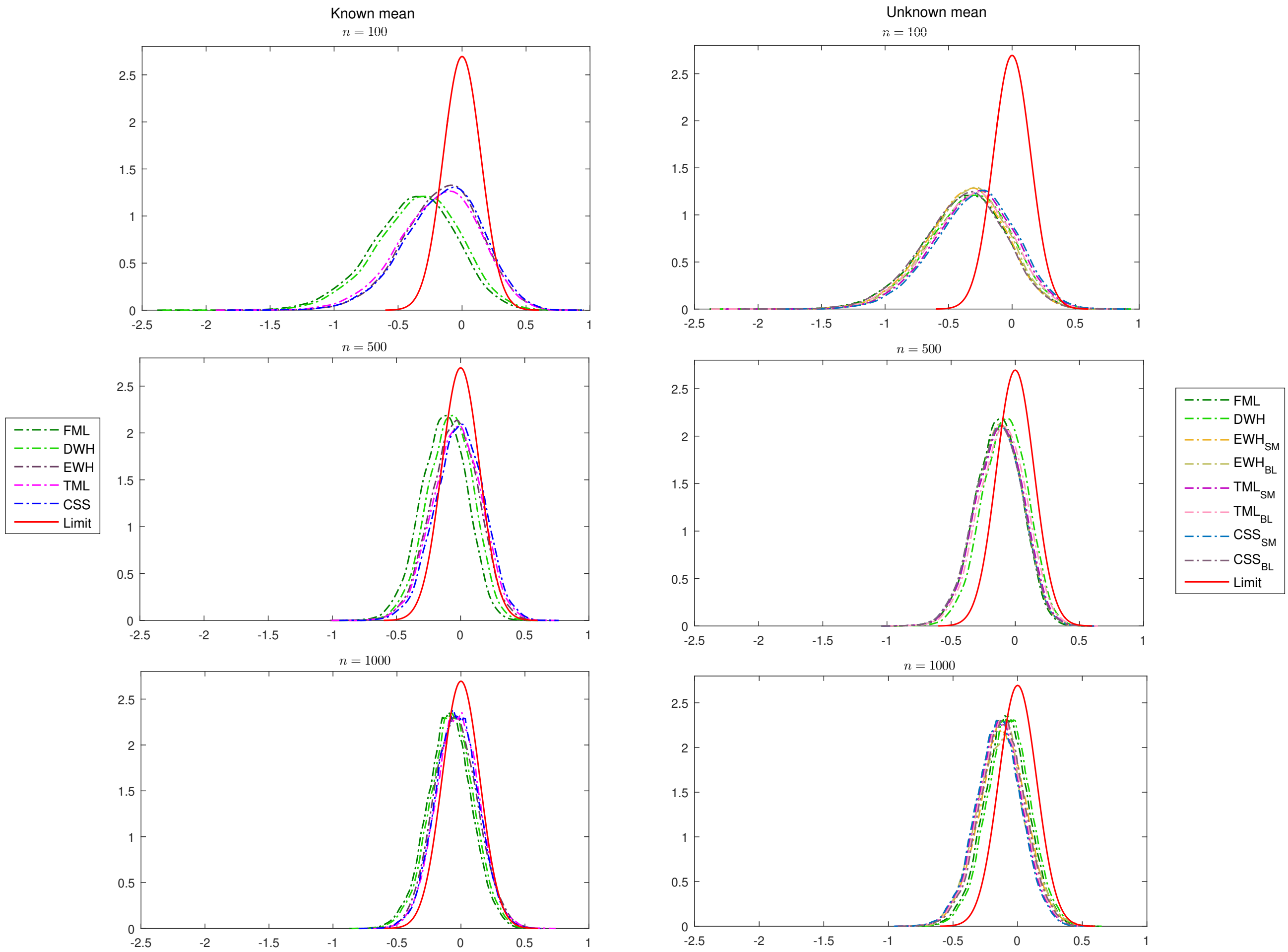

Figure 4.3: Kernel density of $n^{1/2}[\overline{\Lambda}_{dd}]^{-1/2}(\widehat{d}_1 - d_1)$ for an ARFIMA$(0, d_0, 1)$ TDGP with $d_0 = 0.2$ and $\theta_0 = -0.444978$ and an ARFIMA$(0, d, 0)$ MisM, $d^* = 0.25$.

Each panel of Figure 4.3 provides the kernel densities of $n^{1/2}\left[\overline{\Lambda}_d\right]^{-1/2}(\widehat{d}_1 - d_1)$ under the all five estimation methods, under the assumption of known and unknown mean, for a specification as labeled above each plot, plus the limiting distribution given in (4.44). The kernel densities corresponding to all five estimation methods are again positioned to the left of zero, for $n \leq 500$. We also observe once again a clear differentiation between the EWH, TML and CSS estimation methods, and the FML and DWH methods when the mean is known, with CSS being the closest to the limiting distribution for all sample sizes. This discrepancy declines as the sample sizes increases across the estimation methods. Such a differentiation no longer exists when the mean is estimated with sample mean or the BLUE, for any sample size. When the mean is estimated, DWH sits closest to the 'Limit' curve. Again, all finite sample distributions approach the limiting distribution, as the sample size increases.

Case 3: $d_0 - d_1 < 0.25$

The form of the limiting distribution here is as follows.

$$\sqrt{n}\left(\widehat{d}_1 - d_1\right) \rightarrow^D N(0, v^2), \tag{4.45}$$

where, $v^2 = \Lambda_{11}/b^2$, with

$$\begin{aligned} \Lambda_{11} &= 2\pi \int_0^{\pi} \left(\frac{f_0(\lambda)}{f_1(\boldsymbol{\eta}_1,\lambda)}\right)^2 \left(\frac{\partial \log f_1(d_1,\lambda)}{\partial d}\right)^2 d\lambda \\ &= 2\pi \int_0^{\pi} (1+\theta_0^2 + 2\theta_0 \cos(\lambda))^2 (2\sin(\lambda/2))^{-4d^*} (\log(2\sin(\lambda/2)))^2 d\lambda, \end{aligned}$$

and b as given in (4.43) evaluated at $\theta_0 = -0.3$ and $d^* = 0.1736$.

Each panel in Figure 4.4 provides the kernel density estimate of the standardized statistic $\sqrt{n}(\widehat{d}_1 - d_1)$ under the five estimation methods, in both the known and unknown mean cases. All curves correspond to the specifications labeled above in each plot, plus the limiting distribution given in (4.45). Once again, the same distinction across the estimation methods is in evidence (at least for $n < 500$) as was observed in the cases, $d^* > 0.25$ and $d^* = 0.25$, for the known mean case. However, such a distinction does not seem to exist for any sample

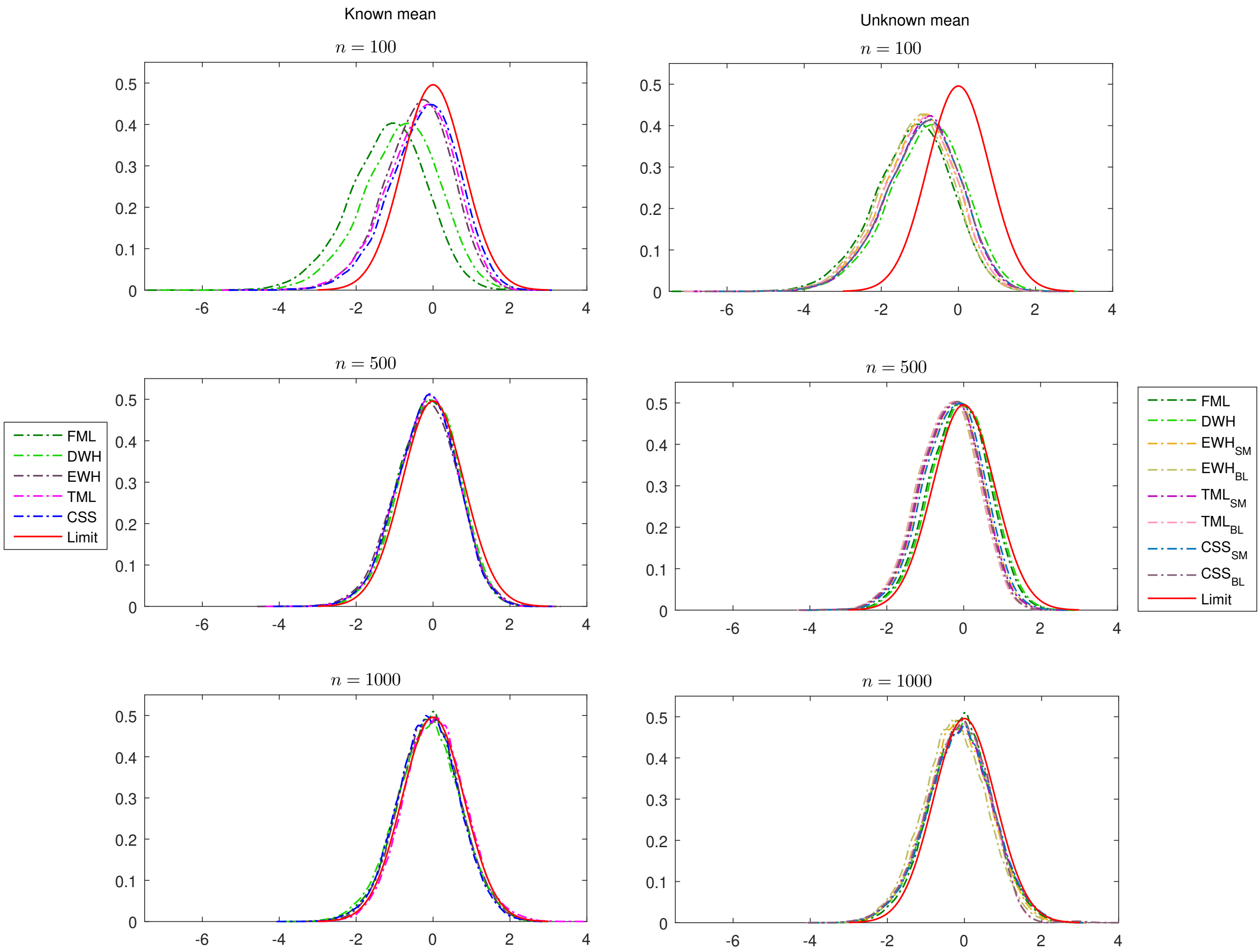

Figure 4.4: Kernel density of $\sqrt{n}(\widehat{d}_1 - d_1)$ for an ARFIMA$(0, d_0, 1)$ TDGP with $d_0 = 0.2$ and $\theta_0 = -0.3$ and an ARFIMA$(0, d, 0)$ MisM, $d^* < 0.25$.

size, for the unknown mean case. Nevertheless, the visual distance between the five estimation methods, in both the known and unknown mean cases declines faster compared to when $d^* \geq 0.25$, due to the $\sqrt{n}$ rate of convergence that obtains when $d^* < 0.25$. Also, all five estimation methods in the known mean case continue to show quicker convergence than in the unknown mean case. Nevertheless, the finite sample distributions under all estimation methods, regardless of the mean being known or unknown, match the limiting distribution very well by the time $n = 1000$. CSS again sits closer to the 'limit' curve for any sample size when the mean is known, whilst DWH is much closer when the mean is unknown, for any sample size.

4.6.2 Finite sample bias and RMSE for estimation of the pseudo-true parameter d_1: mean is known

We document the finite sample bias and RMSE of the five parametric estimators of the pseudo-true parameter d_1 assuming that the process mean is known and its value is zero. Results of bias and RMSE are produced for Example 4.1 in Table 4.1, with selected additional results relevant to the example recorded in Table 4.3. Values of $d^* = d_0 - d_1$ are documented across the key ranges, $d^* \lesseqgtr 0.25$, along with associated values for the MA coefficient in the TDGP, θ_0. The minimum values of bias and RMSE for each parameter setting are highlighted in bold face in all tables for each sample size, n.

Consider first the bias and RMSE results for Example 4.1 with $d_0 = -0.25$, as displayed in the top panel of Table 4.1. As is consistent with the theoretical results, the bias and RMSE of the five parametric estimators, FML, DWH, EWH, TML and CSS, show a clear tendency to decline as the sample size increases, for a fixed value of θ_0, thereby providing evidence that all the estimators are consistent for d_1. Further, as θ_0 declines in magnitude, and the MisM becomes closer to the TDGP, there is a tendency for the absolute values of the bias values and the RMSE to decline. The bias is *negative* for all the estimators across the parameter space of θ_0 considered here, with the (absolute) bias of the three frequency domain estimators (FML, DWH and EWH) being larger than that of the two time domain estimators. The ranking of these five estimators in finite sample performance is as follow, CSS > TML > EWH > DWH > FML, with CSS being the uniformly superior estimator of d_1 and FML being the worst, having the largest bias and RMSE. These findings are consistent with the graphical results presented in the previous section, with respect to the following features; (i) clustering of the sampling distributions of the time domain and frequency domain estimators, and, (ii) the standardized parametric estimators being to the left of zero while the frequency domain estimators tend to be further to the left of zero than those of the time domain estimators. Besides, the rate of decline in the (absolute) bias and RMSE of all estimators, as n increases, is slower for $d^* \geq 0.25$ than for $d^* < 0.25$. This is also consistent with the results given in Theorem 4.4.

Now, consider the bias and RMSE results for Example 4.1 with $d_0 = 0.2$ and 0.45, as displayed in the middle and bottom panels of Table 4.1. The performance of all five estimators under Example 4.1 with $d_0 = 0.2$ and 0.45 remains the same as that for $d_0 = -0.25$ as discussed above. As the value of d_0 associated with the TDGP increases, the bias and RMSE of all the estimators of d_1 increases, despite the estimators still being consistent for d_1. Again the CSS estimator is preferred over the other estimators under mis-specification, and the FML estimator is the worst of all.

Under Example 4.1, when the parameter θ_0 of the TDGP is zero, then the estimated model is the correctly specified (that is, in this case $d^* = 0$ and hence there is no mis-specification). The values of bias and RMSE recorded in Table 4.3 reveal that under the correct specification of the ARFIMA$(0, d, 0)$ model, the TML estimator is now superior under the known mean assumption. This result is consistent with Sowell (1992) and Nielsen and Frederiksen (2005).

Finally, we conclude that under this type of mis-specification, the CSS estimator outperforms the other four parametric estimators in estimating d_1, provided that the process mean is known. The reason for CSS being superior, rather than TML as noted under correct specification, is that when a wrong model is chosen to estimate the set of parameters involved in the model, all the four parametric estimators except the CSS utilize the complete information of mis-specified model. That is, CSS estimation is performed by minimizing the weighted infinite sum of squares of the observations, conditioning on zero values for observations beyond a particular point in time. The weights assigned to each observation reflect the estimated model. For a long run dependent model, the contribution of the weights is still significant for large n. However, since we truncate the weights for each realization (refer equations (4.29) - (4.31)) at some point, we naturally fail to incorporate the whole structure of the incorrectly specified model. Therefore less impact of mis-specification on CSS is observed.

4.6.3 Finite sample bias and RMSE for estimation of the pseudo-true parameter d_1: mean is unknown

In this Section, we document the finite sample performance of the estimators of d_1 when the process mean is unknown. Amongst the five parametric estimators, the FML and DWH estimators are mean invariant as mentioned in Remark 2.1. However, the finite sample performance of the exact Whittle and time domain estimators do alter not only because the mean is unknown, but also with the choice of estimator of the mean. We consider two possible estimators namely, the sample mean and the BLUE as defined in (4.10). The bias and RMSE results for the unknown mean case under the same settings for d^*, θ_0, d_0, and n, are produced for Example 4.1 in Table 4.4 with selected additional results relevant to Example 1 recorded in Table 4.6.

Consider first the bias and RMSE results for any type of mis-specification when the mean is estimated with sample mean. We summarize the key numerical results as follows. For the three estimators (EWH, TML and CSS) of d_1, the decline in (absolute) bias and RMSE as sample size increases, for a given θ_0, indicates that the EWH, TML and CSS estimators are con-

Table 4.1: Estimates of the bias and RMSE for the FML, Whittle, EWH, TML and CSS estimators of d_1 Example 1 - TDGP: ARFIMA$(0, d_0, 1)$ vis-a-vis Mis-M: ARFIMA$(0, d, 0)$. Process mean $\mu = 0$, is known. The estimates are obtained under Gaussian disturbances.

			FML		DWH		EWH		TML		CSS	
d^*	θ_0	n	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
						$d_0 = -0.25$						
0.3723	-0.7	100	-0.2680	0.2736	-0.2275	0.2432	-0.1866	0.2083	-0.1508	0.1829	**-0.1013**	**0.1413**
		200	-0.2157	0.2461	-0.1837	0.2084	-0.1451	0.1519	-0.1269	0.1300	**-0.0845**	**0.1010**
		500	-0.1425	0.1556	-0.1125	0.1334	-0.0865	0.1065	-0.1170	0.0991	**-0.0773**	**0.0867**
		1000	-0.1065	0.1160	-0.0865	0.1073	-0.0792	0.0887	-0.0706	0.0801	**-0.0475**	**0.0698**
0.2500	-0.44	100	-0.1535	0.1698	-0.1328	0.1592	-0.1182	0.1317	-0.0913	0.1142	**-0.0582**	**0.1030**
		200	-0.0994	0.1176	-0.0792	0.1019	-0.0686	0.0798	-0.0535	0.0662	**-0.0257**	**0.0464**
		500	-0.0530	0.0695	-0.0442	0.0540	-0.0420	0.0508	-0.0303	0.0497	**-0.0176**	**0.0301**
		1000	-0.0365	0.0488	-0.0305	0.0428	-0.0339	0.0469	-0.0281	0.0329	**-0.0119**	**0.0171**
0.1736	-0.3	100	-0.0957	0.1385	-0.0650	0.1099	-0.0541	0.0847	-0.0520	0.0766	**-0.0212**	**0.0616**
		200	-0.0605	0.0947	-0.0462	0.0784	-0.0394	0.0699	-0.0224	0.0564	**-0.0112**	**0.0474**
		500	-0.0417	0.0519	-0.0326	0.0567	-0.0250	0.0414	-0.0103	0.0348	**-0.0075**	**0.0215**
		1000	-0.0299	0.0351	-0.0199	0.0284	-0.0108	0.0215	-0.0131	0.0214	**-0.0043**	**0.0198**
						$d_0 = 0.2$						
0.3723	-0.7	100	-0.2771	0.2973	-0.2365	0.2566	-0.1774	0.2072	-0.1473	0.1822	**-0.1207**	**0.1707**
		200	-0.2020	0.2358	-0.1640	0.2031	-0.1451	0.1722	-0.1287	0.1435	**-0.1010**	**0.1207**
		500	-0.1416	0.1508	-0.1267	0.1438	-0.0980	0.1112	-0.0900	0.1015	**-0.0739**	**0.0861**
		1000	-0.1085	0.1155	-0.0859	0.1032	-0.0639	0.0853	-0.0618	0.0789	**-0.0575**	**0.0654**
0.2500	-0.44	100	-0.1489	0.1808	-0.1283	0.1661	-0.1139	0.1307	-0.0789	0.1096	**-0.0473**	**0.0967**
		200	-0.1010	0.1348	-0.0852	0.1117	-0.0637	0.0994	-0.0434	0.0072	**-0.0400**	**0.0052**
		500	-0.0578	0.0725	-0.0443	0.0531	-0.0394	0.0500	-0.0202	0.0413	**-0.0191**	**0.0408**
		1000	-0.0395	0.0506	-0.0296	0.0394	-0.0217	0.0404	-0.0134	0.0376	**-0.0133**	**0.0273**
0.1736	-0.3	100	-0.1042	0.1439	-0.0735	0.1014	-0.0547	0.0901	-0.0342	0.0826	**-0.0240**	**0.0625**
		200	-0.0663	0.0885	-0.0491	0.0664	-0.0309	0.0585	-0.0213	0.0447	**-0.0128**	**0.0345**
		500	-0.0342	0.0530	-0.0242	0.0475	-0.0216	0.0414	-0.0181	0.0317	**-0.0083**	**0.0216**
		1000	-0.0214	0.0357	-0.0114	0.0266	-0.0136	0.0232	-0.0065	0.0198	**-0.0046**	**0.0158**
						$d_0 = 0.45$						
0.3723	-0.7	100	-0.2841	0.3055	-0.2435	0.2947	-0.1979	0.2199	-0.1447	0.1620	**-0.1244**	**0.1491**
		200	-0.2096	0.2501	-0.1942	0.2340	-0.1562	0.1928	-0.1278	0.1457	**-0.1148**	**0.1340**
		500	-0.1566	0.1671	-0.1340	0.1466	-0.1197	0.1223	-0.0989	0.1135	**-0.0894**	**0.0931**
		1000	-0.1141	0.1319	-0.1088	0.1292	-0.0732	0.0955	-0.0628	0.0879	**-0.0532**	**0.0679**
0.2500	-0.44	100	-0.1519	0.1863	-0.1413	0.1758	-0.1273	0.1395	-0.0813	0.1079	**-0.0572**	**0.0963**
		200	-0.1162	0.1337	-0.0927	0.1195	-0.0739	0.0996	-0.0534	0.0703	**-0.0421**	**0.0589**
		500	-0.0628	0.0894	-0.0594	0.0696	-0.0409	0.0570	-0.0378	0.0516	**-0.0209**	**0.0460**
		1000	-0.0554	0.0683	-0.0439	0.0574	-0.0281	0.0352	-0.0227	0.0390	**-0.0182**	**0.0324**
0.1736	-0.3	100	-0.0966	0.1500	-0.0806	0.1295	-0.0429	0.1045	-0.0353	0.0835	**-0.0311**	**0.0700**
		200	-0.0773	0.0929	-0.0505	0.0871	-0.0314	0.0542	-0.0244	0.0357	**-0.0199**	**0.0148**
		500	-0.0494	0.0569	-0.0237	0.0406	-0.0269	0.0422	-0.0178	0.0312	**-0.0153**	**0.0295**
		1000	-0.0278	0.0344	-0.0215	0.0272	-0.0297	0.0381	-0.0146	0.0291	**-0.0062**	**0.0154**

Table 4.2: Estimates of the bias and RMSE for the FML, Whittle, EWH, TML and CSS estimators of d_1 Example 1 - TDGP: ARFIMA$(0, d_0, 1)$ vis-a-vis Mis-M: ARFIMA$(0, d, 0)$. Process mean $\mu = 0$, is known. The estimates are obtained under standardized chi-squared disturbances with 4 degrees of freedom.

			FML		DWH		EWH		TML		CSS	
d^*	θ_0	n	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
						$d_0 = -0.25$						
0.3723	-0.7	100	-0.2722	0.2978	-0.2218	0.2474	-0.1943	0.2281	-0.1608	0.1932	**-0.1255**	**0.1561**
		200	-0.2253	0.2677	-0.2063	0.2218	-0.1800	0.1989	-0.1441	0.1503	**-0.1219**	**0.1386**
		500	-0.1531	0.1762	-0.1231	0.1362	-0.1069	0.1242	-0.0864	0.0986	**-0.0572**	**0.0870**
		1000	-0.1069	0.1264	-0.0969	0.1064	-0.0772	0.0815	-0.0607	0.0700	**-0.0473**	**0.0602**
0.2500	-0.44	100	-0.1571	0.1831	-0.1365	0.1726	-0.1267	0.1529	-0.0920	0.1351	**-0.0602**	**0.1051**
		200	-0.1042	0.1256	-0.0994	0.1115	-0.0776	0.1010	-0.0583	0.0842	**-0.0434**	**0.0709**
		500	-0.0639	0.0846	-0.0528	0.0693	-0.0454	0.0538	-0.0398	0.0493	**-0.0271**	**0.0415**
		1000	-0.0464	0.0551	-0.0364	0.0488	-0.0215	0.0362	-0.0180	0.0226	**-0.0117**	**0.0189**
0.1736	-0.3	100	-0.1090	0.1511	-0.0983	0.1406	-0.0627	0.0954	-0.0529	0.0872	**-0.0229**	**0.0628**
		200	-0.0798	0.0972	-0.0662	0.0851	-0.0522	0.0733	-0.0421	0.0594	**-0.0126**	**0.0335**
		500	-0.0414	0.0615	-0.0314	0.0515	-0.0274	0.0459	-0.0199	0.0344	**-0.0071**	**0.0214**
		1000	-0.0247	0.0376	-0.0199	0.0350	-0.0115	0.0220	-0.0130	0.0211	**-0.0040**	**0.0108**
						$d_0 = 0.2$						
0.3723	-0.7	100	-0.2898	0.3098	-0.2791	0.2992	-0.1986	0.2277	-0.1687	0.1987	**-0.1324**	**0.1724**
		200	-0.2364	0.2675	-0.2153	0.2453	-0.1833	0.2002	-0.1436	0.1551	**-0.1119**	**0.1366**
		500	-0.1511	0.1620	-0.1411	0.1502	-0.1163	0.1345	-0.0991	0.1106	**-0.0730**	**0.0953**
		1000	-0.1186	0.1231	-0.1086	0.1155	-0.0744	0.1054	-0.0614	0.0787	**-0.0571**	**0.0753**
0.2500	-0.44	100	-0.1512	0.1924	-0.1505	0.1818	-0.1254	0.1417	-0.0904	0.1101	**-0.0788**	**0.1079**
		200	-0.1180	0.1462	-0.1010	0.1219	-0.0994	0.1158	-0.0716	0.0972	**-0.0521**	**0.0734**
		500	-0.0575	0.0720	-0.0575	0.0720	-0.0489	0.0596	-0.0297	0.0510	**-0.0195**	**0.0504**
		1000	-0.0442	0.0643	-0.0394	0.0504	-0.0224	0.0415	-0.0211	0.0375	**-0.0164**	**0.0372**
0.1736	-0.3	100	-0.1163	0.1551	-0.1056	0.1446	-0.0562	0.0910	-0.0358	0.0933	**-0.0255**	**0.0933**
		200	-0.0821	0.1082	-0.0629	0.0836	-0.0449	0.0681	-0.0225	0.0526	**-0.0123**	**0.0426**
		500	-0.0439	0.0625	-0.0338	0.0525	-0.0281	0.0411	-0.0175	0.0430	**-0.0077**	**0.0414**
		1000	-0.0244	0.0387	-0.0214	0.0355	-0.0116	0.0240	-0.0082	0.0201	**-0.0044**	**0.0298**
						$d_0 = 0.45$						
0.3723	-0.7	100	-0.2953	0.3161	-0.2746	0.2955	-0.1959	0.2259	-0.1675	0.1942	**-0.1371**	**0.1713**
		200	-0.2462	0.2700	-0.2305	0.2561	-0.1862	0.2021	-0.1406	0.1673	**-0.1162**	**0.1358**
		500	-0.1662	0.1861	-0.1362	0.1461	-0.1275	0.1389	-0.1089	0.1237	**-0.0893**	**0.0952**
		1000	-0.1142	0.1428	-0.1042	0.1119	-0.0880	0.1054	-0.0720	0.0985	**-0.0524**	**0.0684**
0.2500	-0.44	100	-0.1530	0.1868	-0.1424	0.1763	-0.1227	0.1411	-0.1065	0.1298	**-0.0826**	**0.1181**
		200	-0.1272	0.1354	-0.1150	0.1269	-0.0964	0.1105	-0.0793	0.0996	**-0.0586**	**0.0707**
		500	-0.0731	0.0907	-0.0525	0.0690	-0.0416	0.0606	-0.0373	0.0520	**-0.0204**	**0.0464**
		1000	-0.0567	0.0815	-0.0354	0.0481	-0.0290	0.0354	-0.0258	0.0316	**-0.0184**	**0.0232**
0.1736	-0.3	100	-0.1078	0.1404	-0.0973	0.1398	-0.0377	0.1146	-0.0324	0.0945	**-0.0261**	**0.0712**
		200	-0.0804	0.1066	-0.0625	0.0813	-0.0271	0.0656	-0.0203	0.0541	**-0.0163**	**0.0400**
		500	-0.0505	0.0716	-0.0291	0.0506	-0.0263	0.0625	-0.0173	0.0515	**-0.0148**	**0.0399**
		1000	-0.0223	0.0419	-0.0178	0.0342	-0.0105	0.0223	-0.0084	0.0196	**-0.0058**	**0.0161**

Table 4.3: Estimates of the bias and MSE for the FML, Whittle, EWH, TML and CSS estimators of $\widehat{d}_1$ Correct model. Process mean $\mu = 0$, is known.

		FML		DWH		EWH		TML		CSS	
d_0	n	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
						Gaussian					
-0.25	100	-0.0559	0.1053	-0.0451	0.0950	-0.0337	**0.0778**	**-0.0093**	0.0867	-0.0251	0.1050
	200	-0.0326	0.0775	-0.0253	0.0706	-0.0206	0.0624	**-0.0082**	**0.0552**	-0.0105	0.0684
	500	-0.0157	0.0588	-0.0100	0.0447	-0.0036	0.0412	**-0.0028**	**0.0342**	-0.0100	0.0388
	1000	-0.0049	0.0286	-0.0049	0.0261	-0.0016	0.0274	**-0.0014**	**0.0251**	-0.0049	0.0261
0.2	100	-0.0486	0.1069	-0.0378	0.1065	-0.0247	0.0836	**-0.0125**	**0.0771**	-0.0478	0.1065
	200	-0.0351	0.0759	-0.0226	0.0681	-0.0195	0.0562	**-0.0094**	**0.0482**	-0.0186	0.0795
	500	-0.0106	0.0437	-0.0105	0.0336	-0.0096	0.0357	**-0.0031**	**0.0348**	-0.0105	0.0388
	1000	-0.0054	0.0262	-0.0032	0.0212	0.0027	0.0249	**-0.0016**	**0.0245**	-0.0054	0.0268
0.45	100	-0.0513	0.1078	-0.0407	0.0926	-0.0274	0.0665	**-0.0140**	**0.0576**	-0.0207	0.0673
	200	-0.0345	0.0683	-0.0282	0.0558	-0.0216	0.0501	**-0.0112**	**0.0225**	-0.0349	0.0553
	500	-0.0085	0.0560	-0.0085	0.0409	-0.0135	0.0333	**-0.0068**	**0.0287**	-0.0085	0.0359
	1000	-0.0051	0.0252	-0.0031	0.0252	-0.0140	0.0297	**-0.0033**	**0.0215**	-0.0031	0.0246
						Standardized Chi-square					
-0.25	100	-0.0581	0.1065	-0.0478	0.1061	-0.0123	0.0778	-0.0106	0.0867	**-0.0047**	**0.0834**
	200	-0.0325	0.0628	-0.0238	0.0579	-0.0208	0.0486	-0.0152	0.0457	**-0.0054**	**0.0451**
	500	-0.0197	0.0585	-0.0097	0.0484	-0.0075	0.0442	-0.0025	0.0360	**-0.0002**	**0.0356**
	1000	-0.0049	0.0262	-0.0049	0.0262	-0.0022	0.0280	-0.0013	0.0251	**0.0002**	**0.0127**
0.2	100	-0.0506	0.1073	-0.0497	0.1069	-0.0338	0.0838	-0.0144	0.0774	**-0.0036**	**0.0833**
	200	-0.0389	0.0680	-0.0269	0.0592	-0.0192	0.0512	-0.0109	0.0479	**-0.0022**	**0.0378**
	500	-0.0203	0.0486	-0.0103	0.0385	-0.0086	0.0326	-0.0028	0.0317	**0.0000**	**0.0356**
	1000	-0.0052	0.0262	-0.0052	0.0231	-0.0022	0.0181	-0.0014	0.0176	**0.0003**	**0.0149**
0.45	100	-0.0662	0.1176	-0.0519	0.0971	-0.0471	0.0564	-0.0275	0.0598	**-0.0044**	**0.0553**
	200	-0.0479	0.0784	-0.0406	0.0668	-0.0376	0.0607	-0.0229	0.0534	**-0.0037**	**0.0441**
	500	-0.0183	0.0556	-0.0083	0.0456	-0.0143	0.0342	-0.0089	0.0310	**-0.0010**	**0.0278**
	1000	-0.0080	0.0330	-0.0030	0.0252	-0.0034	0.0225	-0.0024	0.0218	**-0.0005**	**0.0153**

sistent for d_1 as stated in Theorem 4.3. As the sample size increases the differences between the bias and RMSE results for the known and estimated mean cases tend to zero – evidence for the consistency of the sample mean for the true (zero) mean (refer (4.7)). However, the rate of decline of all estimators to d_1 is slower when the mean is unknown than when it is known, given the slow rate of convergence of the sample mean to the true mean (see, Hosking, 1996); and, for any given sample size, the bias and RMSE are larger compared to the known mean case (see, Table 4.4). Further, the bias tends to be more negative for all the estimators. This effect on bias and RMSE causes the EWH, TML and CSS estimators to lose their superiority over the FML and DWH estimators. When the process mean is unknown, the DWH estimator outperforms uniformly all four of the other estimators for the values of $d_0 \in (0.2, 0.45)$. On the other hand, when the TDGP has antipersistent memory, the CSS uniformly performs better than the other estimators in the unknown mean case. With reference to Table 4.6, the DWH estimator still performs better than the other estimators under the correct and over-parameterized cases even when the mean is estimated by sample mean. The dominant nature of the DWH estimator observed here over the other four estimators under correct specification of the model is consistent with the findings of Chapter 3 and Nielsen and Frederiksen (2005).

With reference to the BLUE of μ, again we observe similar characteristics for the EWH, TML and CSS estimators, with these methods delivering slightly larger bias and RMSE than the estimators of d_1 based on the sample mean estimate of μ. The reason is that BLUE incorporates the information of the mis-specified model. Here too, DWH is superior to the other estimators in many cases. However, as sample size increases the difference between the finite sample performance of all estimators declines. This is numerical evidence of the theory established in Theorems 4.3 and 4.4.

4.7 Conclusion

Consequences of mis-specification of the short memory dynamics in the class of fractionally integrated autoregressive moving average models have been explored both on a theoretical and simulation basis in Chapter 3, under the assumption that the process mean of the true data generating process is known. In this chapter, we develop new asymptotic results re-

Table 4.4: Estimates of the bias and MSE for the EWH, TML and CSS estimators of d_1 Example 1 - TDGP: ARFIMA$(0, d_0, 1)$ vis-a-vis Mis-M: ARFIMA$(0, d, 0)$. Process mean $\mu = 0$, is unknown. The estimates are obtained under Gaussian disturbances.

		Sample mean						BLUE					
		EWH		TML		CSS		EWH		TML		CSS	
θ_0	n	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
						$d_0 = -0.25$							
-0.7	100	-0.2247	0.2559	-0.1957	0.2211	-0.1174	0.1551	-0.2315	0.2872	-0.2088	0.2410	-0.1721	0.2122
	200	-0.1819	0.2154	-0.1672	0.1833	-0.1056	0.1694	-0.1926	0.2354	-0.1762	0.2255	-0.1617	0.1965
	500	-0.1525	0.1802	-0.1221	0.1335	-0.0844	0.1009	-0.1736	0.1943	-0.1535	0.1626	-0.1394	0.1450
	1000	-0.1087	0.1240	-0.0978	0.1061	-0.0709	0.0831	-0.1248	0.1454	-0.1026	0.1229	-0.0845	0.1068
-0.44	100	-0.1377	0.1678	-0.1145	0.1592	-0.0845	0.1435	-0.1664	0.1909	-0.1527	0.1850	-0.0901	0.1336
	200	-0.0986	0.1128	-0.0883	0.1010	-0.0779	0.0954	-0.1015	0.1357	-0.0975	0.1193	-0.0822	0.1068
	500	-0.0776	0.0875	-0.0529	0.0781	-0.0408	0.0692	-0.0858	0.1027	-0.0615	0.0757	-0.0375	0.0696
	1000	-0.0480	0.0542	-0.0375	0.0488	-0.0291	0.0429	-0.0576	0.0674	-0.0339	0.0495	-0.0143	0.0348
-0.3	100	-0.0766	0.1268	-0.0790	0.1216	-0.0645	0.1092	-0.1293	0.1544	-0.1016	0.1428	-0.0695	0.1159
	200	-0.0597	0.0775	-0.0500	0.0641	-0.0446	0.0525	-0.0678	0.0823	-0.0516	0.0769	-0.0486	0.0594
	500	-0.0422	0.0655	-0.0313	0.0506	-0.0254	0.0469	-0.0553	0.0780	-0.0360	0.0542	-0.0241	0.0471
	1000	-0.0259	0.0363	-0.0206	0.0308	-0.0164	0.0225	-0.0224	0.0416	-0.0113	0.0311	0.0198	0.0272
						$d_0 = 0.2$							
-0.7	100	-0.2739	0.2942	-0.2609	0.2808	-0.2521	0.2715	-0.2847	0.3054	-0.2739	0.2955	-0.2563	0.2765
	200	-0.2257	0.2599	-0.2056	0.0580	-0.1981	0.0551	-0.2394	0.2699	-0.2237	0.2525	-0.2152	0.2440
	500	-0.1573	0.1734	-0.1388	0.1684	-0.1358	0.1555	-0.1675	0.1802	-0.1405	0.1602	-0.1361	0.1557
	1000	-0.1109	0.1377	-0.1063	0.1242	-0.0950	0.1024	-0.1266	0.1453	-0.1053	0.1123	-0.0921	0.1092
-0.44	100	-0.1429	0.1751	-0.1364	0.1692	-0.1296	0.1693	-0.1548	0.1851	-0.1487	0.1724	-0.1378	0.1610
	200	-0.1313	0.1469	-0.1105	0.1379	-0.1009	0.1158	-0.1330	0.1516	-0.1254	0.1376	-0.1166	0.1275
	500	-0.0750	0.0901	-0.0555	0.0879	-0.0454	0.0706	-0.0774	0.0994	-0.0557	0.0884	-0.0455	0.0707
	1000	-0.0434	0.0543	-0.0382	0.0497	-0.0302	0.0396	-0.0579	0.0620	-0.0467	0.0572	-0.0382	0.0497
-0.3	100	-0.0935	0.1435	-0.0843	0.1374	-0.0742	0.1157	-0.1088	0.1677	-0.0949	0.1458	-0.0846	0.1362
	200	-0.0789	0.0964	-0.0603	0.0878	-0.0591	0.0775	-0.0894	0.1002	-0.0705	0.0995	-0.0612	0.0771
	500	-0.0526	0.0717	-0.0342	0.0665	-0.0255	0.0518	-0.0530	0.0769	-0.0430	0.0616	-0.0329	0.0520
	1000	-0.0348	0.0592	-0.0207	0.0352	-0.0117	0.0296	-0.0245	0.0394	-0.0219	0.0325	-0.0120	0.0248
						$d_0 = 0.45$							
-0.7	100	-0.2833	0.3141	-0.2796	0.3012	-0.2683	0.2895	-0.2951	0.3284	-0.2833	0.3158	-0.2701	0.2917
	200	-0.2331	0.2542	-0.2010	0.2469	-0.1969	0.2451	-0.2455	0.2691	-0.2364	0.2506	-0.2230	0.2469
	500	-0.1729	0.2130	-0.1635	0.1939	-0.1432	0.1735	-0.1849	0.2268	-0.1737	0.2041	-0.1534	0.1836
	1000	-0.1320	0.1587	-0.1224	0.1404	-0.1123	0.1302	-0.1463	0.1600	-0.1325	0.1504	-0.1124	0.1303
-0.44	100	-0.1735	0.2071	-0.1637	0.1875	-0.1424	0.1776	-0.1836	0.2195	-0.1641	0.1979	-0.1528	0.1777
	200	-0.1453	0.1588	-0.1372	0.1435	-0.1257	0.1303	-0.1583	0.1781	-0.1425	0.1686	-0.1352	0.1443
	500	-0.0924	0.1086	-0.0724	0.0984	-0.0620	0.0882	-0.1070	0.1242	-0.0971	0.1156	-0.0868	0.1060
	1000	-0.0621	0.0736	-0.0556	0.0679	-0.0454	0.0579	-0.0725	0.0998	-0.0633	0.0827	-0.0531	0.0731
-0.3	100	-0.1119	0.1341	-0.1045	0.1239	-0.0887	0.1129	-0.1110	0.1364	-0.0960	0.1337	-0.0821	0.1224
	200	-0.0774	0.0948	-0.0575	0.0780	-0.0548	0.0679	-0.0849	0.1052	-0.0756	0.0920	-0.0662	0.0854
	500	-0.0597	0.0684	-0.0410	0.0606	-0.0394	0.0500	-0.0554	0.0702	-0.0462	0.0666	-0.0345	0.0582
	1000	-0.0330	0.0423	-0.0294	0.0344	-0.0284	0.0321	-0.0340	0.0411	-0.0272	0.0391	-0.0221	0.0288

Table 4.5: Estimates of the bias and MSE for the EWH, TML and CSS estimators of d_1 Example 1 - TDGP: ARFIMA$(0, d_0, 1)$ vis-a-vis Mis-M: ARFIMA$(0, d, 0)$. Process mean $\mu = 0$, is unknown. The estimates are obtained under standardized chi-squared disturbances with four degrees of freedom.

		Sample mean						BLUE					
		EWH		TML		CSS		EWH		TML		CSS	
θ_0	n	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
						$d_0 = -0.25$							
-0.7	100	-0.2348	0.2674	-0.1994	0.2247	-0.1549	0.1829	-0.2440	0.2855	-0.2813	0.2933	-0.1675	0.1974
	200	-0.2006	0.2232	-0.1672	0.1890	-0.1415	0.1621	-0.2128	0.2359	-0.2037	0.2157	-0.1501	0.1682
	500	-0.1622	0.1842	-0.1224	0.1338	-0.0857	0.1029	-0.1738	0.1994	-0.1529	0.1618	-0.0901	0.1159
	1000	-0.1153	0.1360	-0.0981	0.1063	-0.0718	0.0844	-0.1289	0.1486	-0.1030	0.1122	-0.0736	0.0963
-0.44	100	-0.1405	-0.1613	-0.1175	0.1517	-0.0886	0.1276	-0.1705	0.1964	-0.1553	0.1869	-0.0937	0.1372
	200	-0.1124	0.1385	-0.1053	0.1283	-0.0643	0.1034	-0.1267	0.1458	-0.1173	0.1394	-0.0781	0.0942
	500	-0.0800	0.0922	-0.0526	0.0677	-0.0405	0.0591	-0.0936	0.1157	-0.0610	0.0749	-0.0473	0.0694
	1000	-0.0428	0.0561	-0.0374	0.0486	-0.0289	0.0429	-0.0628	0.0742	-0.0539	0.0693	-0.0446	0.0546
-0.3	100	-0.0785	0.1314	-0.0815	0.1234	-0.0678	0.1119	-0.0997	0.1445	-0.1041	0.1223	-0.0727	0.1185
	200	-0.0604	0.0812	-0.0557	0.0752	-0.0496	0.0558	-0.0696	0.0859	-0.0600	0.0781	-0.0514	0.0669
	500	-0.0349	0.0506	-0.0309	0.0500	-0.0249	0.0465	-0.0408	0.0762	-0.0356	0.0536	-0.0270	0.0488
	1000	-0.0270	0.0388	-0.0204	0.0346	-0.0163	0.0325	-0.0263	0.0456	0.0112	0.0310	-0.0199	0.0371
						$d_0 = 0.2$							
-0.7	100	-0.2765	0.3030	-0.2629	0.2825	-0.2539	0.2730	-0.2875	0.3132	-0.2759	0.2971	-0.2684	0.2983
	200	-0.2031	0.2243	-0.1864	0.2076	-0.1532	0.1859	-0.2143	0.2358	-0.2054	0.2290	-0.1692	0.1995
	500	-0.1644	0.1881	-0.1385	0.1479	-0.1355	0.1451	-0.1713	0.1990	-0.1502	0.1797	-0.1459	0.1654
	1000	-0.1253	0.1390	-0.1069	0.1142	-0.0901	0.1043	-0.1324	0.1584	-0.1153	0.1323	-0.1022	0.1292
-0.44	100	-0.1546	0.1822	-0.1383	0.1703	-0.1224	0.1604	-0.1640	0.1972	-0.1405	0.1735	-0.1397	0.1642
	200	-0.1117	0.1354	-0.0953	0.1268	-0.0782	0.0972	-0.1269	0.1426	-0.1057	0.1399	-0.0684	0.0895
	500	-0.0688	0.0994	-0.0552	0.0703	-0.0427	0.0617	-0.0816	0.1015	-0.0754	0.0945	-0.0515	0.0703
	1000	-0.0449	0.0600	-0.0381	0.0495	-0.0322	0.0399	-0.0551	0.0682	-0.0492	0.0595	-0.0381	0.0494
-0.3	100	-0.1137	0.1476	-0.0961	0.1355	-0.0813	0.1284	-0.1214	0.1563	-0.1067	0.1365	-0.0909	0.1372
	200	-0.0872	0.1054	-0.0656	0.0889	-0.0547	0.0697	-0.0939	0.1122	-0.0861	0.0956	-0.0662	0.0873
	500	-0.0645	0.0784	-0.0323	0.0513	-0.0222	0.0401	-0.0737	0.0892	-0.0327	0.0541	-0.0326	0.0511
	1000	-0.0324	0.0455	-0.0206	0.0350	-0.0187	0.0309	-0.0346	0.0412	-0.0218	0.0346	-0.0218	0.0346
						$d_0 = 0.45$							
-0.7	100	-0.2974	0.3131	-0.2603	0.2815	-0.2590	0.2801	-0.3030	0.3315	-0.2639	0.2862	-0.2609	0.2824
	200	-0.2138	0.2487	-0.1920	0.2386	-0.1495	0.1752	-0.2258	0.2567	-0.2053	0.02483	-0.1520	0.1894
	500	-0.1712	0.1989	-0.1532	0.1535	-0.1329	0.1431	-0.1942	0.2234	-0.1334	0.1437	-0.1431	0.1633
	1000	-0.1157	0.1352	-0.1025	0.1103	-0.0902	0.1031	-0.1595	0.1682	-0.1026	0.1104	-0.1025	0.1102
-0.44	100	-0.1846	0.2024	-0.1548	0.1779	-0.1334	0.1680	-0.1972	0.2213	-0.1351	0.1682	-0.1428	0.1741
	200	-0.1182	0.1305	-0.1069	0.1273	-0.0952	0.1186	-0.1267	0.1452	-0.1183	0.1374	-0.1079	0.1289
	500	-0.0774	0.0957	-0.0624	0.0780	-0.0517	0.0639	-0.0825	0.1026	-0.0568	0.0754	-0.0565	0.0758
	1000	-0.0429	0.0514	-0.0377	0.0438	-0.0354	0.0476	-0.0468	0.0643	-0.0330	0.0525	-0.0428	0.0527
-0.3	100	-0.1210	0.1378	-0.1059	0.1345	-0.0899	0.1332	-0.1157	0.1296	-0.0973	0.1341	-0.0930	0.1408
	200	-0.0639	0.0887	-0.0508	0.0795	-0.0405	0.0618	-0.0775	0.0982	-0.0669	0.0814	-0.0517	0.0719
	500	-0.0448	0.0695	-0.0408	0.0503	-0.0291	0.0497	-0.0576	0.0734	-0.0258	0.0564	-0.0342	0.0581
	1000	-0.0235	0.0402	-0.0214	0.0372	-0.0183	0.0339	-0.0348	0.0420	-0.0169	0.0389	-0.0259	0.0398

Table 4.6: Estimates of the bias and MSE for the EWH, TML and CSS estimators of $\widehat{d_1}$ Correct model. Unknown mean case

		Sample mean						BLUE					
		EWH		TML		CSS		EWH		TML		CSS	
d_0	n	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
							Gaussian						
-0.25	100	-0.0563	0.1120	-0.0422	0.0949	-0.0430	0.1057	-0.0562	0.1123	-0.0357	0.0973	-0.0387	0.1017
	200	-0.0357	0.0759	-0.0246	0.0628	-0.0385	0.0748	-0.0424	0.0775	-0.0332	0.0651	-0.0481	0.0884
	500	-0.0164	0.0572	-0.0127	0.0487	-0.0184	0.0510	-0.0150	0.0567	-0.0097	0.0429	-0.0128	0.0486
	1000	-0.0064	0.0383	-0.0056	0.0358	-0.0090	0.0379	-0.0063	0.0386	-0.0049	0.0270	-0.0060	0.0294
0.2	100	-0.0639	0.1325	-0.0404	0.1121	-0.0501	0.1209	-0.0646	0.1334	-0.0443	0.0997	-0.0516	0.1002
	200	-0.0364	0.0628	-0.0296	0.0566	-0.0320	0.0675	-0.0454	0.0658	-0.0368	0.0595	-0.0486	0.0618
	500	-0.0200	0.0584	-0.0096	0.0351	-0.0119	0.0436	-0.0241	0.0546	-0.0066	0.0398	-0.0087	0.0457
	1000	-0.0089	0.0292	-0.0049	0.0180	-0.0063	0.0280	-0.0097	0.0343	-0.0035	0.0283	-0.0047	0.0223
0.45	100	-0.0697	0.1252	-0.0455	0.1022	-0.0557	0.1136	-0.0700	0.1269	-0.0445	0.1108	-0.0552	0.1109
	200	-0.0445	0.0628	-0.0406	0.0519	-0.0359	0.0696	-0.0526	0.0746	-0.0436	0.0662	-0.0637	0.0885
	500	-0.0287	0.0658	-0.0090	0.0456	-0.0208	0.0523	-0.0264	0.0621	-0.0120	0.0450	-0.0152	0.0561
	1000	-0.0155	0.0328	-0.0039	0.0223	-0.0111	0.0272	-0.0185	0.0304	-0.0141	0.0234	-0.0098	0.0285
							Standardized Chi-square						
-0.25	100	-0.0512	0.1134	-0.0344	0.0962	-0.0326	0.0859	-0.0686	0.1173	-0.0410	0.1022	-0.0384	0.0985
	200	-0.0313	0.0685	-0.0255	0.0438	-0.0336	0.0597	-0.0492	0.0688	-0.0358	0.0509	-0.0454	0.0648
	500	-0.0153	0.0475	-0.0083	0.0377	-0.0080	0.0345	-0.0172	0.0480	-0.0095	0.0382	-0.0083	0.0316
	1000	-0.0089	0.0312	-0.0045	0.0259	-0.0042	0.0274	-0.0103	0.0355	-0.0055	0.0304	-0.0027	0.0250
0.2	100	-0.0645	0.1172	-0.0522	0.0913	-0.0437	0.1023	-0.0731	0.1249	-0.0535	0.1006	-0.0461	0.1000
	200	-0.0354	0.0615	-0.0229	0.0453	-0.0449	0.0628	-0.0427	0.0634	-0.0325	0.0442	-0.0428	0.0579
	500	-0.0216	0.0568	-0.0116	0.0381	-0.0092	0.0329	-0.0334	0.0590	-0.0286	0.0454	-0.0145	0.0375
	1000	-0.0093	0.0100	-0.0062	0.0260	-0.0048	0.0215	-0.0118	0.0353	-0.0046	0.0322	-0.0074	0.0272
0.45	100	-0.0837	0.1156	-0.0775	0.1041	-0.0468	0.0919	-0.0905	0.1235	-0.0867	0.1113	-0.0839	0.1084
	200	-0.0442	0.0628	-0.0357	0.0530	-0.0437	0.0509	-0.0418	0.0693	-0.0338	0.0445	-0.0456	0.0658
	500	-0.0304	0.0686	-0.0207	0.0370	-0.0087	0.0352	-0.0448	0.0517	-0.0249	0.0410	-0.0115	0.0346
	1000	-0.0167	0.0335	-0.0111	0.0252	-0.0038	0.0250	-0.0187	0.0412	-0.0105	0.0354	-0.0142	0.0271

lated to estimation of incorrectly specified models, extending the results of Chapter 3 to the case where the mean is also estimated. We begin by providing some useful results on the best linear unbiased estimator [BLUE] of the mean that allows us to establish the asymptotic properties of the parametric estimators of the dynamic parameters in the incorrectly specified model. The BLUE depends on the specification of the model. If the model is correctly specified, then BLUE holds the same asymptotic properties as the sample mean with a slightly different asymptotic variance. Suppose the fitted model is incorrectly specified in terms of the short memory dynamics, we show that the BLUE is still consistent for true mean. We then show that when the process mean is jointly estimated with the dynamic parameters, all five parametric estimators – the frequency domain maximum likelihood [FML] of Chen and Deo (2006), discrete version of Whittle [DWH], the exact Whittle likelihood [EWH], time domain maximum likelihood [TML], and conditional sum of squares [CSS] – converge to the same pseudo-true value under common mis-specification. Thereby, we establish the asymptotic equivalence of all five parametric estimators under common mis-specification. The theoretical results proven in the chapter are valid for short memory, long memory and antipersistent regions of the differencing parameter corresponding to the TDGP and the estimated model.

The finite sample performance of the aforementioned five parametric estimators are assessed in terms of bias, root mean squared error [RMSE] and sampling distribution, for the case where the mean is known and the case where it is unknown. The finite sampling distribution results reveal that under extreme mis-specification, and when the mean is known, two clusters are formed for smaller sample sizes; one with FML and DWH, and, the other formed by EWH, TML and CSS. This clustering is not in evidence when the mean is estimated. When the process mean is known, the finite sample distribution of CSS is closer to the limiting distribution and when the mean is estimated, that of DWH is the closest. However, as sample size increases, all finite sample distributions move closer together and towards the limiting distribution, whether the mean is estimated or not. Further, the bias and the RMSE estimates demonstrate that if the process mean is known then the CSS estimator is superior to the other four parametric estimators under the same mis-specification. When the mean is estimated with either the sample mean or BLUE, DWH is preferred over the others. Further, Monte-Carlo results enable us to recommend the sample mean as the estimate of the mean of the process for the following reasons: (i) it is very easy to implement, (ii) it is computationally not much more expensive than other estimators, (iii) in finite samples, the bias and RMSE estimates of the estimator of the fractional differencing parameter are quite similar, and, (iv) it is not affected by the distributional assumption of the error terms.

4.A Appendix: Proofs

Proof of Theorem 4.1.

Recalling $\widehat{\mu}_{BLU,0} = \mathbf{1}^\top \mathbf{\Sigma}^{-1}\mathbf{y} / \mathbf{1}^\top \mathbf{\Sigma}^{-1}\mathbf{1} = \sum_{i=1}^{n} a_i y_i$, where a_i is the i^{th} element of $\left[\mathbf{1}^\top \mathbf{\Sigma}^{-1} / \mathbf{1}^\top \mathbf{\Sigma}^{-1}\mathbf{1}\right]$, such that $\sum_{i=1}^{n} a_i = 1$ and $0 \le \sup |a_i| \le 1$.

Denote by $W_j = a_j y_j = \mu_0 + \sum_{k=-\infty}^{\infty} c_k \varepsilon_{j-k}$, where $c_k = a_k b_{0,k}$ such that $a_k = 0$ if $j < k$ otherwise, a_k takes the value as expressed above.

(i) Since $\{y_t\}$ is a stationary time series, $\sum_{k=-\infty}^{\infty} b_{0,k}^2 < \infty$ and $\sum_{k=-\infty}^{\infty} a_k^2 < \infty$, the proof follows following the corrected proof of Theorem 18.6.5 of Ibragimov and Linnik (1971) given in Hosking (1984).

(ii) This result is proven similarly using the above arguments and following the steps of the proof of Theorem 7 in Hosking (1984). Hence we omit the details.

■

Proof of Theorem 4.2. The expected value and variance of BLUE under mis-specification are as follows:

$$E\left(\widehat{\mu}_{BLU}\right)=\frac{\mathbf{1}^{\top}\mathbf{\Sigma}_{\eta}^{-1}}{\mathbf{1}^{\top}\mathbf{\Sigma}_{\eta}^{-1}\mathbf{1}}E\left(\mathbf{y}\right)=\frac{\mathbf{1}^{\top}\mathbf{\Sigma}_{\eta}^{-1}}{\mathbf{1}^{\top}\mathbf{\Sigma}_{\eta}^{-1}\mathbf{1}}\mu_{0}\mathbf{1}=\mu_{0},$$

and,

$$\begin{aligned}
Var\left(\widehat{\mu}_{BLU}\right) &= E\left[\left(\mathbf{1}^{\top}\mathbf{\Sigma}_{\eta}^{-1}\mathbf{1}\right)^{-1}\mathbf{1}^{\top}\mathbf{\Sigma}_{\eta}^{-1}\left(\mathbf{y}-\mu_{0}\mathbf{1}\right)\right]^{2} \\
&= \left(\mathbf{1}^{\top}\mathbf{\Sigma}_{\eta}^{-1}\mathbf{1}\right)^{-2}\left|\mathbf{1}^{\top}\mathbf{\Sigma}_{\eta}^{-1}\mathbf{\Sigma}_{0}\mathbf{\Sigma}_{\eta}^{-1}\mathbf{1}\right| \\
&= \left(\mathbf{1}^{\top}\mathbf{\Sigma}_{\eta}^{-1}\mathbf{1}\right)^{-1}\left\|\mathbf{\Sigma}_{\eta}^{-1/2}\mathbf{\Sigma}_{0}\mathbf{\Sigma}_{\eta}^{-1/2}\right\| \\
&= \left(\mathbf{1}^{\top}\mathbf{\Sigma}_{\eta}^{-1}\mathbf{1}\right)^{-1}\left\|\mathbf{\Sigma}_{\eta}^{-1/2}\mathbf{\Sigma}_{0}^{1/2}\right\|^{2} \\
&\leq Kn^{-1+2d+2d_{0}-2d+\delta}\leq Kn^{-1+2d_{0}+\delta},
\end{aligned} \tag{4.46}$$

following Theorem 5.2 of Adenstedt (1974) and Lemma 2 of Lieberman *et al.* (2010). Here K is a constant independent of $\boldsymbol{\eta}$ and n. ■

Proof of Lemma 4.1. The proof of the lemma is developed by following the arguments of Lemma 1 of Fox and Taqqu (1986) while extending them to the periodogram defined in (4.23) and to the entire region stationary of the fractional differencing parameter.

Note that $\tilde{I}\left(\lambda,\widehat{\mu}\right)$ has Fourier coefficients

$$\int_{-\pi}^{\pi}\exp\left(\imath\lambda k\right)\tilde{I}\left(\lambda,\widehat{\mu}\right)d\lambda=\begin{cases}W\left(k,n\right), & \left|k\right|<n\\ 0, & \left|k\right|>n\end{cases},$$

where

$$\begin{aligned}
W\left(k,n\right) &= \frac{1}{n}\sum_{t=1}^{n-k}\left(y_{t}-\widehat{\mu}\right)\left(y_{t+k}-\widehat{\mu}\right) \\
&= \frac{1}{n}\sum_{t=1}^{n-k}\left(\left(y_{t}-\mu_{0}\right)-\left(\widehat{\mu}-\mu_{0}\right)\right)\left(\left(y_{t+k}-\mu_{0}\right)-\left(\widehat{\mu}-\mu_{0}\right)\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{t=1}^{n-k}(y_t-\mu_0)(y_{t+k}-\mu_0)+\frac{n-k}{n}(\widehat{\mu}-\mu_0)^2 \\
&\quad -\frac{(\widehat{\mu}-\mu_0)}{n}\sum_{t=1}^{n-k}(y_t-\mu_0)-\frac{(\widehat{\mu}-\mu_0)}{n}\sum_{t=k+1}^{n}(y_t-\mu_0). \qquad (4.47)
\end{aligned}
$$

The sequence $\{y_t\}$ is a stationary process with $\frac{\sigma_0^2}{2\pi}f_0(\lambda)$. Following Assumption $(A.6)$, we have that $(\widehat{\mu}-\mu)\to^P 0$, and hence the last three terms of (4.47) converge to zero in probability. Therefore,

$$
\lim_{n\to\infty} W(k,n)=\frac{1}{n}\sum_{t=1}^{n-k}(y_t-\mu_0)(y_{t+k}-\mu_0)=\sigma_0^2\gamma_{0,k}(\boldsymbol{\eta}).
$$

Following the arguments of Hannan (1973) in Lemma 1, the proof can be completed. ■

Proof of Theorem 4.3. In this proof, we use the concentrated objective functions under each estimation method, with respect to σ^2.

The Exact Whittle estimator: Let us firstly consider the *known mean case*. Since

$$
\lim_{n\to\infty}\left|Q_n^{(2)}(\boldsymbol{\eta})-Q_n^{(3)}(\boldsymbol{\eta},\mu_0)\right|=0,
$$

and,

$$
\left|Q_n^{(3)}(\boldsymbol{\eta},\mu_0)-Q(\boldsymbol{\eta})\right|\le\left|Q_n^{(3)}(\boldsymbol{\eta},\mu_0)-Q_n^{(2)}(\boldsymbol{\eta})\right|+\left|Q_n^{(2)}(\boldsymbol{\eta})-Q(\boldsymbol{\eta})\right|,
$$

following the arguments in the proof of Theorem 3.1 in Chapter 3 corresponding to the Whittle estimator, we have that

$$
\lim_{n\to\infty}\left|Q_n^{(3)}(\boldsymbol{\eta},\mu_0)-Q(\boldsymbol{\eta})\right|=0,
$$

and hence $\lim_{n\to\infty}\left\|\widehat{\boldsymbol{\eta}}_1^{(3)}-\boldsymbol{\eta}_1\right\|=0.$

Now, let us consider the *unknown mean case*. Lemma 4.1 immediately gives that

$$
\left|Q_n^{(3)}(\boldsymbol{\eta},\mu_0)-Q(\boldsymbol{\eta})\right|\to^P 0.
$$

and therefore, $\lim_{n\to\infty}\left\|\widehat{\boldsymbol{\eta}}_1^{(3)}-\boldsymbol{\eta}_1\right\|=0.$

The TML estimator: Recall that the likelihood functions for the two cases when the process mean is known and unknown are denoted by $Q_n^{(4)}(\boldsymbol{\eta},\mu_0)$ and $Q_n^{(4)}(\boldsymbol{\eta},\widehat{\mu})$ respectively. We will now show that $\left|Q_n^{(4)}(\boldsymbol{\eta},\widehat{\mu}) - Q_n^{(4)}(\boldsymbol{\eta},\mu_0)\right| \to^P 0$. This distance is equivalent to considering the distance $\left|S_n - \widetilde{S}_n\right|$, where

$$\begin{aligned} S_n &= \frac{1}{n}\left(\mathbf{y}-\mu_0\mathbf{1}\right)^\top \boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\left(\mathbf{y}-\mu_0\mathbf{1}\right) \\ &= \frac{1}{n}\left[\mathbf{y}^\top\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{y} - 2\mu_0\mathbf{1}^\top\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{y} + \mu_0^2\mathbf{1}^\top\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{1}\right], \end{aligned}$$

and

$$\begin{aligned} \widetilde{S}_n &= \frac{1}{n}\left(\mathbf{y}-\widehat{\mu}\mathbf{1}\right)^\top \boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\left(\mathbf{y}-\widehat{\mu}\mathbf{1}\right) \\ &= \frac{1}{n}\left[\mathbf{y}^\top\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{y} - 2\widehat{\mu}\mathbf{1}^\top\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{y} + \widehat{\mu}^2\mathbf{1}^\top\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{1}\right]. \end{aligned}$$

Therefore,

$$\widetilde{S}_n - S_n = \frac{2}{n}\left(\widehat{\mu}-\mu_0\right)\mathbf{1}^\top\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{y} - \frac{2}{n}\left(\widehat{\mu}^2-\mu_0^2\right)\mathbf{1}^\top\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{1} \tag{4.48}$$

and hence,

$$\begin{aligned} \left|\widetilde{S}_n - S_n\right| &\le \frac{2}{n}\left|\widehat{\mu}-\mu_0\right|\left|\mathbf{1}^\top\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{y}\right| + \frac{2}{n}\left|\widehat{\mu}^2-\mu_0^2\right|\left|\mathbf{1}^\top\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{1}\right| \\ &\le \frac{2}{n}\left|\widehat{\mu}-\mu_0\right|\left|\mathbf{1}^\top\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{y}\right| + \frac{2}{n}\left|\widehat{\mu}-\mu_0\right|^2\left|\mathbf{1}^\top\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{1}\right|. \end{aligned} \tag{4.49}$$

This leads to,

$$\left|\widetilde{S}_n - S_n\right|\left|\mathbf{1}^\top\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{1}\right|^{-1} \le \frac{2}{n}\left|\widehat{\mu}-\mu_0\right|\left|\widehat{\mu}_{BLU}\right| + \frac{2}{n}\left|\widehat{\mu}-\mu_0\right|^2.$$

Assumption $(A.6)$ confirms that $\frac{1}{n}\left|\widehat{\mu}-\mu_0\right| = o_p\left(n^{-3/2+d_0}\right)$. Together with Theorem 5.2 of Adenstedt (1974), that is,

$$\sup_{\boldsymbol{\eta}\in\mathbf{E}_\delta} \mathbf{1}^\top\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{1} \le Kn^{1-2d}, \tag{4.50}$$

and Theorem 4.2, we have that

$$\left|\widetilde{S}_n - S_n\right|\left|\mathbf{1}^\top\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{1}\right|^{-1} = o_p\left(n^{-2+2d_0}\right),$$

and hence the left hand side component converges in probability to 1 uniformly in $\boldsymbol{\eta}$. Therefore,

$$\left|\widetilde{S}_n - S_n\right| = o_p\left(n^{-3+2(d_0-d_1)}\right). \tag{4.51}$$

Hence, $\lim_{n\to\infty}\left|\widetilde{S}_n - S_n\right| = 0$, uniformly in $\boldsymbol{\eta}$. Therefore, the limit of the criterion function $Q_n^{(4)}(\boldsymbol{\eta},\widehat{\mu})$ on subset $\mathbb{E}_\delta^0$ is

$$Q^{(4)}(\boldsymbol{\eta},\sigma^2) = \log\sigma^2 + \frac{2Q(\boldsymbol{\eta})}{\sigma^2},$$

uniformly in $\boldsymbol{\eta}$, by Proposition 3.1 of Chapter 3. Hence directly following the arguments in the proof of Theorem 3.1 in Chapter 3 corresponding to the TML estimator, we can conclude that $\lim_{n\to\infty}\widehat{\boldsymbol{\eta}}_1^{(4)} = \boldsymbol{\eta}_1$ and

$$\lim_{n\to\infty}\left\|\widehat{\boldsymbol{\eta}}_1^{(4)} - \widehat{\boldsymbol{\eta}}_1^{(1)}\right\| = 0.$$

The CSS estimator: The CSS objective function in (4.28) can be expressed as follows.

$$Q_n^{(5)}(\boldsymbol{\eta},\widehat{\mu}) = \frac{1}{n}\left(\mathbf{y}-\widehat{\mu}\mathbf{1}\right)^\top \mathbf{T}_\eta\mathbf{T}_\eta^\top\left(\mathbf{y}-\widehat{\mu}\mathbf{1}\right),$$

where, $\mathbf{T}_\eta$ is an $n\times n$ is an upper triangular Toeplitz matrix with non-zero elements $\gamma_{i-j}(\boldsymbol{\eta})$, $i,j=1,\ldots,n$, such that $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1} = \mathbf{T}_\eta\mathbf{T}_\eta^\top + \mathbf{H}_\eta\mathbf{H}_\eta^\top$ with $\mathbf{H}_\eta$ being the $n\times\infty$ reverse Hankel matrix with typical element $\gamma_{n-i+j}(\boldsymbol{\eta})$, $i,j=1,\ldots,\infty$.

Let $Q_n^{(5)}(\boldsymbol{\eta},\mu_0)$ be the CSS objective function when the process mean is known defined as

$$Q_n^{(5)}(\boldsymbol{\eta},\widehat{\mu}) = \frac{1}{n}\left(\mathbf{y}-\mu_0\mathbf{1}\right)^\top \mathbf{T}_\eta\mathbf{T}_\eta^\top\left(\mathbf{y}-\mu_0\mathbf{1}\right). \tag{4.52}$$

Denote by the objective function when the process mean is known

$$Q_n^{(5)}(\boldsymbol{\eta},\mu_0) = \frac{1}{n}\left(\mathbf{y}-\mu_0\mathbf{1}\right)^\top \mathbf{T}_\eta\mathbf{T}_\eta^\top\left(\mathbf{y}-\mu_0\mathbf{1}\right). \tag{4.53}$$

Now, we establish that $\left|Q_n^{(5)}(\boldsymbol{\eta},\widehat{\mu}) - Q_n^{(5)}(\boldsymbol{\eta},\mu_0)\right| \to^P 0$.

$$\left|Q_n^{(5)}(\boldsymbol{\eta},\widehat{\mu}) - Q_n^{(5)}(\boldsymbol{\eta},\mu_0)\right| = \frac{1}{n}\left|\left(\mathbf{y}-\widehat{\mu}\mathbf{1}\right)^\top \mathbf{T}_\eta\mathbf{T}_\eta^\top\left(\mathbf{y}-\widehat{\mu}\mathbf{1}\right) - \left(\mathbf{y}-\mu_0\mathbf{1}\right)^\top \mathbf{T}_\eta\mathbf{T}_\eta^\top\left(\mathbf{y}-\mu_0\mathbf{1}\right)\right|$$

$$
\begin{aligned}
&\leq \frac{2}{n}\left|\widehat{\mu}-\mu_0\right|\left|\mathbf{1}^\top \mathbf{T}_\eta \mathbf{T}_\eta^\top \mathbf{y}\right| + \frac{1}{n}\left|\widehat{\mu}-\mu_0\right|^2\left|\mathbf{1}^\top \mathbf{T}_\eta \mathbf{T}_\eta^\top \mathbf{1}\right| \\
&\leq \frac{2}{n}\left|\widehat{\mu}-\mu_0\right|\left|\mathbf{1}^\top \boldsymbol{\Sigma}_{\eta}^{-1} \mathbf{y}\right| + \frac{1}{n}\left|\widehat{\mu}-\mu_0\right|^2\left|\mathbf{1}^\top \boldsymbol{\Sigma}_{\eta}^{-1} \mathbf{1}\right|. \qquad (4.54)
\end{aligned}
$$

Using a similar approach to that used for the TML estimator, we can conclude that

$$\lim_{n\to\infty}\left|Q_n^{(5)}(\boldsymbol{\eta},\widehat{\mu}) - Q_n^{(5)}(\boldsymbol{\eta},\mu_0)\right| = 0,$$

uniformly in $\boldsymbol{\eta}$. Therefore the limit of the criterion function $Q_n^{(5)}(\boldsymbol{\eta},\widehat{\mu})$ on subsets $\mathbb{E}_\delta^0$ is $Q^{(5)}(\boldsymbol{\eta}) = 2Q(\boldsymbol{\eta})$. Hence directly following the arguments in the proof of Theorem 3.1 in Chapter 3 corresponding to the CSS estimator, we can conclude that $\lim_{n\to\infty}\widehat{\boldsymbol{\eta}}_1^{(5)} = \boldsymbol{\eta}_1$ and

$$\lim_{n\to\infty}\left\|\widehat{\boldsymbol{\eta}}_1^{(5)} - \widehat{\boldsymbol{\eta}}_1^{(1)}\right\| = 0.$$

■

Prior to proving Theorem 4.4, let us prove some important results that will be used to prove the theorem. Let us recalling the definition of $\boldsymbol{\Sigma}_\eta$ as given in (4.11). We define the following matrices associated with $\boldsymbol{\Sigma}_\eta$.

$$
\begin{aligned}
\mathbf{A}_\eta^{(1)} &= \boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Sigma}_{\partial\eta}\boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Sigma}_{\partial\eta}\boldsymbol{\Sigma}_\eta^{-1}, \qquad (4.55)\\
\mathbf{A}_\eta^{(2)} &= \boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Sigma}_{\partial^2\eta}\boldsymbol{\Sigma}_\eta^{-1}, \text{ and,} \qquad (4.56)\\
\mathbf{A}_\eta^{(3)} &= \boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Sigma}_{\partial\eta}\boldsymbol{\Sigma}_\eta^{-1}, \qquad (4.57)
\end{aligned}
$$

where

$$\boldsymbol{\Sigma}_{\partial\eta} := \int_{-\pi}^{\pi}\frac{\partial f_1(\boldsymbol{\eta},\lambda)}{\partial\boldsymbol{\eta}}e^{\iota\lambda\tau}d\lambda.$$

Lemma 4.2 *Suppose Assumptions $(A.2)-(A.4)$ hold. For every $\delta>0$ and the constant K independent of $\boldsymbol{\eta}$ and n,*

$$\left|\mathbf{1}^\top\mathbf{A}_\eta^{(3)}\boldsymbol{\Sigma}_0\mathbf{A}_\eta^{(3)}\mathbf{1}\right| \leq Kn^{1-2d_0+\delta},$$

where $\mathbf{A}_\eta^{(3)}$ is as defined in (4.57).

Proof. Expanding $\left|\mathbf{1}^\top \mathbf{A}_\eta^{(3)} \boldsymbol{\Sigma}_0 \mathbf{A}_\eta^{(3)} \mathbf{1}\right|$ using the form of $\mathbf{A}_\eta^{(3)}$ as given in (4.57), we have that

$$\begin{aligned}
\left|\mathbf{1}^\top \mathbf{A}_\eta^{(3)} \boldsymbol{\Sigma}_0 \mathbf{A}_\eta^{(3)} \mathbf{1}\right| &= \left|\mathbf{1}^\top \boldsymbol{\Sigma}_\eta^{-1} \boldsymbol{\Sigma}_{\partial\eta} \boldsymbol{\Sigma}_\eta^{-1} \boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}_\eta^{-1} \boldsymbol{\Sigma}_{\partial\eta} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{1}\right| \\
&\leq \left|\mathbf{1}^\top \boldsymbol{\Sigma}_0^{-1} \mathbf{1}\right| \left\|\boldsymbol{\Sigma}_\eta^{-1/2} \boldsymbol{\Sigma}_{\partial\eta} \boldsymbol{\Sigma}_\eta^{-1/2}\right\|^2 \\
&\leq K n^{1-2d_0+\delta},
\end{aligned}$$

using (4.50) and Lemma 5.3 of Dahlhaus (1989). ■

Proof of Theorem 4.4. Here we show that the asymptotic distribution of $\widehat{\boldsymbol{\eta}}_1^{(\cdot)}$ corresponding to the TML and CSS estimation methods is the same even when the mean is estimated. Let $\partial Q_n^{(i)}(\boldsymbol{\eta}, \widehat{\mu}) = (\partial/\partial\boldsymbol{\eta})\, Q_n^{(i)}(\boldsymbol{\eta}, \widehat{\mu})$ and $\partial^2 Q_n^{(i)}(\boldsymbol{\eta}, \widehat{\mu}) = \left(\partial^2/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top\right) Q_n^{(i)}(\boldsymbol{\eta}, \widehat{\mu})$ for $i = 3, 4, 5$. Application of the mean value theorem yields,

$$\partial Q_n^{(i)}(\widehat{\boldsymbol{\eta}}_1^{(i)}, \widehat{\mu}) - \partial Q_n^{(i)}(\boldsymbol{\eta}_1, \widehat{\mu}) = \partial^2 Q_n^{(i)}(\overline{\boldsymbol{\eta}}, \widehat{\mu}) \left(\widehat{\boldsymbol{\eta}}_1^{(i)} - \boldsymbol{\eta}_1\right), \tag{4.58}$$

with $|\overline{\boldsymbol{\eta}} - \boldsymbol{\eta}_1| \leq |\widehat{\boldsymbol{\eta}}_1^{(i)} - \boldsymbol{\eta}_1|$. Since $\widehat{\boldsymbol{\eta}}_1^{(i)}$ is the minimizer of $Q_n^{(i)}(\boldsymbol{\eta}, \widehat{\mu})$,

$$-\partial Q_n^{(i)}(\boldsymbol{\eta}_1, \widehat{\mu}) = \partial^2 Q_n^{(i)}(\overline{\boldsymbol{\eta}}, \widehat{\mu}) \left(\widehat{\boldsymbol{\eta}}_1^{(i)} - \boldsymbol{\eta}_1\right).$$

Henceforth writing $\mathbf{A} = \left[\partial^2 Q_n^{(i)}(\overline{\boldsymbol{\eta}}, \widehat{\mu})\right]$, we have that

$$R_n \left(\widehat{\boldsymbol{\eta}}_1^{(i)} - \boldsymbol{\eta}_1\right) = -\left[\partial^2 Q_n^{(i)}(\overline{\boldsymbol{\eta}}, \widehat{\mu})\right]^{-1} \left[R_n \times \partial Q_n^{(i)}(\boldsymbol{\eta}_1, \widehat{\mu})\right], \tag{4.59}$$

where R_n is the rate of convergence that takes different forms depending on $(d_0 - d_1)$ as given in Theorem 4.4.

The expression given in (4.59) can be re-arranged as follows.

$$R_n \left(\widehat{\boldsymbol{\eta}}_1^{(i)} - \boldsymbol{\eta}_1\right) = -\left\{\left(\left[\partial^2 Q_n^{(i)}(\overline{\boldsymbol{\eta}}, \widehat{\mu})\right]^{-1} - \left[\partial^2 Q_n^{(i)}(\overline{\boldsymbol{\eta}}, \mu_0)\right]^{-1}\right) + \left[\partial^2 Q_n^{(i)}(\overline{\boldsymbol{\eta}}, \mu_0)\right]^{-1}\right\}$$

$$
\begin{aligned}
& \times \left[R_n \times \partial Q_n^{(i)}(\boldsymbol{\eta}_1, \mu_0)\right] \\
= & -\left(\left[\partial^2 Q_n^{(i)}(\overline{\boldsymbol{\eta}}, \widehat{\mu})\right]^{-1} - \left[\partial^2 Q_n^{(i)}(\overline{\boldsymbol{\eta}}, \mu_0)\right]^{-1}\right) \times \left[R_n \times \partial Q_n^{(i)}(\boldsymbol{\eta}_1, \mu_0)\right] \\
& - \left[\partial^2 Q_n^{(i)}(\overline{\boldsymbol{\eta}}, \mu_0)\right]^{-1} \times \left[R_n \times \partial Q_n^{(i)}(\boldsymbol{\eta}_1, \mu_0)\right]. \qquad (4.60)
\end{aligned}
$$

We shall show that the first component disappears for large n, by proving prove the following.

(i) $R_n \left[\partial Q_n^{(i)}(\boldsymbol{\eta}_1, \widehat{\mu}) - \partial Q_n^{(i)}(\boldsymbol{\eta}_1, \mu_0)\right] \rightarrow^P 0,$

(ii) $\sup_{\boldsymbol{\eta} \in \mathbb{E}_\delta} \left|\left(\partial^2 Q_n^{(i)}(\boldsymbol{\eta}, \widehat{\mu}) - \partial^2 Q_n^{(i)}(\boldsymbol{\eta}, \mu_0)\right)\right| \rightarrow^P 0,$

(iii) $\left|\partial^2 Q_n^{(i)}(\overline{\boldsymbol{\eta}}, \mu_0) - \partial^2 Q_n^{(i)}(\boldsymbol{\eta}_1, \mu_0)\right| \rightarrow^P 0,$

The second component on the right-hand side of (4.60) has the asymptotic distribution given in Theorem 3.3 of Chapter 3 as the process mean is known. Thereby, we can show that the limiting distribution of the TML and CSS estimators of $\boldsymbol{\eta}$, when the mean is estimated, is the same limiting distribution under the known mean case. We shall prove the above three points for the TML and CSS estimation methods.

The TML estimator:

(i) As mentioned in the proof of convergence of the TML estimator, investigating the likelihood function $Q_n^{(4)}(.)$ is similar to investigating S_n. Hence we explore the behaviour of the first derivative of $\left(\widetilde{S}_n - S_n\right)$ given in (4.48):

$$
\begin{aligned}
\partial \widetilde{S}_n - \partial S_n = & \frac{2}{n}\left[\mathbf{1}^\top \boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1} \mathbf{y} \times \partial\left(\widehat{\mu} - \mu_0\right) + \left(\widehat{\mu} - \mu_0\right) \times \partial\left(\mathbf{1}^\top \boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1} \mathbf{y}\right)\right. \\
& \left. - \partial\left(\widehat{\mu}^2 - \mu_0^2\right) \times \left(\mathbf{1}^\top \boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1} \mathbf{1}\right) - \left(\widehat{\mu}^2 - \mu_0^2\right) \times \partial\left(\mathbf{1}^\top \boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1} \mathbf{1}\right)\right] \\
= & \frac{2}{n}\left[\mathbf{1}^\top \boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1} \mathbf{y} \times \partial\left(\widehat{\mu} - \mu_0\right) + \left(\widehat{\mu} - \mu_0\right) \times \mathbf{1}^\top \mathbf{A}_{\boldsymbol{\eta}}^{(3)} \mathbf{y}\right. \\
& \left. - \partial\left(\widehat{\mu}^2 - \mu_0^2\right) \times \left(\mathbf{1}^\top \boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1} \mathbf{1}\right) - \left(\widehat{\mu}^2 - \mu_0^2\right) \times \mathbf{1}^\top \mathbf{A}_{\boldsymbol{\eta}}^{(3)} \mathbf{1}\right],
\end{aligned}
$$

where $\partial\widehat{\mu} = 0$ if the mean estimator is sample mean. If it is estimated by BLUE,

$$\begin{aligned}
\partial\widehat{\mu} &= \frac{\left(\mathbf{1}^\top \boldsymbol{\Sigma}_\eta^{-1}\mathbf{1}\right) \times \left(\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\mathbf{y}\right) - \left(\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\mathbf{1}\right) \times \left(\mathbf{1}^\top \boldsymbol{\Sigma}_\eta^{-1}\mathbf{y}\right)}{\left(\mathbf{1}^\top \boldsymbol{\Sigma}_\eta^{-1}\mathbf{1}\right)^2} \\
&= \frac{1}{\left(\mathbf{1}^\top \boldsymbol{\Sigma}_\eta^{-1}\mathbf{1}\right)} \left(\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\mathbf{y}\right) - \frac{\left(\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\mathbf{1}\right)}{\left(\mathbf{1}^\top \boldsymbol{\Sigma}_\eta^{-1}\mathbf{1}\right)^2} \left(\mathbf{1}^\top \boldsymbol{\Sigma}_\eta^{-1}\mathbf{y}\right), \text{ and,} \\
E\left[\partial\widehat{\mu}\right] &= \frac{1}{\left(\mathbf{1}^\top \boldsymbol{\Sigma}_\eta^{-1}\mathbf{1}\right)} E\left(\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\mathbf{y}\right) - \frac{\left(\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\mathbf{1}\right)}{\left(\mathbf{1}^\top \boldsymbol{\Sigma}_\eta^{-1}\mathbf{1}\right)^2} E\left(\mathbf{1}^\top \boldsymbol{\Sigma}_\eta^{-1}\mathbf{y}\right) \\
&= \frac{\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\mathbf{1}}{\left(\mathbf{1}^\top \boldsymbol{\Sigma}_\eta^{-1}\mathbf{1}\right)} \mu_0 - \frac{\left(\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\mathbf{1}\right)}{\left(\mathbf{1}^\top \boldsymbol{\Sigma}_\eta^{-1}\mathbf{1}\right)} \mu_0 = 0. \qquad (4.61)
\end{aligned}$$

This leads to,

$$R_n \left|\partial\widetilde{S}_n - \partial S_n\right| \leq \frac{2R_n}{n} \left|\widehat{\mu} - \mu_0\right| \left|\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\mathbf{y}\right| + \frac{2R_n}{n} \left(\widehat{\mu} - \mu_0\right)^2 \left|\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\mathbf{1}\right|. \qquad (4.62)$$

Lemma 5.4(d) of Dahlhaus (1989) and Jensen's inequality imply

$$\sup_{\eta \in \mathbb{E}_\delta^0} \left|\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\mathbf{1}\right| \leq K n^{1-2d_1+\delta}, \qquad (4.63)$$

and

$$\begin{aligned}
E\left|\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\mathbf{y}\right| &\leq E\left|\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\left(\mathbf{y} - \mu_0\mathbf{1}\right)\right| + \mu_0 \left|\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\mathbf{1}\right|^{1/2} \\
&\leq \left|\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\boldsymbol{\Sigma}_0 \mathbf{A}_\eta^{(3)}\mathbf{1}\right|^{1/2} + \mu_0 \left|\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\mathbf{1}\right|^{1/2} \\
&\leq K\left|\mathbf{a}^\top \mathbf{a}\right|^{1/2-d_0} n^\delta + \mu_0 K n^{1/2-d_1+\delta} \text{ for some } \delta > 0 \\
&\leq K n^{1/2-\max(d_0,d_1)+\delta}, \qquad (4.64)
\end{aligned}$$

following Lemma 5.4(c and d) of Dahlhaus. Therefore, the first and second components in the right-hand side of (4.62) reduce to,

$$\begin{aligned}
\frac{R_n}{n}\left|\widehat{\mu} - \mu_0\right| \left|\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\mathbf{y}\right| &\leq \frac{R_n}{n}\left|\widehat{\mu} - \mu_0\right| \left|\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\mathbf{y}\right| \\
&\leq K R_n n^{-1+d_0-\max(d_0,d_1)+\delta},
\end{aligned}$$

and

$$\frac{R_n}{n}\left(\widehat{\mu}-\mu_0\right)^2\left|\mathbf{1}^\top\mathbf{A}_\eta^{(3)}\mathbf{1}\right|\leq KR_nn^{-1+2(d_0-d_1)+\delta},$$

using Assumption $(A.6)$, (4.63) and (4.64). This simplifies (4.62) as,

$$R_n\left|\partial\widetilde{S}_n-\partial S_n\right|\leq KR_nn^{-1+2(d_0-d_1)+\delta}.$$

Since R_n is the rate of convergence that is slower than $\sqrt{n}$, we therefore obtain (i).

(ii) The second derivative of $\left(\widetilde{S}_n-S_n\right)$ is

$$\begin{aligned}
\partial^2\widetilde{S}_n-\partial^2 S_n &= \frac{2}{n}\Big[\mathbf{1}^\top\mathbf{\Sigma}_\eta^{-1}\mathbf{y}\times\partial^2\left(\widehat{\mu}-\mu_0\right)+\mathbf{1}^\top\mathbf{A}_\eta^{(3)}\mathbf{y}\times\partial\left(\widehat{\mu}-\mu_0\right)\\
&\quad+\left(\widehat{\mu}-\mu_0\right)\times\mathbf{1}^\top\mathbf{A}_\eta^{(1)}\mathbf{y}+\partial\left(\widehat{\mu}-\mu_0\right)\times\mathbf{1}^\top\mathbf{A}_\eta^{(3)}\mathbf{y}\\
&\quad-\partial\left(\widehat{\mu}^2-\mu_0^2\right)\times\left(\mathbf{1}^\top\mathbf{A}_\eta^{(3)}\mathbf{1}\right)-\partial^2\left(\widehat{\mu}^2-\mu_0^2\right)\times\left(\mathbf{1}^\top\mathbf{\Sigma}_\eta^{-1}\mathbf{1}\right)\\
&\quad-\left(\widehat{\mu}^2-\mu_0^2\right)\times\mathbf{1}^\top\mathbf{A}_\eta^{(1)}\mathbf{1}-\partial\left(\widehat{\mu}^2-\mu_0^2\right)\times\mathbf{1}^\top\mathbf{A}_\eta^{(3)}\mathbf{1}\Big]\\
&= \frac{2}{n}\Big[\mathbf{1}^\top\mathbf{\Sigma}_\eta^{-1}\mathbf{y}\times\partial^2\left(\widehat{\mu}-\mu_0\right)+2\mathbf{1}^\top\mathbf{A}_\eta^{(3)}\mathbf{y}\times\partial\left(\widehat{\mu}-\mu_0\right)\\
&\quad+\left(\widehat{\mu}-\mu_0\right)\times\mathbf{1}^\top\mathbf{A}_\eta^{(1)}\mathbf{y}-2\partial\left(\widehat{\mu}^2-\mu_0^2\right)\times\left(\mathbf{1}^\top\mathbf{A}_\eta^{(3)}\mathbf{1}\right)\\
&\quad-\partial^2\left(\widehat{\mu}^2-\mu_0^2\right)\times\left(\mathbf{1}^\top\mathbf{\Sigma}_\eta^{-1}\mathbf{1}\right)-\left(\widehat{\mu}^2-\mu_0^2\right)\times\mathbf{1}^\top\mathbf{A}_\eta^{(1)}\mathbf{1}\Big].
\end{aligned}$$

Henceforth writing $\sup_{\mathbb{E}_\delta^0}=\sup_\eta$:

$$\begin{aligned}
\sup_\eta\left|\partial^2\widetilde{S}_n-\partial^2 S_n\right| &\leq \frac{2}{n}\sup_\eta\left|\widehat{\mu}-\mu_0\right|\sup_\eta\left|\mathbf{1}^\top\mathbf{A}_\eta^{(1)}\mathbf{y}\right|\\
&\quad+\frac{4}{n}\sup_\eta\left|\mathbf{1}^\top\mathbf{A}_\eta^{(3)}\mathbf{y}\right|\sup_\eta\left|\partial\left(\widehat{\mu}-\mu_0\right)\right|\\
&\quad+four\ other\ terms. \qquad (4.65)
\end{aligned}$$

Suppose the mean is estimated by the sample mean, $\partial^2\widehat{\mu}=0$, and if it is estimated by BLUE,

$$\partial^2\widehat{\mu} = \frac{\left(\mathbf{1}^\top\mathbf{A}_\eta^{(1)}\mathbf{y}\right)}{\left(\mathbf{1}^\top\mathbf{\Sigma}_\eta^{-1}\mathbf{1}\right)}-\frac{\left(\mathbf{1}^\top\mathbf{A}_\eta^{(3)}\mathbf{y}\right)^2}{\left(\mathbf{1}^\top\mathbf{\Sigma}_\eta^{-1}\mathbf{1}\right)^2}-\frac{\left(\mathbf{1}^\top\mathbf{A}_\eta^{(3)}\mathbf{1}\right)}{\left(\mathbf{1}^\top\mathbf{\Sigma}_\eta^{-1}\mathbf{1}\right)^2}\left(\mathbf{1}^\top\mathbf{A}_\eta^{(3)}\mathbf{y}\right)$$

$$-\frac{\left(\mathbf{1}^\top \mathbf{A}_\eta^{(1)}\mathbf{1}\right)}{\left(\mathbf{1}^\top \mathbf{\Sigma}_\eta^{-1}\mathbf{1}\right)^2}\left(\mathbf{1}^\top \mathbf{\Sigma}_\eta^{-1}\mathbf{y}\right)+2\frac{\left(\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\mathbf{1}\right)^2}{\left(\mathbf{1}^\top \mathbf{\Sigma}_\eta^{-1}\mathbf{1}\right)^3}\left(\mathbf{1}^\top \mathbf{\Sigma}_\eta^{-1}\mathbf{y}\right),$$

with

$$\begin{aligned} E\left|\partial^2\widehat{\mu}\right| \;\leq\;& \frac{\left|\mathbf{1}^\top \mathbf{A}_\eta^{(1)}\mathbf{y}\right|}{\left|\mathbf{1}^\top \mathbf{\Sigma}_\eta^{-1}\mathbf{1}\right|}+\frac{E\left(\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\mathbf{y}\right)^2}{\left(\mathbf{1}^\top \mathbf{\Sigma}_\eta^{-1}\mathbf{1}\right)^2}+\frac{\left|\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\mathbf{1}\right|}{\left(\mathbf{1}^\top \mathbf{\Sigma}_\eta^{-1}\mathbf{1}\right)^2}E\left|\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\mathbf{y}\right| \\ &+\frac{\left|\mathbf{1}^\top \mathbf{A}_\eta^{(1)}\mathbf{1}\right|}{\left(\mathbf{1}^\top \mathbf{\Sigma}_\eta^{-1}\mathbf{1}\right)^2}E\left|\mathbf{1}^\top \mathbf{\Sigma}_\eta^{-1}\mathbf{y}\right|+2\frac{\left(\mathbf{1}^\top \mathbf{A}_\eta^{(3)}\mathbf{1}\right)^2}{\left(\mathbf{1}^\top \mathbf{\Sigma}_\eta^{-1}\mathbf{1}\right)^3}E\left|\mathbf{1}^\top \mathbf{\Sigma}_\eta^{-1}\mathbf{y}\right|. \end{aligned} \tag{4.66}$$

Application of Lemma 5.4 of Dahlhaus (1989) gives

$$\begin{aligned} \left|\mathbf{1}^\top \mathbf{A}_\eta^{(1)}\left(\mathbf{y}-\mu_0\mathbf{1}\right)\right| \;\leq\;& \left(\sum_{j,k=1}^{p+q+1}\left|\left(\mathbf{1}^\top \mathbf{A}_\eta^{(1)}\mathbf{1}\right)_{jj}\right|\left|\left(\mathbf{y}-\mu_0\mathbf{1}\right)^\top \mathbf{A}_\eta^{(1)}\left(\mathbf{y}-\mu_0\mathbf{1}\right)_{kk}\right|\right)^{1/2} \\ \leq\;& Kn^{1/2-(d_0-d_1)}\left(\sum_{t=1}^{n}\left(y_t-\mu_0\right)^2\right)^{1/2} \\ \leq\;& Kn^{1-(d_0-d_1)}\left(\gamma_0\left(\boldsymbol{\eta}\right)\right)^{1/2}. \end{aligned}$$

The other terms in (4.66) can be treated similarly.

Then considering the first term of $\sup_\eta\left|\partial^2\widetilde{S}_n-\partial^2 S_n\right|$ given in (4.65).

$$\frac{2}{n}\sup_\eta\left|\widehat{\mu}-\mu_0\right|\sup_\eta\left|\mathbf{1}^\top \mathbf{A}_\eta^{(1)}\mathbf{y}\right|\leq n^{-1/2+d_1+\delta},$$

and the other five terms can be deduced in a similar way. Hence we prove (ii).

(iii) Immediately from Theorem 4.3, we have that $\widehat{\boldsymbol{\eta}}_1^{(i)} \rightarrow^{a.s} \boldsymbol{\eta}_1$ and that when the mean is replaced by its true mean it is established that for $i=3,4$ and 5,

$$\left|\partial^2 Q_n^{(i)}\left(\overline{\boldsymbol{\eta}},\mu_0\right)-\partial^2 Q_n^{(i)}\left(\boldsymbol{\eta}_1,\mu_0\right)\right|\rightarrow^P 0,$$

following Theorem 3.3 of Chapter 3.

Hence, we have that the limiting distribution of the TML estimator when the mean is also estimated along with the dynamic parameters under mis-specification, is the same as that is obtained for the known mean case. Following similar steps to those given in the proof of Theorem 3.3 of Chapter 3 immediately gives that the TML and DWH methods are asymptotically equivalent.

The CSS estimator: Using the steps used for TML estimator, the limiting distribution of the CSS estimation method when the mean is estimated is the same as that under the known mean case. Hence we omit the details here. Further,

$$\begin{aligned}\left|Q_n^{(5)}(\boldsymbol{\eta}_1,\widehat{\mu}) - Q_n^{(1)}(\boldsymbol{\eta}_1)\right| &\leq \left|Q_n^{(5)}(\boldsymbol{\eta}_1,\widehat{\mu}) - Q_n^{(5)}(\boldsymbol{\eta}_1,\mu_0)\right| + \left|Q_n^{(5)}(\boldsymbol{\eta}_1,\mu_0) - Q_n^{(1)}(\boldsymbol{\eta}_1)\right| \\ &\leq O\left(n^{-1}\right) + o_p\left(n^{-1/2}\right),\end{aligned}$$

following the expression in (4.54). Therefore, the FML and CSS estimation methods are asymptotically equivalent.

The Exact Whittle estimator: Since $\left|Q_n^{(2)}(\boldsymbol{\eta}) - Q_n^{(3)}(\sigma^2,\boldsymbol{\eta})\right|$ converges to zero almost surely when $\boldsymbol{\eta} \in \mathbb{E}_\delta$, the proof of Theorem 3.3 of Chapter 3 immediately gives the asymptotic equivalence of the DWH and EWH estimators. ∎

Chapter 5

Optimal bias correction of the log-periodogram regression estimator: A jackknife approach

5.1 Introduction

Data on many climate, hydrological, economic and financial variables exhibit dynamic patterns characterized by a long lasting response to past shocks. Notable examples include, water levels in rivers (Hurst, 1951), rainfall (Gil-Alana, 2012), aggregate output (Diebold and Rudebusch, 1989), inflation (Hassler and Wolters, 1995), interest rates (Baillie, 1996), exchange rates (Cheung, 2016) and stock market volatility (Bollerslev and Mikkelsen, 1996; Andersen *et al.*, 2003). Such 'long memory processes' are characterized by non-summable autocovariances that decline at a (slow) hyperbolic rate, in contrast to the usual exponential, and summable, decay associated with a short memory process; the fractionally integrated autoregressive moving average [ARFIMA] model of Adenstedt (1974), Granger and Joyeux (1980) and Hosking (1981) being a popular representation. Equivalently, a stationary (potentially) long memory process, $\{Y_t\}, t \in \mathbb{Z} := \{0, \pm 1, \pm 2, \ldots\}$ can be represented by the spectral density,

$$f_{YY}(\lambda) = (2\sin(\lambda/2))^{-2d} g_{YY}(\lambda), \ -\pi < \lambda < \pi, \tag{5.1}$$

where the fractional differencing parameter d satisfies $d \in (-0.5, 0.5)$, and $g_{YY}(\cdot)$ is an even function that is continuous on $(-\pi, \pi)$, is bounded above and bounded away from zero, and satisfies $\int_{-\pi}^{\pi} \log g_{YY}(\lambda)\, d\lambda = 0$. The process is said to have *long* memory when $d \in (0, 0.5)$, *intermediate* memory when $d \in (-0.5, 0)$ and *short* memory when $d = 0$. The factor $g_{YY}(\cdot)$ controls the (remaining) short memory behaviour associated with the process. For detailed expositions of processes described by (5.1), including applications, see, Beran (1994), Doukhan *et al.* (2003) and Robinson (2004).

In estimating the parameter d, the semi-parametric log-periodogram regression [LPR] estimator of Geweke and Porter-Hudak (1983) and (Robinson, 1995a,b) has been widely used, due to the simplicity of its construction as an ordinary least squares [OLS] estimator, and its avoidance of a potentially incorrect specification for the short memory component. Consistency is only achieved, however, at the cost of both a slower rate of convergence than the usual parametric rate and substantial finite sample bias in the presence of ignored short run dynamics (see, for example, Agiakloglou *et al.*, 1993 and Nielsen and Frederiksen, 2005).

Given this well-documented bias, *bias reduction* of the LPR estimator has been a focus of the literature. Andrews and Guggenberger (2003), for example, include additional frequencies, to degree $2r$ for $r \geq 0$, in the log-periodogram regression that defines the LPR estimator, producing an estimator (denoted hereafter by $\widehat{d}_r^{AG}$) whose bias goes to zero at a faster rate than that of the unadjusted procedure (recovered by setting $r = 0$), when $r > 1$. Alternative analytical procedures appear in Moulines and Soulier (1999), Hurvich and Brodsky (2001) and Robinson and Henry (2003), whilst a method based on the pre-filtered sieve bootstrap has been introduced by Poskitt *et al.* (2016). Critically, all such bias-correction methods come at a cost: namely, an increase in asymptotic variance. Notably, Guggenberger and Sun (2006) produce a weighted average of LPR estimators over different bandwidths that achieves the same degree of bias reduction as $\widehat{d}_r^{AG}$ for any given r, but with less variance inflation. This estimator, along with that of Poskitt *et al.* (2016), serve as important comparators for the alternative bias-corrected estimator that we develop herein.

The approach to bias adjustment adopted in this chapter applies the jackknife principle, with the bias-corrected estimator constructed as a weighted average of LPR estimators computed, in turn, from the full sample and m sub-samples of a given length. The sub-samples may be created by using either the non-overlapping or the moving-block method. Motivated by the jackknife technique proposed by Chen and Yu (2015) in a unit root setting, weights are chosen to remove bias up to a given order and, at the same time, to minimize the increase in asymptotic variance. The weights are 'optimal' in this sense and the associated jackknife estimator referred to as 'optimal' accordingly. In the fractional setting, with the LPR estimator being the method to be adjusted, these optimal weights involve two types of covariance terms: (i) covariances between the full-sample and sub-sample log-periodogram ordinates (to be defined in (5.15)), and, (ii) covariances between distinct sub-sample log-periodogram values (to be defined in (5.16)). These covariance terms may, in turn, be represented by cumulants of the discrete Fourier transform [DFT] of the time series. Building on results in Brillinger (1981, Chapters 2 and 4), we firstly derive closed-form expressions for the association between the corresponding discrete Fourier transforms [DFTs] in terms of cumulants. Under mild conditions on the regularity of $g_{YY}(.)$ in (5.1), we prove that the periodograms (at a given ordinate or at different ordinates) associated with the full sample and the sub-samples are asymptotically independent $\chi^2_{(2)}$ random variables. We then obtain closed-form expressions for the covariances in (5.15) and (5.16), that are required to evaluate the optimal weights.

Under regularity, we prove the consistency and asymptotic normality of the optimal jackknife estimator. Most notably, we establish that the convergence rate and asymptotic variance are equivalent to those of the unadjusted LPR estimator. That is, there is *no* inflation in asymptotic efficiency compared to the *unadjusted* LPR estimator of d, despite the bias reduction that is achieved. This compares with Guggenberger and Sun (2006), in which the goal is to produce an estimator (for a given value of r) with an asymptotic variance that is smaller than that of the corresponding bias-adjusted estimator of Andrews and Guggenberger (2003), as based on the same value of r, $\widehat{d}_r^{AG}$. In particular, in the case where $r = 0$, and no bias adjustment is achieved (with $\widehat{d}_r^{AG}$ equivalent to the raw LPR estimator), the estimator of Guggenberger and Sun is still biased, but with a (possibly) reduced asymptotic variance. In addition, in contrast with Guggenberger and Sun, and the other analytical bias adjustment methods cited above, our theoretical results do not rely on the assumption of Gaussianity. Specifically, expressions for the dominant bias term and variance of the LPR estimator - needed in the construction of the jackknife estimator and as originally derived by Hurvich *et al.* (1998) for fractional *Gaussian* processes - are shown to hold under non-Gaussian assumptions. Hence, all theoretical results for the bias-adjusted estimator hold under similar generality.[1]

The remainder of the chapter is organized as follows. In Section 5.2, we introduce two log-periodogram regression estimators; namely, the LPR estimator originally proposed by Geweke and Porter-Hudak (1983) and the particular bias-reduced estimator of Guggenberger and Sun (2006). In Section 5.3, we develop the new jackknife estimator that accommodates both bias correction and variance minimization via the appropriate choice of weights. All theoretical results pertaining to the construction of the afore-mentioned covariance terms, and

[1]We refer the reader to Hahn and Newey (2013), Chen and Yu (2015), Robinson and Kaufmann (2015) and Chambers (2013) for other applications of the jackknife in time series settings. To our knowledge the technique has been used only once in a long memory setting *per se*, namely in the numerical work of Ekonomi and Butka (2011), where the method of Chambers (2013) is adopted for the purpose of reducing the bias of the LPR estimator to the first order. However, no rigorous proofs of the properties of the estimator are provided, and no attempt at yielding an optimal estimator in the sense given in the current chapter, is made.

the resultant asymptotic properties of the optimal estimator, are given in Section 5.4. Section 5.5 documents the finite sample performance of the estimator by means of a Monte Carlo study. The simulation results show that versions of the optimally bias-corrected jackknife estimator outperform the alternative bias-adjusted estimators of Guggenberger and Sun and Poskitt *et al.* (2016), in terms of bias-reduction and root mean squared error [RMSE], with the RMSE being somewhat close to, or even smaller than, that of the LPR in some cases. This qualitative result holds under both Gaussian and Student t errors and for both autoregressive and moving average structures for the short run dynamics. In the empirically realistic case where the true values of the parameters - required in order to evaluate all relevant covariances - are unknown, we implement the jackknife estimator using an iterative procedure. This feasible version of the estimator does not consistently outperform either the bootstrap-based estimator of Poskitt *et al.* or (a feasible version of) the method of Guggenberger and Sun, but is not substantially inferior, in terms of either bias or RMSE, and is sometimes still the least biased estimator of all.

The proofs of all results are contained in Appendix 5.A, while Appendix 5.B provides various technical results, including the evaluation of the covariances required for the construction of the weights for the optimal jackknife estimator. The following notation is used throughout: "$\rightarrow^{P}$" denotes convergence in probability, "$\rightarrow^{D}$" denotes convergence in distribution, and "$\rightarrow$" is used to indicate the limit as $n \rightarrow \infty$, (unless otherwise stated). The k^{th}-order spectral density function of the time series $\{X_t\}$ is denoted by $f_{X \ldots X}(\lambda_1, \lambda_2, \ldots, , \lambda_{k-1})$, where $\lambda_1, \lambda_2, \ldots, , \lambda_{k-1}$ are fundamental frequencies. For instance, the density function given in (5.1) is the second-order spectral density of $\{Y_t\}$.[2]

[2]As noted in Chapter 1, this chapter has been written as a draft for a self-contained article for journal submission. Hence, there is a certain amount of repetition of material presented in other chapters. There are also some

5.2 Log-periodogram regression estimation methods

In this section we briefly review two log-periodogram regression estimators; namely, the raw (unadjusted) LPR estimator and the bias-reduced weighted-average estimator of Guggenberger and Sun (2006) [GS]. These estimators are used as benchmarks for later comparisons, and the raw LPR estimator, of course, underpins the jackknife method developed in Section 5.3. We summarize the asymptotic properties of these estimators and the assumptions underlying those properties. In contrast to earlier proofs related to the LPR estimator (e.g. Hurvich *et al.*, 1998) we do not assume that the data generating process [DGP] is Gaussian. This extension to non-Gaussian processes means that the properties subsequently derived for the optimal jackknife estimator are also applicable for this general case.

5.2.1 The log-periodogram regression estimator

Let $\mathbf{y}^\top = (y_1, y_2, ..., y_n)$ be a sample of n observations from a process with a spectral density as given in (5.1). The LPR estimator, $\widehat{d}_n$, is motivated by the following simple linear regression model that is formed directly from the spectral density given in (5.1),

$$\log I_Y^{(n)}(\lambda_j) = (\log g_{YY}(0) - C) - 2d\log(2\sin(\lambda_j/2)) + \xi_j, \tag{5.2}$$

where

$$I_Y^{(n)}(\lambda) = |D_Y^{(n)}(\lambda)|^2; \quad D_Y^{(n)}(\lambda) = \frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n} y_t \exp(-\imath\lambda t), \tag{5.3}$$

and $D_Y^{(n)}(\lambda_j)$ is the DFT of the vector of realizations, $\mathbf{y}$, measured at Fourier frequencies, $\lambda_j = 2\pi j/n$; $(j = 1, 2..., N_n)$, $N_n = \lfloor n^\alpha \rfloor$, for $0 < \alpha < 1$, and $\imath = \sqrt{-1}$ is the imaginary unit. Here, the error terms $\xi_j = \log\left(I_Y^{(n)}(\lambda_j)/f_{YY}(\lambda_j)\right) + C + V_j$, $j = 1, 2, ..., N_n$, where

$$V_j = \log\left(g_{YY}(\lambda_j)/g_{YY}(0)\right), \tag{5.4}$$

small differences in notation from the other chapters, due to the notational requirements of the material in this chapter.

are assumed to be asymptotically independently and identically distributed (*i.i.d.*) and C is the Euler constant. The LPR estimator of d is simply the OLS estimator of the slope parameter in (5.2) and is given by

$$\widehat{d}_n = \frac{-0.5 \sum_{j=1}^{N_n} \left(x_j - \overline{x}\right) z_j}{\sum_{j=1}^{N_n} (x_j - \overline{x})^2}, \tag{5.5}$$

where $z_j = \log I_Y^{(n)}(\lambda_j)$, $x_j = \log(2\sin(\lambda_j/2))$, and $\overline{x} = \frac{1}{N_n}\sum_{j=1}^{N_n} x_j$. The subscript n is introduced here in order to distinguish this full-sample version of the estimator from that computed subsequently from sub-samples, in the process of applying the jackknife.

Certain statistical properties of the LPR estimator such as its bias, variance, mean-squared-error [MSE] and asymptotic distribution have been derived by Hurvich *et al.* (1998) under given regularity conditions, and with certain approximations invoked. Alternative expressions for the bias and variance of the LPR estimator are provided in Theorem 1 of Andrews and Guggenberger (2003), plus in Theorem 3.1 of Guggenberger and Sun (2006), by setting $r = 0$. Lieberman (2001) also provides a formula for the expectation of the LPR estimator under the same conditions as Hurvich *et al.*; however, his expression is an infinite sum of a quantity that depends on the true values of d and the short memory parameters, which renders a feasible version of the jackknife technique using his expression more cumbersome.

With all results cited above derived under the assumption of Gaussianity, we now extend the results stated in Theorems 1 and 2 of Hurvich *et al.* (1998) to the general (potentially non-Gaussian) case. In particular, the resultant expression for the expectation of the LPR estimator is used in the specification of the optimal jackknife estimator, and in the proof of its properties.

We begin with the following assumptions on the DGP:

$(A.1)$ There exists $G > 0$, such that

$$f_{YY}(\lambda) = G\lambda^{-2d} + O(\lambda^{2-2d}) \text{ as } \lambda \to 0+,$$

where '$\to 0+$' denotes an approach from above.

$(A.2)$ In a neighbourhood $(0, \varepsilon)$ of the origin, $f_{YY}(\lambda)$ is differentiable and

$$\left| \frac{d}{d\lambda} \log f_{YY}(\lambda) \right| = O(\lambda^{-1}), \text{ as } \lambda \to 0+.$$

In addition, $g'_{YY}(0) = 0$, $|g''_{YY}(\lambda)| < \widetilde{B}_2 < \infty$ and $|g'''_{YY}(\lambda)| < \widetilde{B}_3 < \infty$, where $g'_{YY}(\lambda)$, $g''_{YY}(\lambda)$ and $g'''_{YY}(\lambda)$ denote, respectively, the first-, second- and third-order derivatives of g_{YY} with respect to λ.

$(A.3)$ $\{Y_t\}$, $t \in \mathbb{Z} := \{0, \pm 1, \pm 2, \cdots\}$ satisfies

$$Y_t - \mu_Y = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} b_j^2 < \infty, \quad \left| \frac{d}{d\lambda} b(\lambda) \right| = O(\lambda^{-1}), \quad \text{as } \lambda \to 0+,$$

where $b(\lambda) = \sum_{j=0}^{\infty} b_j \exp(\imath j \lambda)$ and $\{\varepsilon_t\}$ is a strictly stationary *i.i.d.* process with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = 1$.

$(A.4)$ The innovation process $\{\varepsilon_t\}$ satisfies the conditions in $(A.3)$. In addition, $E(\varepsilon_t)^3 < \infty$ and $E(\varepsilon_t)^4 < \infty$ are assumed.

Assumptions $(A.1) - (A.3)$ are standard in the long memory literature and are satisfied by the class of ARFIMA models. The boundedness of the first three derivatives of g_{YY} in Assumption $(A.2)$ is required to control the fourth-order moment of the sine and cosine components of the standardized DFTs that are used to derive the bias term of the LPR. Assumption $(A.4)$ specifies the third and fourth moments of $\{\varepsilon_t\}$ to be finite, as we do not invoke Gaussianity.

The boundedness imposed on the higher-order moments of $\{\varepsilon_t\}$ ensures the asymptotic normality of the DFTs associated with the process $\{Y_t\}$. The asymptotic normality of the DFTs is, in turn, used in proving Theorems 5.1 – 5.5.

We now state Theorem 5.1, which gives the mean, variance and asymptotic distribution of the LPR estimator. We subsequently exploit these results to construct the optimal jackknife estimator, and to prove its properties, in Section 5.3.

Theorem 5.1 *Let Assumptions $(A.1)-(A.3)$ hold. Given $N_n \to \infty$, $n \to \infty$, with $\frac{N_n \log N_n}{n} \to 0$,*

$$E(\widehat{d}_n) = d_0 - \frac{2\pi^2}{9}\frac{g''_{YY}(0)}{g_{YY}(0)}\frac{N_n^2}{n^2} + o\left(\frac{N_n^2}{n^2}\right) + O\left(\frac{\log^3 N_n}{N_n}\right), \tag{5.6}$$

$$Var(\widehat{d}_n) = \frac{\pi^2}{24 N_n} + o\left(\frac{1}{N_n}\right) \tag{5.7}$$

and $\widehat{d}_n \to^P d_0$, where d_0 is the true value of d. Given that $(A.4)$ also holds and if $N_n = o\left(n^{4/5}\right)$ and $\log^2 n = o\left(N_n\right)$, then,

$$\sqrt{N_n}(\widehat{d}_n - d_0) \to^D N\left(0, \tfrac{\pi^2}{24}\right). \tag{5.8}$$

5.2.2 The weighted-average log-periodogram regression estimator

The motivation for the estimator of Guggenberger and Sun (2006) stems from the work of Andrews and Guggenberger (2003). With (5.4) being the term that causes the dominant bias in the LPR estimator, Andrews and Guggenberger use a Taylor series expansion around $j = 0$ to approximate (5.4) as an even polynomial in the frequencies of order r.[3] Including the first $2r$ terms (with $r \geq 1$) in the log-periodogram regression in (5.2) as additional regressors leads to

$$\ln I_Y^{(n)}(\lambda_j) = (\log g_{YY}(0) - C) - 2d\log(2\sin(\lambda_j/2)) + \sum_{k=1}^{r}\frac{b_{2k}}{(2k)!}\lambda_j^{2k} + \zeta_j, \tag{5.9}$$

[3]The odd-order terms of the Taylor's expansion around zero are exactly zero. This leads to the expansion with only even-order terms.

where $\zeta_j = \xi_j - \sum_{k=1}^{r} \frac{b_{2k}}{(2k)!}\lambda_j^{2k}$. Application of OLS to (5.9) then yields an estimator of d, $\widehat{d}_r^{AG}$, with reduced bias relative to the raw LPR estimator, $\widehat{d}_n$. The bias-adjusted estimator is shown to be $\sqrt{N_n}$- consistent, with an asymptotic variance equal to $\frac{\pi^2}{24}c_r$, with $c_r > 1$ for $r \geq 1$ and $c_r = 1$ for $r = 0$.

Guggenberger and Sun (2006) proceed to show that an appropriate weighted average of raw LPR estimators, as based on different bandwidths, $N_{n,i} = \lfloor q_i N_n \rfloor$; $i = 1, \ldots, K$, for fixed numbers q_i chosen suitably, has the same asymptotic bias as $\widehat{d}_r^{AG}$ (constructed using N_n), but with a reduced asymptotic variance. That is, bias reduction is achieved at a smaller cost than is the original method of Andrews and Guggenberger (2003). Further, for the case of $r = 0$, the bias of the raw LPR estimator is retained but with reduced asymptotic variance. The authors also demonstrate that the weighted-average estimator, denoted by $\widehat{d}_r^{GS}$ hereafter, can be implemented via a simple two-step procedure. In the first step, a series of K LPR estimates are obtained using the regression model in (5.2) and for bandwidths, $N_{n,i}, i = 1, \ldots, K$. Then, in the second step, the following pseudo-regression is estimated, using the K estimates produced in the first step as observations of the dependent variable in the regression,

$$\widehat{d}_{N_{n,i}} = d + \sum_{j=1}^{r} \beta_{2j} q_i^{2j} + \beta_{2+2r}\left(q_i^{2+2r} - \delta \sum_{p=1}^{K} q_p^{2+2r}\right) + u_i, \; i = 1, \ldots, K, \tag{5.10}$$

where u_i is the error term, and $\mathbf{u}^\top = (u_1, u_2, ..., u_K)$ has a zero (vector) mean and asymptotic variance-covariance matrix,

$$\mathbf{\Omega} = \left(\Omega_{i,j}\right) \in \mathbb{R}^{K \times K}, \text{ with } \Omega_{i,j} = \frac{1}{\max\left(q_i, q_j\right)}.$$

The tuning parameter δ on the right-hand-side of (5.10) is a fixed non-zero constant that is used to control the multiplicative constant of the dominant bias term and render that term equivalent to the dominant bias term of $\widehat{d}_r^{AG}$. The estimator, $\widehat{d}_r^{GS}$, is then defined as the first component of the GLS estimator of $\left(d, \boldsymbol{\beta}^\top\right)^\top$, where $\boldsymbol{\beta}^\top = (\beta_2, \beta_4, \ldots, \beta_{2+2r})$, that is,

$$\left(\widehat{d}_r^{GS}, \widehat{\boldsymbol{\beta}}^\top\right)^\top = \left(\mathbf{Z}^\top \boldsymbol{\Omega}^{-1} \mathbf{Z}\right)^{-1} \mathbf{Z}^\top \boldsymbol{\Omega}^{-1} \widehat{\mathbf{d}}, \tag{5.11}$$

where $\widehat{\mathbf{d}}$ is the $(K \times 1)$ dimensional vector with i^{th} element $\widehat{d}_{N_{n,i}}$, and

$$\mathbf{Z}^\top = (\mathbf{z}_1, \ldots, \mathbf{z}_K) \in \mathbb{R}^{(2+r)\times K}, \text{ with } \mathbf{z}_i^\top = \left(1, q_i^2, \ldots, q_i^{2r}, \left(q_i^{2+2r} - \delta \sum_{p=1}^{K} q_p^{2+2r}\right)\right).$$

Both the raw LPR estimator, $\widehat{d}_n$, and the weighted-average estimator, $\widehat{d}_r^{GS}$, with $r = 1$, are used as comparators of our proposed jackknife procedure in the Monte Carlo simulation exercises in Section 5.5.

5.3 The optimal jackknife log-periodogram regression estimator

5.3.1 Definition of the jackknife estimator

The idea behind jackknifing is to generate a set of sub-samples, by deleting one or more observations of the original sample, while preserving the structure of dependence within the sub-samples; the aim being to use (weighted) sub-sample estimates to produce a bias-corrected estimator of the parameter of interest. Let $\mathbf{y}_i$ $(i = 1, 2, ..., m)$ denote a set of m sub-samples of $\mathbf{y}$, each of which has equal length, l, such that $n = l \times m$. If sub-samples are chosen using the 'non-overlapping' method, then $\mathbf{y}_i^\top = \left(y_{(i-1)l+1}, \ldots, y_{il}\right)$ for $i = 1, \ldots, m$; alternatively if the sub-sampling scheme is 'moving-block' then $\mathbf{y}_i^\top = (y_i, \ldots, y_{i+l-1})$ for all i. In the current context we use the jackknife technique to bias correct the LPR estimator. Hence, we need to produce the full-sample estimator, $\widehat{d}_n$, and the LPR estimators produced by applying OLS to the model in (5.2), using the relevant sub-sample. We denote these m sub-sample estimators (based on either the non-overlapping or moving-block method) by $\widehat{d}_i$, $i = 1, 2, ..., m$. We summarize notation corresponding to the full-sample estimation and both forms of sub-sample estimation in Table 5.1, for ease of subsequent referencing.

Table 5.1: Quantities related to the full sample and the sub-samples used in the construction of the jackknife estimator

		Full sample	i^{th} sub-sample
(i)	Frequency	$\lambda_j = 2\pi j/n$	$\mu_j = 2\pi j/l = 2\pi jm/n = m\lambda_j$
(ii)	Frequency range	$j = 1, ..., N_n$	$j = 1, ..., N_l$
(iii)	Spectral density	$f_{YY}(\lambda) = (2\sin(\lambda/2))^{-2d} g_{YY}(\lambda)$	$f_{Y_iY_i}(\mu) = (2\sin(\mu/2))^{-2d} g_{Y_iY_i}(\mu)$
(iv)	DFT	$D_Y^{(n)}(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} y_t \exp(-\imath\lambda t)$	$D_{Y_i}^{(l)}(\mu) = \frac{1}{\sqrt{2\pi l}} \sum_{t=1}^{l} y_{t+i'} \exp(-\imath\mu t)$
(v)	Periodogram	$I_Y^{(n)}(\lambda) = \lvert D_Y^{(n)}(\lambda)\rvert^2$	$I_{Y_i}^{(l)}(\mu) = \lvert D_{Y_i}^{(l)}(\mu)\rvert^2$
(vi)	Error term	$\xi_j = \log\left(I_Y^{(n)}(\lambda_j) / f_{YY}(\lambda_j)\right)$	$\xi_j^{(i)} = \log\left(I_{Y_i}^{(l)}\left(\mu_j\right) / f_{Y_iY_i}\left(\mu_j\right)\right)$
	Other notation:		
(vii)		$x_j = \ln(2\sin(\lambda_j/2))$	$x_j' = \ln(2\sin\left(\mu_j/2\right))$
$(viii)$		$\overline{x} = \sum_{t=1}^{N_n} x_j \Big/ N_n$	$\overline{x'} = \sum_{t=1}^{N_l} x_j' \Big/ N_l$
(ix)		$a_j = x_j - \overline{x}$	$a_j' = x_j' - \overline{x'}$
(x)		$S_{xx} = \sum_{j=1}^{N_n} a_j^2$	$S_{xx}' = \sum_{j=1}^{N_l} a_j'^2$

Note, regarding the sub-sample notation in point (iv), if the sub-samples are drawn with the non-overlapping scheme then, $i' = (i-1)l$. If the moving-block scheme is used then, $i' = i - 1$.

Define the jackknife estimator, $\widehat{d}_{J,m}$, as

$$\widehat{d}_{J,m} = w_n \widehat{d}_n - \sum_{i=1}^{m} w_i \widehat{d}_i, \tag{5.12}$$

where w_n and $\{w_i\}_{i=1}^{m}$ are the weights assigned to the full-sample estimator and the sub-sample estimators, respectively. Re-iterating, $\widehat{d}_n$ is the LPR estimator obtained from the full sample (as defined directly in (5.5)) and $\widehat{d}_i$ $(i = 1, 2, ..., m)$ denotes the i^{th} sub-sample LPR estimator. Under the conditions of Theorem 5.1, it is straightforward to show that

$$\begin{aligned} E(\widehat{d}_{J,m}) &= (w_n - \sum_{i=1}^{m} w_i) d_0 - \Big(\frac{2\pi^2}{9} \frac{g_{YY}''(0)}{g_{YY}(0)} \frac{N_n^2}{n^2} w_n - \frac{2\pi^2}{9} \frac{g_{Y_iY_i}''(0)}{g_{Y_iY_i}(0)} \frac{N_l^2}{l^2} \sum_{i=1}^{m} w_i\Big) \\ &\quad + o\Big(\frac{N_n^2}{n^2}\Big) + O\Big(\frac{\log^3 N_n}{N_n}\Big), \end{aligned} \tag{5.13}$$

and

$$Var(\widehat{d}_{J,m}) = \frac{\pi^2}{24N_n} w_n^2 + \frac{\pi^2}{24N_l} \sum_{i=1}^{m} w_i^2 + 2\sum_{i=1}^{m-1} \sum_{j=i+1}^{m} w_i w_j Cov(\widehat{d}_i, \widehat{d}_j)$$

$$-2w_n \sum_{i=1}^{m} w_i Cov(\widehat{d}_n, \widehat{d}_i) + o\Big(\frac{1}{N_n}\Big). \tag{5.14}$$

The covariance between the full-sample LPR estimator and each sub-sample LPR estimator, $Cov\left(\widehat{d}_n, \widehat{d}_i\right)$, and the covariances between the different sub-sample LPR estimators, $Cov\left(\widehat{d}_i, \widehat{d}_j\right)$, for $i \neq j,\, i, j = 1, 2, ..., m$, are given respectively by,

$$\begin{aligned} Cov(\widehat{d}_n, \widehat{d}_i) &= \frac{1}{4S_{xx}} \frac{1}{S'_{xx}} \sum_{j=1}^{N_n} \sum_{k=1}^{N_l} a_j a_k^{(i)} Cov\big(\log I_Y^{(n)}\,(\lambda_j)\,, \log I_{Y_i}^{(l)}\,(\mu_k)\,\big), & (5.15)\\ Cov(\widehat{d}_i, \widehat{d}_{i'}) &= \frac{1}{4} \frac{1}{(S'_{xx})^2} \sum_{j=1}^{N_l} \sum_{k=1}^{N_l} a'_j a'_k Cov\big(\log I_{Y_i}^{(l)}\,\left(\mu_j\right)\,, \log I_{Y_{i'}}^{(l)}\,(\mu_k)\,\big), & (5.16) \end{aligned}$$

with all notation as defined in Table 5.1.

Our aim is to obtain the set of weights, $\{w_n, w_1, \ldots, w_m\}$, such that $\widehat{d}_{J,m}$ has the following properties:

$(P.1)$ $\widehat{d}_{J,m}$ is an asymptotically unbiased estimator of d_0, with bias reduced to an order of $o(N_n^2/\,n^2)$, and,

$(P.2)$ $\widehat{d}_{J,m}$ achieves minimum variance among all such bias-reduced estimators.

The ‘optimal’ jackknife estimator so defined is derived via the Lagrangian method in the following section. In Section 4, the asymptotic properties of the covariances in (5.15) and (5.16) that determine the asymptotic behaviour of the estimator are derived, and the asymptotic efficiency of the estimator then proven.

5.3.2 Derivation of the optimal estimator

The minimization problem is formulated as follows. Produce weights, $\{w_n, w_1, \ldots, w_m\}$, that satisfy:

$$\min_{w_n, \{w_i\}_{i=1}^m} Var(\widehat{d}_{J,m}), \tag{5.17}$$

subject to two constraints

$$g^1(w_n, w_1, \ldots, w_m) = w_n - \sum_{i=1}^{m} w_i - 1 = 0, \tag{5.18}$$

$$g^2(w_n, w_1, ..., w_m) = \frac{N_n^2}{n^2} w_n - m^2 \frac{N_l^2}{l^2} \sum_{i=1}^{m} w_i = 0. \tag{5.19}$$

We refer to the optimal estimator so produced as $\widehat{d}_{J,m}^{Opt}$ hereinafter.

Constraints (5.18) and (5.19) ensure that Property $(P.1)$ holds for the resultant estimator. Specifically, (5.18) ensures that $\widehat{d}_{J,m}^{Opt}$ is asymptotically unbiased for d_0, as can be seen by inspection of (5.13). The dominant bias term of $\widehat{d}_{J,m}^{Opt}$ will be eliminated if and only if the second component appearing in (5.13) is set to zero; that is, if and only if

$$\frac{2\pi^2}{9} \frac{g_{YY}''(0)}{g_{YY}(0)} \frac{N_n^2}{n^2} w_n - \frac{2\pi^2}{9} \frac{g_{Y_iY_i}''(0)}{g_{Y_iY_i}(0)} \frac{N_l^2}{l^2} \sum_{i=1}^{m} w_i = 0. \tag{5.20}$$

Using Point (iii) of Table 5.1, we have that $g_{Y_iY_i}(0) = g_{YY}(0)$ and $g_{Y_iY_i}''(0) = m^2 g_{YY}''(0)$. Hence, the condition in (5.20) collapses to constraint (5.19). Given (5.17), Property $(P.2)$ is satisfied by construction.

Henceforth writing, $Cov(\widehat{d}_n, \widehat{d}_i) = c_{n,i}^*$ and $Cov(\widehat{d}_i, \widehat{d}_{i'}) = c_{i,j}^\dagger$, such that $c_{i,j}^\dagger = c_{j,i}^\dagger$, the Lagrangian function is given by,

$$\begin{aligned}\tilde{L}(w_n, w_1, \ldots, w_m, \delta_1, \delta_2) = {} & \frac{\pi^2}{24N_n} w_n^2 + \frac{\pi^2}{24N_l} \sum_{i=1}^{m} w_i^2 + 2\sum_{i=1}^{m-1} \sum_{j=i+1}^{m} w_i w_j c_{i,j}^\dagger \\ & -2w_n \sum_{i=1}^{m} w_i c_{n,i}^* + \delta_1 \big(w_n - \sum_{i=1}^{m} w_i - 1\big) \\ & +\delta_2 \Big(\frac{N_n^2}{n^2} w_n - m^2 \frac{N_l^2}{l^2} \sum_{i=1}^{m} w_i\Big).\end{aligned} \tag{5.21}$$

The first-order conditions [FOCs] are thus given by,

$$\frac{\partial \tilde{L}}{\partial \delta_1} = 0 \Rightarrow w_n - \sum_{i=1}^{m} w_i = 1,$$

$$\begin{aligned}
\frac{\partial \tilde{L}}{\partial \delta_2} &= 0 \Rightarrow \frac{N_n^2}{n^2} w_n - m^2 \frac{N_l^2}{l^2} \sum_{i=1}^{m} w_i = 0, \\
\frac{\partial \tilde{L}}{\partial w_n} &= 0 \Rightarrow \frac{2\pi^2}{24 N_n} w_n - 2 \sum_{i=1}^{m} w_i c_{n,i}^* + \delta_1 + \frac{N_n^2}{n^2} \delta_2 = 0, \\
\frac{\partial \tilde{L}}{\partial w_{i,m}} &= 0 \Rightarrow -2 w_n c_{n,i}^* + \frac{2\pi^2}{24 N_l} w_i + 2 \sum_{j=1, j \neq i}^{m} w_j c_{i,j}^{\dagger} - \delta_1 - m^2 \frac{N_l^2}{l^2} \delta_2 = 0;\ i = 1, \ldots, m.
\end{aligned}$$

Defining

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & \ldots & -1 & 0 & 0 \\ \frac{N_n^2}{n^2} & -m^2 \frac{N_l^2}{l^2} & \ldots & -m^2 \frac{N_l^2}{l^2} & 0 & 0 \\ \frac{\pi^2}{12 N_n} & -2c_{n,1}^* & \ldots & -2c_{n,m}^* & 1 & \frac{N_n^2}{n^2} \\ -2c_{n,1}^* & \frac{\pi^2}{12 N_l} & \ldots & 2c_{1,m}^{\dagger} & -1 & -m^2 \frac{N_l^2}{l^2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ -2c_{n,m}^* & 2c_{1,m}^{\dagger} & \ldots & \frac{\pi^2}{12 N_l} & -1 & -m^2 \frac{N_l^2}{l^2} \end{bmatrix}, \ \mathbf{w} = \begin{bmatrix} w_n \\ w_1 \\ \vdots \\ w_m \\ \delta_1 \\ \delta_2 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \tag{5.22}$$

the optimal solution, $\mathbf{w}^* = \begin{bmatrix} w_n^* & w_1^* & \ldots & w_m^* & \delta_1^* & \delta_2^* \end{bmatrix}^\top$, is given by

$$\mathbf{w}^* = \mathbf{A}^{-1} \mathbf{b}. \tag{5.23}$$

Given the structure of $\mathbf{b}$ this means that the solutions for the weights are given by the elements of the first column of $\mathbf{A}^{-1}$, and the optimal jackknife estimator is accordingly given as:

$$\widehat{d}_{J,m}^{Opt} = w_n^* \widehat{d}_n - \sum_{i=1}^{m} w_i^* \widehat{d}_i, \tag{5.24}$$

where $w_n^* = \left[1 - \left(N_n l / \left(N_l m n\right)\right)^2\right]^{-1}$, given immediately by solving the first two FOCs.

To complete the result we need to show that (5.23) is a local minimizer of $\tilde{L}(\cdot)$. To do so, we need to show that: (i) the constraint qualification – that the rank of the matrix formed by the first-order derivatives at the solution of the constraints with respect to parameters, except the Lagrangian parameters, is equal to the number of conditions – is met, (ii) the solution of the Lagrangian function satisfies the FOCs, and, (iii) the leading principal minors of the bordered Hessian matrix, $\mathbf{H}^B_{(m+3)\times(m+3)}$, all take the same sign of $(-1)^k$, where k is the number of constraints (see, Chapter 12 of Chiang and Wainwright, 2005, for more details).

In our problem, the number of constraints equals 2 and

$$Rank\begin{bmatrix} \frac{\partial g^1}{\partial w_n} & \frac{\partial g^2}{\partial w_n} \\ \frac{\partial g^1}{\partial w_1} & \frac{\partial g^2}{\partial w_1} \\ \vdots & \vdots \\ \frac{\partial g^1}{\partial w_m} & \frac{\partial g^2}{\partial w_m} \end{bmatrix} = Rank\begin{bmatrix} 1 & 1 \\ \frac{N_n^2}{n^2} & m^2\frac{N_l^2}{l^2} \\ \vdots & \vdots \\ \frac{N_n^2}{n^2} & m^2\frac{N_l^2}{l^2} \end{bmatrix} = 2.$$

Hence, the rank condition is met. The second condition is met by default. The important condition is the third one, where we need to show that the leading principal minors of $\mathbf{H}^B_{(m+3)\times(m+3)}$, exceed zero for every $m = 2, 3, \ldots$. The bordered Hessian matrix for our case is given by

$$\mathbf{H}^B_{(m+3)\times(m+3)} = \begin{bmatrix} 0 & 0 & 1 & -1 & \cdots & -1 \\ 0 & 0 & \frac{N_n^2}{n^2} & -m^2\frac{N_l^2}{l^2} & \cdots & -m^2\frac{N_l^2}{l^2} \\ 1 & \frac{N_n^2}{n^2} & \frac{\pi^2}{12N_n} & -2c^*_{n,1} & \cdots & -2c^*_{n,m} \\ -1 & -m^2\frac{N_l^2}{l^2} & -2c^*_{n,1} & \frac{\pi^2}{12N_l} & \cdots & 2c^\dagger_{1,m} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -m^2\frac{N_l^2}{l^2} & -2c^*_{n,m} & 2c^\dagger_{1,m} & \cdots & \frac{\pi^2}{12N_l} \end{bmatrix}.$$

The proof of positivity of the principal minors of the above matrix is given in Appendix 5.B. Hence, the solution in (5.23) is a local minimizer of $\tilde{L}(\cdot)$.

We complete this section with three remarks:

Remark 5.1 *If we consider only bias reduction to the order* N_n^2/n^2, *without concurrent variance reduction; that is, we produce an estimator that satisfies only (P.1), and not (P.2), then the formulae for the weights are*

$$w_n^* = \left[1 - \left(\frac{N_n}{N_l}\frac{l}{nm}\right)^2\right]^{-1} \text{ and } w_i^* = \frac{1}{m}\left(w_n^* - 1\right), \text{ for } i = 1, \ldots, m. \tag{5.25}$$

These weights mimic those of Chambers (2013) in the short memory setting (under a non-overlapping sub-sampling scheme), in which variance minimization was not a consideration.

Remark 5.2 *When Chambers (2013) considers the moving-block sub-sampling scheme (again, in the short memory setting), he chooses the sub-sample length to be* $l = n - m + 1$. *In this case, when* n

is large and m is small, the sub-sample length is $l \approx n$, and the impact of bias correction is reduced as a consequence; something that is in evidence in the Monte Carlo simulation results reported by that author. As a result of this observation, in our investigations we use the common sub-sample length of $l = n/m$, under both the non-overlapping and moving-block schemes.

Remark 5.3 *Condition* 3.3 *of Guggenberger and Sun (2006) has a similar purpose to our (5.19). The difference is that we eliminate the $O\left(N_n^2/n^2\right)$ term from the bias of the LPR estimator, whereas they eliminate bias up to an order of N_n^{2r}/n^{2r}, for some $r \geq 1$. The role played by (5.17) is somewhat different from that played by Condition* 3.4 *of Guggenberger and Sun (2006). The latter condition is imposed mainly to link the bias and variance of $\widehat{d}_r^{GS}$ to that of $\widehat{d}_r^{AG}$, for any given r; this link occurring via the introduction of the tuning parameter, δ (see (5.10) above), on which the finite sample performance of their estimator depends. In our method, (5.17) is used to control the increase in variance that occurs due to the reduction in bias, with the optimal weights determined by (5.17)-(5.19) not depending on any arbitrary quantities.*

5.4 Asymptotic results

The asymptotic properties of the optimal jackknife estimator depend on the optimal weights which, in turn, are functions of the covariance terms between the log-periodograms associated with the full sample and the sub-samples, as seen in (5.15) and (5.16). Provided that the DGP satisfies assumptions $(A.1) - (A.3)$, Lahiri (2003) has shown that periodogram ordinates are asymptotically independent when the frequencies are at a sufficient distance apart, provided that the set of observations remain the same. However, in our case, we are dealing with periodograms calculated both for the full set of observations, and for subsets of the full set.

Thus, two questions that arise here are: (i) Are the periodograms of the full sample and the sub-samples at different frequency ordinates asymptotically independent? and, (ii) When $d \neq 0$, do the periodograms still converge to a Chi-squared distribution as they do when $d = 0$ (see Theorem 5.2.6 of Brillinger, 1981)? We address both questions in Section 5.4.1 and provide formulae for calculating the relevant covariance terms algebraically, adopting the procedure used in Brillinger (1981). In Section 5.4.2 we then use these results to derive the asymptotic properties of the optimal jackknife estimator.

5.4.1 Stochastic properties of periodograms in the full sample and in sub-samples

We begin by defining $\{X_1, X_2, \ldots, X_h\}$ as an arbitrary set of h stationary time series. We link these series to the full sample and the m sub-samples of observations below. Our use of notation in this section mimics, in large part, that of Brillinger (1981, §. 2.6).

Definition 5.1 *Suppose $\{X_1, X_2, \ldots, X_h\}$ is a set of h stationary time series. The k^{th}-order cumulant $\kappa_{X_{a_1},\ldots,X_{a_k}}(u_1, \ldots, u_{k-1})$, for $k = 1, 2, \ldots, h$, and $u_j = 0, \pm 1, \pm 2 \ldots$ for $j = 1, 2, \ldots, k-1$, is defined as follows,*

$$\kappa_{X_{a_1},\ldots,X_{a_k}}(u_1, \ldots, u_{k-1}) = \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \exp\Big(-\imath \sum_{j=1}^{k-1} \lambda_j u_j\Big) f_{X_{a_1},\ldots,X_{a_k}}(\lambda_1, \ldots, \lambda_{k-1})\, d\lambda_1 \ldots d\lambda_{k-1}, \tag{5.26}$$

where $f_{X_{a_1},\ldots,X_{a_k}}(\lambda_1, \ldots, \lambda_{k-1})$ is the k^{th}-order joint spectral density of $\{X_{a_1}, \ldots, X_{a_k}\}$, for $-\pi < \lambda_j < \pi$, $j = 1, 2, \ldots, k-1$, with $a_1, \ldots, a_k = 1, 2, \ldots, h$, and $k = 1, 2, \ldots$.

For $\sum_{u_1=-\infty}^{\infty} \cdots \sum_{u_{k-1}=-\infty}^{\infty} \left|\kappa_{X_{a_1},\ldots,X_{a_k}}(u_1, \ldots, u_{k-1})\right| < \infty$, then the inverse form of (5.26) is given by,

$$f_{X_{a_1},\ldots,X_{a_k}}(\lambda_1, \ldots, \lambda_{k-1}) = (2\pi)^{-k+1} \sum_{u_1=-\infty}^{\infty} \cdots \sum_{u_{k-1}=-\infty}^{\infty} \kappa_{X_{a_1},\ldots,X_{a_k}}(u_1, \ldots, u_{k-1}) \exp\Big(-\imath \sum_{j=1}^{k-1} \lambda_j u_j\Big). \tag{5.27}$$

Now let $X_1 = \mathbf{y}$ denote the full sample of n observations on the random variable following the model in (5.1); whilst $X_{1+i} = \mathbf{y}_i$ denotes the vector of observations for the sub-sample $i = 1, 2, \ldots, m$, with length l. Set $h = m + 1$ in Definition 5.1. Let $D_{X_1}^{(n)}(.)$ and $D_{X_{1+i}}^{(l)}(.)$ respectively be the DFT of the full sample and i^{th} sub-sample at some frequency. Set

$$L_i = \begin{cases} n & if\; i = 1 \\ l & otherwise \end{cases} . \tag{5.28}$$

In Proposition 5.1 we give the expression for the k^{th}-order joint cumulant of the DFTs of the $h = m + 1$ series associated with the full sample and the m sub-samples.

Proposition 5.1 *Suppose Assumptions $(A.1) - (A.3)$ hold. The k^{th}-order cumulant of $\{D_{X_{a_1}}^{(L_1)}(\lambda_1)$, $D_{X_{a_2}}^{(L_2)}(\lambda_2)$, ..., $D_{X_{a_k}}^{(L_k)}(\lambda_k)\}$, for $k = 1, 2, \ldots$, is given by,*

$$\kappa_{D_{X_{a_1}},\ldots,D_{X_{a_k}}}(\lambda_1, \ldots, \lambda_{k-1}) = L^{-\frac{k}{2}} (2\pi)^{\frac{k}{2}-1} \Delta^{(L)}\Big(\sum_{j=1}^{k} \lambda_j\Big) f_{X_{a_1},\ldots,X_{a_k}}(\lambda_1, \ldots, \lambda_{k-1}) + o\Big(L^{1-2d-\frac{k}{2}}\Big), \tag{5.29}$$

where, $L = \min\{L_1, \ldots, L_k\}$.[4]

From Proposition 5.1 we can derive the relationship between the DFTs corresponding to full sample and the m sub-samples as the sample size increases. The result is given in the following theorem:

Theorem 5.2 *Suppose Assumptions $(A.1) - (A.4)$ hold, and suppose $\lambda = 2\pi r / L_i$ and $\omega = 2\pi s / L_j$ for integers r and s. Then for a fixed value of L_i and L_j, $D_{X_{a_i}}^{(L_i)}(\lambda)$ and $D_{X_{a_j}}^{(L_j)}(\mu)$ are asymptotically independent, whenever $\max\{L_i\lambda, L_j\mu\} \to \infty$, for $i \neq j$.*

[4]The k^{th}-order cumulant of associated with the DFTs should, for completeness, be denoted by $\kappa_{D_{X_{a_1}}^{(L_1)},\ldots,D_{X_{a_k}}^{(L_k)}}(., \ldots, .)$. For notational ease, however, we express the cumulant without making explicit the relevant sample sizes.

Theorem 5.2 immediately implies the asymptotic independence of the periodograms of the full sample and all sub-samples. However, in finite samples, the dependence structure across these periodograms may play an important role in determining the variance of the jackknife estimator in (5.14), through the form of the covariances in (5.15) and (5.16). Expressions for the covariances between the periodograms corresponding to the full sample and the sub-samples are provided in the following theorem, from which further insights on this point can be gleaned.

Theorem 5.3 *Let $I_{X_{a_i}}^{(L_i)}(\lambda)$ and $I_{X_{a_j}}^{(L_j)}(\lambda)$ be the periodograms associated with DFTs $D_{X_{a_i}}^{(L_i)}(\lambda)$ and $D_{X_{a_j}}^{(L_j)}(\mu)$ respectively. Suppose Assumptions $(A.1)-(A.3)$ hold. Then,*

$$\begin{aligned} Cov\left(I_{X_{a_i}}^{(L_i)}(\lambda), I_{X_{a_j}}^{(L_j)}(\mu)\right) &= \frac{2\pi}{L} f_{X_{a_i},X_{a_i},X_{a_j},X_{a_j}}(\lambda,-\lambda,\mu) + \frac{2\pi}{L}\left[\eta(\lambda-\mu)+\eta(\lambda+\mu)\right] \\ &\quad \times \left\{f_{X_{a_i}X_{a_j}}(\lambda)\right\}^2 + 2\pi\left[\eta(\lambda-\mu)+\eta(\lambda+\mu)\right], \\ &\quad \times f_{X_{a_i}X_{a_j}}(\lambda)\, o\left(L^{-2d}\right) + o\left(L^{-1-2d}\right), \end{aligned} \tag{5.30}$$

where $\eta(\omega) = \lim_{T\to\infty}\frac{1}{2\pi}\sum_{t=-T}^{T}\exp\{-\imath\omega t\}$, and L is as defined in Proposition 5.1. When Assumption $(A.4)$ also holds, the periodogram ordinates $I_{X_{a_i}}^{(L_i)}(\mu)$ and $I_{X_{a_j}}^{(L_j)}(\omega)$ with $i \neq j$, are asymptotically $f_{X_1X_1}(\cdot)\,\chi^2_{(2)}\big/2$ random variables.

Theorem 5.3 is a generalization of the result of Theorem 5.2.6 of Brillinger (1981) to the context of jackknifing. Equation (5.30) provides the first few dominant terms of the covariance between the periodograms associated with the full sample and a particular sub-sample, or between distinct sub-samples, at various frequency ordinates. Further, (5.30) reflects the fact that, for finite n, the relevant periodograms are positively correlated. This result is to be anticipated given that the sub-samples are subsets of the full sample and, hence, retain the same dependence structure as the full sample. Furthermore, the theorem states that the periodogram ordinates (for either the full sample and a given sub-sample, or between sub-samples) have a limiting joint distribution of the form, $f_{X_1X_1}(\lambda)\ \chi^2_{(2)}\Big/2$, where $f_{X_1X_1}(.)$ is the spectral density of the time series from which the full sample is generated.

Using the covariance terms and the distribution of the periodograms provided in the above theorem, we can find the joint distribution of the log-periodograms associated with the full sample and any sub-sample (or for two distinct sub-samples). Using the joint distribution of the log-periodograms, we can derive the moment generating function of the joint distribution. This leads to the derivation of the covariance terms for the log-periodogram. This result is provided in Appendix 5.B. The covariances between log-periodograms allow us to obtain the covariances between the full-sample and sub-sample LPR estimators given in (5.15) and (5.16). Exploiting the relationship between the different LPR estimators, we then establish the consistency and asymptotic normality of the optimal jackknife estimator in the following section.

5.4.2 Asymptotic properties of the optimal jackknife estimator

Using the results established in the previous section, we state the relationship between the full-sample and sub-sample LPR estimators in Theorem 5.4. The asymptotic properties of the optimal jackknife estimator are then established in Theorem 5.5.

Theorem 5.4 *Let $\widehat{d_n}$ and $\widehat{d_i}$ be the LPR estimators for the full-sample and the i^{th} sub-sample with sub-sample length, l. Suppose Assumptions $(A.1)-(A.4)$ hold. Then, for a fixed value of m,*

(i) *$\widehat{d_n}$ and $\widehat{d_i}$ are asymptotically independent.*

(ii) $\widehat{d}_i$ *and* $\widehat{d}_j$ *for* $i \neq j$, $i, j = 1, \ldots, m$, *are asymptotically independent.*

From Theorem 5.1, the LPR estimator constructed from the full sample is consistent and satisfies (5.8). Similarly, allowing the number of sub-samples, m, to be fixed (hence l changes as n changes such that $n = m \times l$), as $l \to \infty$, $\widehat{d}_i \to^P d_0$, and $\sqrt{N_l}(\widehat{d}_i - d_0) \to^D N\left(0, \frac{\pi^2}{24}\right)$. This implies the sub-sample LPR estimators have the same limiting distribution as the full-sample estimator. The asymptotic properties of $\widehat{d}^{Opt}_{J,m}$ are given in the following theorem.

Theorem 5.5 *Under the same assumptions and conditions given in Theorem 5.1, for a fixed value of* m*,*

$$\widehat{d}^{Opt}_{J,m} \to^P d_0, \text{ and } \sqrt{N_n}(\widehat{d}^{Opt}_{J,m} - d_0) \to^D N\left(0, \frac{\pi^2}{24}\right),$$

where d_0 *is the true value of* d *and* $\widehat{d}^{Opt}_{J,m}$ *is as given in (5.24).*

Thus, it follows from Theorem 5.5 that $\widehat{d}^{Opt}_{J,m}$ is consistent for d_0 and achieves a limiting normal distribution with the same variance as the base LPR estimator itself. Further, the rate of convergence of the optimal jackknife estimator, $\sqrt{N_n}$, is also the same as that of the LPR estimator. That is, there is no loss of asymptotic efficiency compared to $\widehat{d}_n$. Importantly, these asymptotic properties of the jackknife estimator do not depend on the number of sub-samples or the sub-sample length, as long as the former is fixed and the latter increases with n such that $n = m \times l$.

5.5 Simulation exercise

In this section, Monte Carlo simulation is used to compare the finite sample performance of the proposed jackknife estimator with: (i) the weighted-average estimator of Guggenberger and Sun (2006), $\widehat{d}_r^{GS}$, with $r = 1$, (ii) the bias-corrected prefiltered sieve bootstrap-based estimator of Poskitt *et al.* (2016), $\widehat{d}^{PFSB}$, and, (iii) the unadjusted LPR estimator, $\widehat{d}_n$. Performance is assessed in terms of bias and RMSE, and under a variety of DGPs. All numerical results are produced using MATLAB 2015*b*, version 8.6.0.267246.

5.5.1 Monte Carlo design

Data are generated from two stationary fractional processes where, without loss of generality, it is assumed that the process mean is zero. The two processes considered are the ARFIMA$(1, d_0, 0)$ and ARFIMA$(0, d_0, 1)$ models, given respectively by

$$(1 + \phi_0 B)\,(1 - B)^{d_0}\, Y_t = \varepsilon_t, \text{ and } (1 - B)^{d_0}\, Y_t = (1 + \theta_0 B)\,\varepsilon_t, \tag{5.31}$$

where B is the backward shift operator, $B^k x_t = x_{t-k}$, for $k = 1, 2, \ldots$, and $\varepsilon_t \sim i.i.d\ (0, 1)$. We consider two alternative distributions for ε_t, namely, (i) Gaussian, and, (ii) Student t with 5 degrees of freedom. For the parameter of interest, d, we consider true values, $d_0 = \{-0.25, 0, 0.25, 0.45\}$. Values from the set $\{-0.9, -0.4, 0.4\}$ are adopted for both ϕ_0 and θ_0.

Sample sizes $n \in \{96, 576\}$ are considered. These values are chosen to reflect the size of samples used in real world examples (see, for example, Diebold *et al.*, 1991, Delgado and Robinson, 1994, Gil-Alana and Robinson, 1997, and Reisen and Lopes, 1999). However, one should note that, in general, the size of data sets from finance, in particular those recorded at high frequency (for example, Granger and Hyung, 2004), or from biology (for example, the tree-ring data set of Contreras-reyes and Palma, 2013), or in certain other of the examples mentioned in the Introduction, are much larger than the sample sizes considered here. On the other hand, these sample sizes are large enough to enable a range of values for the number of sub-samples, m, to be explored, with the chosen range of m being $\{2, 3, 4, 6, 8\}$. We also consider only sub-samples that have equal length, $l = n/m$, under both sub-sampling approaches.

We adopt the following procedure in implementing the jackknife bias-adjustment technique:

Step 1: Generate the full sample of size n, $\mathbf{y}$, from the relevant stationary ARFIMA(p, d_0, q) model.

Step 2: Compute the LPR estimator of d_0, $\widehat{d}_n$ using (5.5).

Step 3: Draw the sub-samples, $\mathbf{y}_i$ $(i = 1, 2, ..., m)$, from the full sample based on the relevant sub-sampling technique (non-overlapping or moving-block) and compute the LPR estimator of d_0, $\widehat{d}_i$, for each sub-sample.

Step 4: Depending on the sub-sample selection method chosen in Step 3, obtain the optimal weights for the corresponding method and compute the optimal jackknife estimator, $\widehat{d}_{J,m}^{Opt}$.

Step 5: Repeat Steps $1 - 4$ $100,000$ times and compute estimates of the bias and RMSE of the optimal jackknife estimator.

In Steps 2 and 3, the number of frequencies used to calculate the relevant LPR estimator is set to $N_L = \lfloor L^{\alpha} \rfloor$, with $\alpha = 0.65$, where L is as defined in (5.28). The optimal jackknife estimators calculated using the non-overlapping (abbreviated to Opt-NO), and moving-block (abbreviated to Opt-MB) schemes, are denoted by $\widehat{d}_{J,m}^{Opt-NO}$ and $\widehat{d}_{J,m}^{Opt-MB}$, respectively.

The weighted-average estimator of Guggenberger and Sun (2006) is computed as described in Section 5.2.2, with the following additional details. For a given N_n, the set of bandwidths used to calculate the constituent estimators in (5.10) are $N_{n,i} = \lfloor q_i N_n \rfloor$, where $\mathbf{q}^\top = (q_1, q_2, \ldots, q_K) = (1, 1.05, \ldots, 2)$. We produce the GS estimator (based on $r = 1$) using two different choices of N_n: (i) $N_n = \lfloor n^\alpha \rfloor$, with $\alpha = 0.65$ (denoting this estimator by $\widehat{d}_1^{GS}$), and (ii) the optimal choice of N_n as suggested in Guggenberger and Sun (2006, page 876) (denoting this version by $\widehat{d}_1^{Opt-GS}$). Importantly, this optimal choice of bandwidth depends on knowledge of the true values of the short memory parameters. The parameter δ, required for both versions of the GS estimator, is evaluated using the formula $\delta = \tau_r / (\tau_r^* \sum_{i=1}^K q_k^{2+2r})$, where $\tau_{r-1}^* = -\, (2\pi)^{2r} r \Big/ [(2r)!\,(2r+1)^2]$ and the number τ_r is as defined in Andrews and Guggenberger (2003). Details regarding the construction of the pre-filtered sieve bootstrap estimator ($\widehat{d}^{PFSB}$) can be found in Poskitt *et al.* (2016). In implementing this method, we set the number of bootstrap samples to $B = 1000$.[5]

5.5.2 Finite sample bias and RMSE

In this section, we document the relative performance of the jackknife method in two scenarios: (i) when the true parameters are assumed to be known and are used in the construction of the optimal jackknife weights, and, (ii) when they are unknown. The relevant finite sample results are presented in Section 5.5.2.1 and Section 5.5.2.2 respectively. In case (i) we compare the jackknife estimator with the GS estimator obtained with the optimal choice of N_n ($\widehat{d}_1^{Opt-GS}$) - which, of course, relies on the known values of the short memory parameters - and with the sub-optimal estimator, $\widehat{d}_1^{GS}$. In case (ii) results for only $\widehat{d}_1^{GS}$ are included, as $\widehat{d}_1^{Opt-GS}$ is infeasible.[6] An iterative method is used to produce a feasible version of the jackknife estimator in

[5] Certain simulation results based on $\alpha = 0.5$ have also been produced but are not presented in the main text. These results are presented in Appendix 5.C.

[6] Note that in the case where the short-memory dynamics are unknown Guggenberger and Sun (2006) suggest that an adaptive procedure for the local Whittle-based estimator that they propose could be extended to the weighted-average estimator based on LPR. Since the adaptive method is not provided in detail in their paper, we do not pursue this option here.

this case. Note that the finite sample results for the (raw) LPR and PFSB estimators remain the same in both scenarios, as the construction of neither estimator relies on knowledge of any of the true parameters. To save on space, results for $\widehat{d}_{J,m}^{Opt-NO}$ are recorded for the full range of values for m, whilst results for $\widehat{d}_{J,m}^{Opt-MB}$ based on only $m = 2$ are documented. We do note that the patterns exhibited (in terms of both bias and RMSE) for $\widehat{d}_{J,m}^{Opt-MB}$, across m, are similar to those exhibited for $\widehat{d}_{J,m}^{Opt-NO}$.

5.5.2.1 Case 1: True parameters are known

Tables 5.2 and 5.3 record the bias and RMSE of the various optimal jackknife estimators, the two different GS estimators, and the LPR and PFSB estimators, for case where the DGP is ARFIMA$(1, d_0, 0)$ and the short memory parameter ϕ_0 is known and the bandwidth corresponding to $\alpha = 0.65$. The corresponding results for the ARFIMA$(0, d_0, 1)$ DGP are presented in Tables 5.4 and 5.5. The top panel of each table displays the results based on Gaussian errors and the bottom panel of each, the results based on Student t errors with 5 degrees of freedom (denoted by Student t_5 hereafter). The lowest biases and RMSEs for each design are marked in boldface. Similarly, the results for $\alpha = 0.5$ are presented in Tables 5.10 – 5.13 for ARFIMA$(1, d_0, 0)$ and ARFIMA$(0, d_0, 1)$ DGPs.

We shall begin the discussion on the bias and RMSE results based on $\alpha = 0.65$. With reference to Tables 5.2 and 5.3: as is consistent with existing results (see, for example, Agiakloglou *et al.*, 1993, Nielsen and Frederiksen, 2005 and Poskitt *et al.*, 2016) when short memory dynamics are present, the raw, unadjusted, LPR estimator is biased, as the low frequencies are contaminated by the spectral density of the short run dynamics, particularly for *negative* values of ϕ (which corresponds to positive autocorrelation). As is evident from the recorded results, the bias is particularly large when there is a large negative value for ϕ_0 in (5.31), and it decreases as this value increases. Further, both bias and RMSE decline as the sample size increases, illustrating the consistency of the estimator. These characteristics of the LPR estimator are in evidence for both error processes: Gaussian and Student t_5.

We shall now comment on the performance of all nine bias-corrected estimators under the ARFIMA $(1, d_0, 0)$ process. With reference to Table 5.2, for the great majority of designs, $\widehat{d}_{J,m}^{Opt-NO}$ with $m = 2$, has the smallest bias of all, and uniformly for $\phi_0 = -0.9$. For $\phi_0 = -0.9$ and $n = 96$, the bias reduction of $\widehat{d}_{J,m}^{Opt-NO}$ ($m = 2$), relative to the raw LPR estimator is up to 3.6%, and when $n = 576$, this rises to 5.7%.[7] For the larger values of ϕ_0, when $n = 96$, the bias reduction ranges from 27% to 82%, and from 67% to 97% when $n = 576$. Only occasionally is this particular version of the jackknife estimator inferior to an alternative bias-adjusted estimator. Importantly, however, an increase in m leads to an increase in bias for $\widehat{d}_{J,m}^{Opt-NO}$ and, hence, a reduction in its superiority over all alternatives, including the raw LPR method. The reason is that the increase in m leads to a smaller sub-sample length and, hence, increases the finite sample impact of the dominant bias term on the sub-sample estimators used in the construction of the jackknife estimator.

Now referencing the results in Table 5.3, we see that despite the lack of variance inflation in the asymptotic distribution of the optimal jackknife estimator, the reduction in bias does cause some finite sample increase in variance, leading to RMSEs for $\widehat{d}_{J,m}^{Opt-NO}$ that are occasionally slightly larger than the RMSE of the raw LPR estimator. That said, in the vast majority of cases $\widehat{d}_{J,m}^{Opt-NO}$ with $m = 8$, has the smallest RMSE of all estimators (including the raw LPR) and, in many cases, the RMSE of the jackknife estimator with the smallest bias ($\widehat{d}_{J,m}^{Opt-NO}$, $m = 2$)

[7] We remind the reader that when $\phi_0 = -0.9$ all estimators remain very biased.

has a RMSE which remains less than that of the raw estimator. In addition, all versions of the jackknife estimator (including the moving-block version) tend to have smaller RMSEs than the three alternative bias-corrected methods ($\widehat{d}_1^{GS}$, $\widehat{d}_1^{Opt-GS}$ and $\widehat{d}^{PFSB}$), most notably for the smaller sample size $(n = 96)$. As befits the optimality of the estimator, in almost all cases, $\widehat{d}_1^{Opt-GS}$ out-performs $\widehat{d}_1^{GS}$, in terms of both bias and RMSE, although both estimators, as already noted, are virtually always out-performed by a version of the jackknife procedure.

The bias and RMSE features of the LPR estimator and all nine bias-corrected estimators do hold even for the bandwidth associated with $\alpha = 0.5$, (see Tables 5.10 and 5.11). Importantly, when $\alpha = 0.5$ all the estimators exhibit smaller bias and larger RMSE compared to case that of $\alpha = 0.65$ for all ϕ_0 and d_0. Due to the very large RMSE that comes with the choice of smaller bandwidth, we recommend using larger bandwidth.

The broad conclusions drawn above obtain under both specifications for the innovations, and also under the ARFIMA$(0, d_0, 1)$ DGP, as seen from the results recorded in Tables 5.4 and 5.5.

5.5.2.2 Case 2: True parameters are unknown

Evaluation of the optimal weights in (5.23), required for the construction of the optimal jackknife estimator, depends on the covariances between both the different sub-sample LPR estimators and between the full-sample and sub-sample estimators, as given in (5.15) and (5.16). These covariances depend, in turn, on covariances between the various log-periodograms and, hence, on the values of the parameters that underpin the true DGP, as is made explicit in (5.30) and Appendix 5.B. Hence, implementation of the optimal bias-correction procedure via the jackknife is not feasible in practice, without modification. To this end, we propose the

Table 5.2: Bias estimates of the unadjusted LPR estimator, the optimal jackknife estimator based on 2,3,4,6,8 non-overlapping (NO) sub-samples, the optimal jackknife estimator based on 2 moving block (MB) sub-samples, both versions of the GS estimator, and the prefiltered sieve bootstrap estimator, for the DGP: ARFIMA$(1, d_0, 0)$. The estimates are obtained under Gaussian and Student t_5 innovations, with $\alpha = 0.65$.

ϕ_0	d_0	n	$\widehat{d}_n$	$\widehat{d}_{J,2}^{Opt-NO}$	$\widehat{d}_{J,3}^{Opt-NO}$	$\widehat{d}_{J,4}^{Opt-NO}$	$\widehat{d}_{J,6}^{Opt-NO}$	$\widehat{d}_{J,8}^{Opt-NO}$	$\widehat{d}_{J,2}^{Opt-MB}$	$\widehat{d}_1^{GS}$	$\widehat{d}_1^{Opt-GS}$	$\widehat{d}^{PFSB}$
							Gaussian					
-0.9	-0.25	96	0.8145	**0.7852**	0.7903	0.7995	0.8072	0.8120	0.8156	0.8002	0.7902	0.7908
		576	0.5945	**0.5614**	0.5682	0.5726	0.5804	0.5946	0.5841	0.5724	0.5657	0.5898
	0	96	0.8053	**0.7865**	0.7945	0.7988	0.8042	0.8169	0.7927	0.8015	0.7957	0.7955
		576	0.5912	**0.5581**	0.5627	0.5699	0.5773	0.5843	0.5608	0.5761	0.5630	0.5888
	0.25	96	0.7752	**0.7477**	0.7515	0.7694	0.7747	0.7804	0.7799	0.7673	0.7517	0.7685
		576	0.5883	**0.5553**	0.5622	0.5687	0.5731	0.5816	0.5673	0.5716	0.5628	0.5638
	0.45	96	0.7006	**0.6783**	0.6842	0.6905	0.7046	0.7172	0.6945	0.6946	0.6846	0.6705
		576	0.5748	**0.5423**	0.5487	0.5535	0.5586	0.5629	0.5567	0.5659	0.5580	0.5451
-0.4	-0.25	96	0.1756	**0.1223**	0.1344	0.1459	0.1563	0.1660	0.1560	0.1367	0.1286	0.1435
		576	0.0607	**0.0043**	0.0429	0.0534	0.0585	0.0599	0.0599	0.0304	0.0245	0.0286
	0	96	0.1653	**0.1203**	0.1216	0.1395	0.1596	0.1674	0.1674	0.1304	0.1276	0.1353
		576	0.0560	**0.0127**	0.0253	0.0307	0.0479	0.0569	0.0369	0.0264	0.0152	0.0249
	0.25	96	0.1629	**0.1190**	0.1274	0.1314	0.1508	0.1665	0.0731	0.1329	0.1276	0.1294
		576	0.0571	**0.0179**	0.0243	0.0341	0.0431	0.0599	0.0599	0.0289	0.0181	0.0251
	0.45	96	0.1653	**0.1154**	0.1226	0.1353	0.1560	0.1702	0.1702	0.1400	0.1245	0.1277
		576	0.0625	**0.0203**	0.0325	0.0495	0.0518	0.0667	0.0667	0.0359	0.0217	0.0261
0.4	-0.25	96	-0.0363	-0.0194	-0.0136	-0.0259	-0.0323	-0.0493	-0.0393	**-0.0047**	-0.0068	-0.0147
		576	-0.0056	**-0.0004**	-0.0037	-0.0046	-0.0057	-0.0076	-0.0076	0.0056	-0.0027	-0.0004
	0	96	-0.0534	-0.0114	-0.0145	-0.0298	-0.0360	-0.0449	-0.0549	**-0.0089**	-0.0092	-0.0175
		576	-0.0125	**-0.0007**	-0.0049	-0.0038	-0.0031	-0.0028	-0.0128	-0.0008	-0.0007	-0.0040
	0.25	96	-0.0559	-0.0121	-0.0188	-0.0281	-0.0350	-0.0458	-0.0558	-0.0068	**-0.0050**	-0.0153
		576	-0.0115	-0.0003	-0.0014	-0.0024	-0.0079	-0.0100	-0.0100	0.0017	**-0.0008**	-0.0027
	0.45	96	-0.0501	-0.0091	-0.0092	-0.0302	-0.0460	-0.0486	-0.0486	**0.0032**	0.0090	-0.0111
		576	-0.0058	**-0.0003**	-0.0037	-0.0054	-0.0062	-0.0078	-0.0028	0.0089	-0.0061	0.0004
							Student t_5					
-0.9	-0.25	96	0.8123	**0.7739**	0.7895	0.7921	0.7993	0.8042	0.7913	0.7914	0.7856	0.7847
		576	0.5952	**0.5621**	0.5693	0.5740	0.5805	0.5873	0.5746	0.5863	0.5775	0.5770
	0	96	0.8034	**0.7749**	0.7816	0.7895	0.7927	0.7988	0.7822	0.7843	0.7763	0.7830
		576	0.5915	**0.5516**	0.5644	0.5716	0.5780	0.5853	0.5769	0.5642	0.5640	0.5539
	0.25	96	0.7726	**0.7457**	0.7564	0.7622	0.7693	0.7749	0.7626	0.7633	0.7536	0.7572
		576	0.5883	**0.5453**	0.5631	0.5684	0.5733	0.5798	0.5657	0.5633	0.5532	0.5472
	0.45	96	0.7002	**0.6714**	0.6719	0.6781	0.6829	0.6941	0.6870	0.6849	0.6780	0.6731
		576	0.5758	**0.5434**	0.5511	0.5584	0.5612	0.5679	0.5548	0.5602	0.5587	0.5514
-0.4	-0.25	96	0.1764	**0.1326**	0.1341	0.1457	0.1566	0.1467	0.1632	0.1371	0.1263	0.1422
		576	0.0611	**0.0140**	0.0233	0.0238	0.0281	0.0302	0.0244	0.0305	0.0246	0.0289
	0	96	0.1662	**0.1205**	0.1215	0.1295	0.1301	0.1384	0.1269	0.1307	0.1259	0.1340
		576	0.0565	0.0230	0.0258	0.0312	0.0374	0.0472	0.0347	0.0266	**0.0175**	0.0252
	0.25	96	0.1640	**0.1196**	0.1279	0.1319	0.1374	0.1381	0.1276	0.1334	-0.1237	0.1282
		576	0.0575	0.0184	0.0149	**0.0128**	0.0176	0.0201	0.0128	0.0292	-0.0163	0.0254
	0.45	96	0.1666	**0.1033**	0.1060	0.1100	0.1163	0.1214	0.1228	0.1405	-0.1374	0.1270
		576	0.0627	**0.0206**	0.0229	0.0300	0.0414	0.0466	0.0402	0.0359	-0.0142	0.0627
0.4	-0.25	96	-0.0357	-0.0116	-0.0180	-0.0232	-0.0016	**-0.0014**	-0.0035	-0.0054	-0.0089	-0.0132
		576	-0.0052	-0.0023	-0.0045	-0.0081	-0.0106	-0.0075	-0.0081	-0.0054	-0.0024	**0.0003**
	0	96	-0.0525	**-0.0148**	-0.0192	-0.0179	-0.0158	-0.0141	-0.0144	-0.0081	-0.0077	-0.0164
		576	-0.0121	-0.0036	-0.0045	-0.0082	-0.0095	-0.0116	-0.0093	**-0.0006**	-0.0038	-0.0033
	0.25	96	-0.0641	**-0.0034**	-0.0076	-0.0178	-0.0143	-0.0244	-0.0175	-0.0062	-0.0056	-0.0165
		576	-0.0182	0.0016	0.0014	**0.0002**	-0.0083	-0.0098	-0.0027	-0.0019	-0.0039	-0.0045
	0.45	96	-0.0489	-0.0198	-0.0085	-0.0197	-0.0258	-0.0274	-0.0166	**-0.0040**	-0.0100	-0.0097
		576	-0.0055	**-0.0008**	-0.0031	-0.0060	-0.0025	-0.0029	-0.0016	-0.0087	-0.0027	0.0008*

Table 5.3: RMSE estimates of the unadjusted LPR estimator, the optimal jackknife estimator based on 2,3,4,6,8 non-overlapping (NO) sub-samples, the optimal jackknife estimator based on 2 moving block (MB) sub-samples, both versions of the GS estimator, and the prefiltered sieve bootstrap estimator, for the DGP: ARFIMA$(1, d_0, 0)$. The estimates are obtained under Gaussian and Student t_5 innovations, with $\alpha = 0.65$.

ϕ_0	d_0	n	$\widehat{d}_n$	$\widehat{d}_{J,2}^{Opt-NO}$	$\widehat{d}_{J,3}^{Opt-NO}$	$\widehat{d}_{J,4}^{Opt-NO}$	$\widehat{d}_{J,6}^{Opt-NO}$	$\widehat{d}_{J,8}^{Opt-NO}$	$\widehat{d}_{J,2}^{Opt-MB}$	$\widehat{d}_1^{GS}$	$\widehat{d}_1^{Opt-GS}$	$\widehat{d}^{PFSB}$
							Gaussian					
-0.9	-0.25	96	1.0359	1.0627	1.0532	1.0596	1.0358	**1.0286**	1.1837	1.3386	1.1864	1.2885
		576	0.7398	0.7490	0.7403	0.7372	0.7325	**0.7299**	0.7382	0.7371	0.7200	0.7359
	0	96	1.1148	1.1398	1.1275	1.1158	1.1080	**1.0966**	1.1576	1.1819	1.1120	1.2167
		576	0.8288	0.8370	0.8311	0.8294	0.8216	0.8157	0.8215	0.8173	0.8173	**0.8053**
	0.25	96	1.1618	1.1857	1.1066	1.0971	1.0944	**1.0913**	1.1162	1.1484	1.1285	1.2299
		576	0.9175	0.9250	0.9203	0.9186	0.9128	**0.9076**	0.9115	1.1171	1.0172	1.1130
	0.45	96	1.1286	1.1552	1.1325	1.1294	1.1200	**1.1168**	1.1132	1.4331	1.3331	1.5385
		576	0.9708	0.9781	0.9732	0.9650	0.9558	**0.9546**	0.9687	1.1124	1.0524	1.1647
-0.4	-0.25	96	0.2568	0.2292	0.2568	0.2422	0.2384	**0.2376**	0.2576	0.2594	0.2441	0.3028
		576	0.1098	0.0978	0.0974	0.0884	**0.0873**	0.0896	0.1096	0.1118	0.0995	0.1272
	0	96	0.2498	0.2395	0.2284	0.2146	0.2138	**0.2117**	0.2517	0.2560	0.2416	0.2930
		576	0.1069	0.0837	0.0879	0.0819	0.0787	**0.0778**	0.1078	0.1104	0.0967	0.1247
	0.25	96	0.2490	0.2678	0.2574	0.2435	0.2354	**0.2254**	0.3254	0.2580	0.2404	0.2879
		576	0.1079	0.1036	0.0965	0.0901	0.0819	**0.0797**	0.1097	0.1115	0.1029	0.1239
	0.45	96	0.2506	0.2615	0.2563	0.2434	0.2390	**0.2243**	0.2544	0.2616	0.2511	0.2506
		576	0.1115	0.0963	0.0878	0.0808	0.0777	**0.0742**	0.1142	0.1143	0.1005	0.1230
0.4	-0.25	96	0.1917	0.1721	0.1654	0.1629	0.1544	**0.1529**	0.1929	0.2212	0.2157	0.2717
		576	0.0919	0.0762	0.0747	0.0665	0.0632	**0.0624**	0.0924	0.1081	0.0695	0.1198
	0	96	0.1946	0.1726	0.1717	0.1631	0.1569	**0.1557**	0.1957	0.2203	0.2162	0.2546
		576	0.0920	0.0890	0.0793	0.0751	0.0730	**0.0724**	0.0924	0.1073	0.0684	0.1166
	0.25	96	0.1960	0.2107	0.2063	0.2008	**0.1913**	0.1966	0.1966	0.2209	0.2091	0.2482
		576	0.0922	0.0705	0.0696	0.0644	0.0627	**0.0624**	0.0924	0.1076	0.0688	0.1158
	0.45	96	**0.1955**	0.2178	0.2140	0.2085	0.2061	0.2058	0.1958	0.2218	0.2143	0.2453
		576	0.0926	0.0710	0.0684	0.0667	0.0634	**0.0569**	0.0929	0.1089	0.0701	0.1149
							Student t_5					
-0.9	-0.25	96	1.0321	1.0600	0.9961	0.9820	0.9723	**0.9641**	0.9862	1.1741	1.0708	1.0570
		576	0.7408	0.7501	0.7415	0.7386	0.7159	**0.7004**	0.7439	0.7406	0.7215	0.7309
	0	96	1.1120	1.1373	1.1120	1.1085	1.0958	**1.0767**	1.0946	1.2792	1.1728	1.2542
		576	0.8291	0.8376	0.8216	0.8173	0.8066	**0.7914**	0.8264	0.8484	0.8181	0.8367
	0.25	96	1.1577	1.1822	1.1648	1.1432	1.1257	**1.1169**	1.1384	1.2620	1.1624	1.2967
		576	0.9173	0.9248	0.9155	0.9048	0.8937	**0.8845**	0.8762	0.9174	0.9076	0.9133
	0.45	96	1.1272	1.1533	1.1762	1.1520	1.1344	**1.1159**	1.1254	1.2314	1.2058	1.2848
		576	0.9720	0.9793	0.9640	0.9536	0.9428	**0.9342**	0.9595	0.9643	0.9532	0.9755
-0.4	-0.25	96	0.2562	0.2901	0.2659	0.2613	0.2579	0.2553	0.3156	0.2587	**0.2415**	0.3008
		576	0.1096	0.1078	0.1075	0.1083	0.1090	0.1093	**0.0961**	0.1109	0.1064	0.1264
	0	96	0.2492	0.2403	0.2376	0.2337	0.2330	**0.2315**	0.2643	0.2552	0.2476	0.2912
		576	0.1069	0.0939	0.0982	0.0920	0.0884	**0.0875**	0.0919	0.1095	0.0900	0.1241
	0.25	96	0.2487	0.2384	0.2367	0.2327	0.2246	**0.2216**	0.2550	0.2567	0.2418	0.2865
		576	0.1078	0.1040	0.1367	0.1201	0.1116	**0.1023**	0.1040	0.1106	0.1095	0.1233
	0.45	96	0.2509	0.2475	0.2464	0.2335	0.2293	**0.2247**	0.2346	0.2610	0.2549	0.2881
		576	0.1115	0.1067	0.1032	0.1009	0.0974	**0.0941**	0.0965	0.1137	0.1010	0.1228
0.4	-0.25	96	**0.1907**	0.2142	0.2075	0.2038	0.1958	0.1956	0.2001	0.2202	0.2112	0.2698
		576	0.0915	0.0955	0.0944	0.0860	0.0832	**0.0820**	0.1178	0.1076	0.0943	0.1190
	0	96	0.1930	0.1853	0.1756	0.1625	0.1550	**0.1543**	0.1915	0.2181	0.2004	0.2532
		576	0.0915	0.0955	0.0944	0.0860	0.0832	**0.0820**	0.1178	0.1076	0.0997	0.1190
	0.25	96	0.1977	0.1889	0.1844	0.1792	0.1758	**0.1750**	0.2016	0.2193	0.2019	0.2361
		576	0.0927	0.0998	0.0953	0.0940	0.0925	**0.0918**	0.1116	0.1072	0.0981	0.1216
	0.45	96	0.1942	0.1864	0.1724	0.1671	0.1655	**0.1546**	0.2147	0.2201	0.2048	0.2440
		576	0.0924	0.0887	0.0804	0.0764	0.0733	**0.0728**	0.1009	0.1082	0.0942	0.1142

Table 5.4: Bias estimates of the unadjusted LPR estimator, the optimal jackknife estimator based on 2,3,4,6,8 non-overlapping (NO) sub-samples, the optimal jackknife estimator based on 2 moving block (MB) sub-samples, both versions of the GS estimator, and the prefiltered sieve bootstrap estimator, for the DGP: ARFIMA$(0, d_0, 1)$. The estimates are obtained under Gaussian and Student t_5 innovations, with $\alpha = 0.65$.

θ_0	d_0	n	$\widehat{d}_n$	$\widehat{d}_{J,2}^{Opt-NO}$	$\widehat{d}_{J,3}^{Opt-NO}$	$\widehat{d}_{J,4}^{Opt-NO}$	$\widehat{d}_{J,6}^{Opt-NO}$	$\widehat{d}_{J,8}^{Opt-NO}$	$\widehat{d}_{J,2}^{Opt-MB}$	$\widehat{d}_1^{GS}$	$\widehat{d}_1^{Opt-GS}$	$\widehat{d}^{PFSB}$
						Gaussian						
-0.9	-0.25	96	-0.5671	**-0.5276**	-0.5348	-0.5429	-0.5574	-0.5653	-0.5536	-0.5450	-0.5329	-0.5466
		576	-0.4527	**-0.4149**	-0.4266	-0.4357	-0.4404	-0.4595	-0.4375	-0.4385	-0.4248	-0.4285
	0	96	-0.7042	**-0.6416**	-0.6502	-0.6642	-0.6743	-0.6869	-0.6724	-0.6575	-0.6476	-0.6664
		576	-0.5594	**-0.5112**	-0.5259	-0.5384	-0.5469	-0.5572	-0.5346	-0.5256	-0.5156	-0.5375
	0.25	96	-0.7763	**-0.7299**	-0.7345	-0.7466	-0.7547	-0.7681	-0.7367	-0.7524	-0.7425	-0.7661
		576	-0.5880	**-0.5299**	-0.5374	-0.5450	-0.5581	-0.5623	-0.5348	-0.5473	-0.5373	-0.5621
	0.45	96	-0.8004	**-0.7414**	-0.7588	-0.7615	-0.7741	-0.7878	-0.7649	-0.7600	-0.7501	-0.7854
		576	-0.5880	**-0.5061**	-0.5127	-0.5349	-0.5457	-0.5537	-0.5224	-0.5351	-0.5151	-0.5527
-0.4	-0.25	96	-0.1437	**-0.1013**	-0.1152	-0.1105	-0.1211	-0.1371	-0.1271	-0.1120	-0.1057	-0.1240
		576	-0.0476	-0.0342	-0.0234	-0.0139	-0.0234	-0.0303	-0.0303	-0.0187	**-0.0123**	-0.0271
	0	96	-0.1653	**-0.1199**	-0.1213	-0.1293	-0.1394	-0.1472	-0.1472	-0.1305	-0.1209	-0.1248
		576	-0.0560	**-0.0226**	-0.0353	-0.0407	-0.0579	-0.0570	-0.0370	-0.0265	-0.0274	-0.0307
	0.25	96	-0.1692	**-0.1136**	-0.1273	-0.1292	-0.1398	-0.1496	-0.1496	-0.1297	-0.1170	-0.1200
		576	-0.0552	**-0.0122**	-0.0366	-0.0475	-0.0529	-0.0543	-0.0443	-0.0243	-0.0160	-0.0287
	0.45	96	-0.1630	**-0.0712**	-0.1374	-0.1510	-0.1605	-0.1620	-0.1420	-0.1190	-0.1036	-0.1118
		576	-0.0493	**-0.0155**	-0.0177	-0.0314	-0.0436	-0.0436	-0.0268	-0.0169	-0.0126	-0.0244
0.4	-0.25	96	0.0637	**0.0036**	0.0475	0.0563	0.0628	0.0637	0.0437	0.0154	0.0092	0.0651
		576	0.0175	**0.0037**	0.0092	0.0068	0.0141	0.0161	0.0061	0.0049	0.0040	0.0132
	0	96	0.0525	0.0202	0.0234	0.0288	0.0351	0.0340	0.0340	0.0081	**0.0077**	0.0603
		576	0.0125	0.0088	0.0148	0.0137	0.0130	0.0128	0.0088	**0.0006**	0.0007	0.0100
	0.25	96	0.0504	0.0164	0.0397	0.0511	0.0566	0.0535	0.0335	0.0110	**0.0095**	0.0574
		576	0.0136	**0.0028**	0.0048	0.0072	0.0083	0.0157	0.0057	0.0031	0.0030	0.0108
	0.45	96	0.0549	0.0192	0.0375	0.0474	0.0641	0.0592	0.0393	0.0204	**0.0112**	0.0570
		576	0.0192	**0.0049**	0.0072	0.0069	0.0073	0.0129	0.0119	0.0103	0.0050	0.0132
						Student t_5						
-0.9	-0.25	96	-0.5754	**-0.5194**	-0.5249	-0.5357	-0.5486	-0.5549	-0.5375	-0.5479	-0.5373	-0.5553
		576	-0.4589	**-0.3941**	-0.4043	-0.4129	-0.4261	-0.4384	-0.4158	-0.4275	-0.4196	-0.4103
	0	96	-0.7073	**-0.6270**	-0.6368	-0.6425	-0.6574	-0.6682	-0.6466	-0.6427	-0.6379	-0.6638
		576	-0.5613	**-0.5139**	-0.5259	-0.5340	-0.5473	-0.5583	-0.5242	-0.5366	-0.5291	-0.5570
	0.25	96	-0.7814	**-0.7172**	-0.7272	-0.7364	-0.7468	-0.7514	-0.7344	-0.7373	-0.7216	-0.7477
		576	-0.5876	**-0.5294**	-0.5341	-0.5448	-0.5527	-0.5643	-0.5409	-0.5478	-0.5378	-0.5532
	0.45	96	-0.8032	**-0.7449**	-0.7562	-0.7662	-0.7749	-0.7828	-0.7641	-0.6661	-0.7654	-0.7880
		576	-0.5875	**-0.5151**	-0.5247	-0.5384	-0.5466	-0.5571	-0.5349	-0.5364	-0.5159	-0.5438
-0.4	-0.25	96	-0.1442	-0.1119	-0.1264	-0.1342	-0.1465	-0.1551	-0.1482	-0.1117	**-0.0985**	-0.1224
		576	-0.0477	**-0.0103**	-0.0264	-0.0342	-0.0463	-0.0582	-0.0462	-0.0187	-0.0106	-0.0208
	0	96	-0.1646	**-0.1101**	-0.1183	-0.1242	-0.1375	-0.1462	-0.1558	-0.1299	-0.1140	-0.1259
		576	-0.0559	**-0.0127**	-0.0326	-0.0257	-0.0365	-0.0486	-0.0462	-0.0265	-0.0157	-0.0264
	0.25	96	-0.1686	**-0.1122**	-0.1264	-0.1358	-0.1467	-0.1582	-0.1432	-0.1290	-0.1154	-0.1211
		576	-0.0548	**-0.0123**	-0.0257	-0.0299	-0.0306	-0.0397	-0.0267	-0.0242	-0.0179	-0.0248
	0.45	96	-0.1621	**-0.0698**	-0.0712	-0.0793	-0.0862	-0.0944	-0.0885	-0.1183	-0.1043	-0.1071
		576	-0.0492	**-0.0157**	-0.0254	-0.0332	-0.0397	-0.0453	-0.0262	-0.0169	-0.0178	-0.0209
0.4	-0.25	96	0.0648	**0.0037**	0.0099	0.0176	0.0224	0.0346	0.0448	0.0159	0.0103	0.0187
		576	0.0179	**0.0025**	0.0158	0.0193	0.0247	0.0331	0.0134	0.0051	0.0030	0.0074
	0	96	0.0529	0.0193	0.0415	0.0481	0.0516	0.0564	0.0442	0.0084	**0.0060**	0.0145
		576	0.0122	0.0059	0.0086	0.0056	0.0095	0.0103	0.0186	0.0008	**0.0006**	0.0038
	0.25	96	0.0505	0.0111	0.0168	0.0193	0.0215	0.0397	0.0375	0.0116	**0.0082**	0.0151
		576	0.0140	**0.0024**	0.0064	0.0095	0.0119	0.0168	0.0081	0.0033	0.0030	0.0053
	0.45	96	0.0561	**0.0097**	0.0276	0.0334	0.0415	0.0483	0.0382	0.0209	0.0100	0.0187
		576	0.0194	**0.0043**	0.0096	0.0126	0.0143	0.0177	0.0122	0.0103	0.0051	0.0076

Table 5.5: RMSE estimates of the unadjusted LPR estimator, the optimal jackknife estimator based on 2,3,4,6,8 non-overlapping (NO) sub-samples, the optimal jackknife estimator based on 2 moving block (MB) sub-samples, both versions of the GS estimator, and the prefiltered sieve bootstrap estimator, for the DGP: ARFIMA$(0, d_0, 1)$. The estimates are obtained under Gaussian and Student t_5 innovations, with $\alpha = 0.65$.

θ_0	d_0	n	$\widehat{d}_n$	$\widehat{d}_{J,2}^{Opt-NO}$	$\widehat{d}_{J,3}^{Opt-NO}$	$\widehat{d}_{J,4}^{Opt-NO}$	$\widehat{d}_{J,6}^{Opt-NO}$	$\widehat{d}_{J,8}^{Opt-NO}$	$\widehat{d}_{J,2}^{Opt-MB}$	$\widehat{d}_1^{GS}$	$\widehat{d}_1^{Opt-GS}$	$\widehat{d}^{PFSB}$
							Gaussian					
-0.9	-0.25	96	0.6233	0.6345	0.6275	0.6177	0.6112	**0.6020**	0.6284	0.6385	0.6086	0.8247
		576	0.4794	0.4812	0.4723	0.4662	0.4553	**0.4492**	0.4671	0.4885	0.4686	0.4977
	0	96	0.7361	0.8081	0.7972	0.7875	0.7726	0.7642	0.7815	0.8413	0.7214	0.8510
		576	0.5687	0.5919	0.5822	0.5719	0.5641	**0.5527**	0.5637	0.5838	0.5639	0.5942
	0.25	96	0.7996	0.8096	0.7918	0.7872	0.7716	**0.7615**	0.7715	0.8268	0.7869	0.8430
		576	0.5951	0.6193	0.6022	0.5976	0.5843	**0.5693**	0.5826	0.6219	0.6019	0.6590
	0.45	96	0.8219	0.8410	0.8325	0.8224	0.8135	**0.8064**	0.8231	0.8590	0.8190	0.8327
		576	0.5950	0.6066	5953	0.5871	0.5763	**0.5642**	0.5783	0.6298	0.6198	0.6487
-0.4	-0.25	96	0.2376	0.2253	0.2218	0.2198	0.2133	**0.2102**	0.2401	0.2488	0.2255	0.3103
		576	0.1037	0.0923	0.0895	0.0745	0.0672	**0.0652**	0.1052	0.1098	0.1004	0.1254
	0	96	0.2497	0.2385	0.2278	0.2142	0.2136	**0.2015**	0.2514	0.2559	0.2512	0.2883
		576	0.1070	0.0936	0.0979	0.0819	0.0887	**0.0778**	0.1078	0.1105	0.0845	0.1215
	0.25	96	0.2527	0.2451	0.2425	0.2379	0.2343	**0.2335**	0.2535	0.2560	0.2495	0.2782
		576	0.1068	0.0987	0.1052	0.1057	0.0964	**0.0867**	0.1067	0.1103	0.0934	0.1199
	0.45	96	0.2496	0.2524	0.2459	0.2476	0.2493	0.2495	0.2495	0.2518	**0.2441**	0.2725
		576	0.1047	0.0928	0.0900	0.0855	0.0830	**0.0740**	0.1040	0.1098	0.0991	0.1188
0.4	-0.25	96	0.1982	0.1894	0.1875	0.1825	0.1793	**0.1687**	0.1987	0.2212	0.2153	0.2809
		576	0.0932	0.0858	0.0988	0.0947	0.0935	**0.0933**	0.0933	0.1078	0.0812	0.1268
	0	96	0.1944	0.1826	0.1815	0.1729	0.1666	**0.1654**	0.1955	0.2203	0.2146	0.2701
		576	0.0919	0.0890	0.0893	0.0850	0.0829	**0.0824**	0.0924	0.1072	0.0930	0.1243
	0.25	96	0.1947	0.1945	0.1918	0.1878	0.1780	**0.1762**	0.1962	0.2213	0.2048	0.2663
		576	0.0925	0.0942	0.1079	0.0983	0.0942	**0.0832**	0.0932	0.1077	0.0924	0.1238
	0.45	96	0.1964	0.1769	0.1649	0.1544	**0.1407**	0.1483	0.1984	0.2223	0.2175	0.2643
		576	0.0943	0.0902	0.0831	0.0846	0.0772	**0.0756**	0.0955	0.1090	0.0939	0.1229
							Student t_5					
-0.9	-0.25	96	0.6316	0.6469	0.6328	0.6284	0.6117	**0.6045**	0.6286	0.6421	0.6236	0.6643
		576	0.4858	0.4911	0.4872	0.4769	0.4681	**0.4573**	0.4822	0.5262	0.5192	0.5985
	0	96	0.7387	0.7513	0.7404	0.7318	0.7264	**0.7128**	0.7391	0.7614	0.7162	0.7848
		576	0.5709	0.5950	0.5802	0.5741	0.5662	**0.5586**	0.5940	0.6045	0.5873	0.5838
	0.25	96	0.8053	0.8175	0.8026	0.7925	0.7816	**0.7726**	0.7921	0.8387	0.8297	0.8414
		576	0.5948	0.6390	0.6204	0.6349	0.6482	**0.6598**	0.6415	0.5124	0.5122	0.5694
	0.45	96	0.8249	0.8345	0.8237	0.8182	0.8034	**0.7925**	0.8164	0.8646	0.8040	0.8333
		576	0.5948	0.6060	0.5913	0.5872	0.5713	**0.5652**	0.5873	0.6112	0.5907	0.5639
-0.4	-0.25	96	0.2377	0.2447	0.2353	0.2216	0.2153	**0.2064**	0.2347	0.2484	0.2318	0.3067
		576	0.1036	0.1024	0.0982	0.0913	0.0826	**0.0762**	0.0856	0.1091	0.0992	0.1205
	0	96	0.2483	0.2371	0.2264	**0.2145**	0.2264	0.2375	0.2536	0.2549	0.2464	0.2892
		576	0.1064	0.1433	0.1375	0.1246	0.1162	0.1123	0.1348	0.1095	**0.0933**	0.1171
	0.25	96	0.2510	0.2529	0.2457	0.2364	0.2254	**0.2176**	0.2620	0.2543	0.2486	0.2779
		576	0.1063	0.1180	0.1103	0.1096	0.1002	**0.0927**	0.1126	0.1093	0.0945	0.1169
	0.45	96	0.2477	0.2511	0.2486	0.2401	0.2365	**0.2274**	0.2495	0.2499	0.2344	0.2652
		576	0.1047	0.1126	0.1082	0.1010	0.0985	**0.0919**	0.1123	0.1092	0.0920	0.1169
0.4	-0.25	96	0.1970	0.2275	0.2033	0.1972	0.1861	**0.1804**	0.2166	0.2202	0.2127	0.2208
		576	0.0927	0.1054	0.1002	0.0982	0.0935	**0.0876**	0.1011	0.1069	0.0922	0.1041
	0	96	0.1936	0.2096	0.2024	0.1975	0.1912	**0.1876**	0.2153	0.2193	0.2058	0.2110
		576	0.0918	0.1188	0.1113	0.1054	0.1069	**0.0984**	0.1068	0.1062	0.0915	0.1134
	0.25	96	0.1935	0.2235	0.2175	0.2141	0.2097	**0.1822**	0.1972	0.2196	0.1905	0.2126
		576	0.0920	0.1040	0.0973	0.0902	0.0846	**0.0824**	0.1066	0.1067	0.0913	1056
	0.45	96	0.1962	0.2168	0.2101	0.2046	0.1972	**0.1903**	0.1847	0.2211	0.1906	0.2176
		576	0.0943	0.1165	0.1112	0.1055	0.0946	**0.0812**	0.755	0.1084	0.0908	0.1154

following iterative method for obtaining a feasible version of the jackknife-based estimator.

An iterative version of the optimal jackknife estimator

1. **Prerequisite:** Estimate the short memory parameter, in either the ARFIMA$(1, d_0, 0)$ or ARFIMA$(0, d_0, 1)$ model, by estimating an AR(1) or MA(1) model (respectively) using pre-filtered data based on $d^f = \widehat{d}_n$.

2. **Initialization:** Set $k = 1$ and tolerance level $\tau = \tau^{(0)}$.

3. **Recursive step:** For the k^{th} recursion, perform the optimal jackknife bias-correction procedure of Section 5.3.2 with the estimates of the short memory parameters from step 1, and $d^f = \widehat{d}_n$, inserted into the formulae for the covariance terms in (5.15) and (5.16). Denote the resulting estimator by $\widehat{d}_{J,m}^{Opt(k)}$.

4. **Stopping rule:** If $\left|\widehat{d}_{J,m}^{Opt,(k+1)} - \widehat{d}_{J,m}^{Opt,(k)}\right| > \tau$ set $k = k + 1$ and $\tau = \tau^{(k)}$, and repeat steps 1 and 3 after updating $d^f = \widehat{d}_{J,m}^{Opt,(k)}$.

The basic idea behind the algorithm is as follows: estimation of the short memory parameter requires pre-filtering via some preliminary estimate of d_0. An obvious initial (consistent) choice is $d^f = \widehat{d}_n$; however $\widehat{d}_n$ is known to be biased in finite samples. Hence, iteration of the above algorithm, which involves replacing the initial pre-filtering value with successively less biased values, $d^f = \widehat{d}_{J,m}^{opt,(k)}$, is expected to yield a final feasible version of the jackknife estimator, $\widehat{d}_{J,m}^{Opt,(k+1)}$, based on accurate estimates of all unknown parameters. (See also Poskitt *et al.*, 2016 for a related application of this form of iterative procedure). The feasible version of the jackknife statistic is denoted hereafter by $\widehat{d}_{J,m}^{NO}$ if the sub-sampling method is non-overlapping and $\widehat{d}_{J,m}^{MB}$ if the sub-sampling method is moving-block.

Tables 5.6 and 5.7 display the bias and RMSE results of the feasible jackknife estimator, the feasible GS estimator, $\widehat{d}_1^{GS}$, and the PFSB estimator, for the ARFIMA$(1, d_0, 0)$ process. The corresponding results for the ARFIMA$(0, d_0, 1)$ process are presented in Table 5.8 and 5.9. Once again, the two panels in each table record the results for the two different error processes, and the minimum bias and RMSE is shown in bold font.

Consider the results for the ARFIMA$(1, d_0, 0)$ process. The (various versions of the) feasible jackknife estimators show similar characteristics to the corresponding optimal estimators, except for exhibiting larger bias and RMSE. This is to be expected given that the optimal weights are now functions of estimates of both d and the autoregressive coefficient. The increase in bias (relative to the known parameter case) is particularly marked when $\phi_0 = -0.9$, with the feasible jackknife estimators seen to be more biased overall than the raw LPR estimator itself, even for the larger sample size. However, for $\phi_0 = -0.4$ and 0.4, the feasible jackknife estimators still often show reduction in bias compared to the LPR estimator, especially for the smaller values of m. For example, when $\phi_0 = -0.4$ and $n = 96$, the bias reduction of $\widehat{d}_{J,m}^{NO}$ with $m = 2$ compared to the raw LPR estimator is up to 26% and when $n = 576$, the bias reduction rises to 62%. Overall, however, the estimators with the least bias are the feasible GS estimator and the PFBS estimator, where, as noted earlier, the latter does not depend on knowledge of the true DGP.

The RMSE results in Table 5.7 confirm the consistency of the feasible jackknife estimators. However, neither the feasible jackknife estimators, nor the alternative bias-adjusted methods, now out-perform the raw LPR estimator in terms of RMSE. The feasible $\widehat{d}_{J,m}^{NO}$ with $m = 8$ and $\widehat{d}_1^{GS}$ compete for second place in terms of RMSE, with the feasible jackknife estimator preferable overall, in particular when one considers the results in the lower panel of Table 5.7.

The results in Tables 5.8 and 5.9, for the ARFIMA$(0, d_0, 1)$ process, tell a very similar story to those for the ARFIMA$(1, d_0, 0)$ case.

5.6 Conclusion

With the fractionally integrated autoregressive moving-average model being one of the key model classes for describing long memory processes, much effort has been expended on producing accurate estimates of the fractional differencing parameter, d, in particular. This quest has been hampered by certain problems, for both parametric and semi-parametric approaches. Specifically, the need to fully specify the model for parametric estimation means that any incorrect specification of the short memory dynamics has serious consequences, in terms of both finite sample and asymptotic properties (see, for example, Chen and Deo, 2006 and Chapters 3 and 4 of this thesis). On the other hand, the semi-parametric estimators, whilst not requiring explicit modelling of the short memory component, can suffer substantial finite sample bias in the presence of unaccounted for short memory dynamics. It is bias correction of this latter class of estimator that has been the focus of this chapter.

A natural way of producing a bias-corrected version of the commonly used the log-periodogram regression [LPR] estimator is suggested in this chapter, based on the jackknife technique. Optimality is achieved by allocating weights within the jackknife that are adjusted for the bias to a particular order, and that minimize the increase in variance caused by the reduction in bias. The construction of the optimally bias-corrected estimator requires expressions for the dominant bias term and variance of the unadjusted LPR estimator. We show that the statistical properties of the LPR estimator, as originally established by Hurvich *et al.* (1998), are valid for a more general class of fractional process that is not necessarily Gaussian. Hence, the jackknife

Table 5.6: Bias estimates of the unadjusted LPR estimator, the optimal jackknife estimator based on 2,3,4,6,8 non-overlapping (NO) sub-samples, the feasible jackknife estimator based on 2 moving block (MB) sub-samples, both versions of the GS estimator, and the prefiltered sieve bootstrap estimator, for the DGP: ARFIMA$(1, d_0, 0)$. The estimates are obtained under Gaussian and Student t_5 innovations, with $\alpha = 0.65$.

ϕ_0	d_0	n	$\widehat{d}_n$	$\widehat{d}_{J,2}^{NO}$	$\widehat{d}_{J,3}^{NO}$	$\widehat{d}_{J,4}^{NO}$	$\widehat{d}_{J,6}^{NO}$	$\widehat{d}_{J,8}^{NO}$	$\widehat{d}_{J,2}^{MB}$	$\widehat{d}_1^{GS}$	$\widehat{d}^{PFSB}$
						Gaussian					
-0.9	-0.25	96	0.8145	0.8456	0.8514	0.8612	0.8523	0.8669	0.8351	0.8002	**0.7908**
		576	0.5945	0.6076	0.5982	0.5816	0.5943	0.6166	0.6057	**0.5724**	0.5898
	0	96	0.8053	0.8532	0.8421	0.8337	0.8214	0.8377	0.8269	0.8015	**0.7955**
		576	0.5912	0.6634	0.6517	0.6428	0.6363	0.6278	0.6379	**0.5761**	0.5888
	0.25	96	0.7752	0.7982	0.7843	0.7716	0.7886	0.8130	0.7975	**0.7673**	0.7685
		576	0.5883	0.6062	0.5996	0.5904	0.5855	0.5963	0.5846	0.5716	**0.5638**
	0.45	96	0.7006	0.7236	0.7173	0.7003	0.7394	0.7226	0.7139	0.6946	**0.6705**
		576	0.5748	0.5994	0.5912	0.5845	0.5748	0.5830	0.5759	0.5659	**0.5451**
-0.4	-0.25	96	0.1756	0.1687	0.1699	0.1700	0.1791	0.1866	0.1630	**0.1367**	0.1435
		576	0.0607	**0.0226**	0.0389	0.0497	0.0886	0.0664	0.0442	0.0304	0.0286
	0	96	0.1653	0.1367	0.1388	0.1442	0.1641	0.1776	0.1542	**0.1304**	0.1353
		576	0.0560	0.0355	0.0462	0.0586	0.0641	0.0663	0.0432	0.0264	**0.0249**
	0.25	96	0.1629	**0.1223**	0.1374	0.1442	0.1594	0.1777	0.1302	0.1329	0.1294
		576	0.0571	0.0370	0.0446	0.0581	0.0665	0.0718	0.0660	0.0289	**0.0251**
	0.45	96	0.1653	**0.1233**	0.1395	0.1468	0.1699	0.1881	0.1730	0.1400	0.1277
		576	0.0625	0.0421	0.0562	0.0664	0.0782	0.0882	0.0728	0.0359	**0.0261**
0.4	-0.25	96	-0.0363	-0.0221	-0.0348	-0.0461	-0.0594	-0.0667	-0.0416	**-0.0047**	-0.0147
		576	-0.0056	-0.0106	-0.0064	-0.0073	-0.0097	-0.0102	-0.0060	0.0056	**-0.0004**
	0	96	-0.0534	-0.0316	-0.0215	-0.0113	-0.0297	-0.0419	-0.0435	**-0.0089**	-0.0175
		576	-0.0125	-0.0030	-0.0052	-0.0065	-0.0078	-0.0086	-0.0076	**-0.0008**	-0.0040
	0.25	96	-0.0559	-0.0201	-0.0220	-0.0292	-0.0340	-0.0414	-0.0420	**-0.0068**	-0.0153
		576	-0.0115	-0.0024	-0.0043	-0.0052	-0.0084	-0.0121	-0.0070	**0.0017**	-0.0027
	0.45	96	-0.0501	-0.0111	-0.0129	-0.0210	-0.0337	-0.0549	-0.0185	**0.0032**	-0.0111
		576	-0.0058	-0.0018	-0.0026	-0.0045	-0.0069	-0.0095	-0.0056	0.0089	**0.0004**
						Student t_5					
-0.9	-0.25	96	0.8123	0.8045	0.8164	0.8203	0.8272	0.8300	0.8135	0.7914	**0.7847**
		576	0.5952	0.5861	0.5912	0.5985	0.6015	0.6098	0.5861	0.5863	**0.5770**
	0	96	0.8034	0.8026	0.8176	0.8219	0.8283	0.8311	0.8042	0.7843	**0.7830**
		576	0.5915	0.6135	0.6294	0.6347	0.6428	0.6483	0.6254	0.5642	**0.5539**
	0.25	96	0.7726	0.7992	0.8034	0.8088	0.8126	0.8195	0.7938	0.7633	**0.7572**
		576	0.5883	0.6172	0.6221	0.6279	0.6334	0.6386	0.6154	0.5633	**0.5472**
	0.45	96	0.7002	0.6997	0.7042	0.7088	0.7126	0.7184	0.6955	0.6849	**0.6731**
		576	0.5758	0.5724	0.5846	0.5875	0.5901	0.5978	0.5849	0.5602	**0.5514**
-0.4	-0.25	96	0.1764	0.1454	0.1590	0.1627	0.1796	0.1879	0.1662	**0.1371**	0.1422
		576	0.0611	**0.0168**	0.0315	0.0432	0.0469	0.0620	0.0524	0.0305	0.0289
	0	96	0.1662	0.1379	0.1408	0.1485	0.1532	0.1658	0.1423	**0.1307**	0.1340
		576	0.0565	0.0365	0.0493	0.0522	0.0598	0.0673	0.0474	0.0266	**0.0252**
	0.25	96	0.1640	**0.1255**	0.1397	0.1462	0.1613	0.1731	0.1416	0.1334	0.1282
		576	0.0575	**0.0246**	0.0366	0.0429	0.0557	0.0634	0.0329	0.0292	0.0254
	0.45	96	0.1666	**0.1261**	0.1430	0.1532	0.1638	0.1721	0.1562	0.1405	0.1270
		576	0.0627	0.0385	0.0468	0.0554	0.0622	0.0594	0.0667	**0.0359**	0.0627
0.4	-0.25	96	-0.0357	-0.0246	-0.0365	-0.0413	-0.0522	-0.0567	-0.0345	**-0.0054**	-0.0132
		576	-0.0052	-0.0066	-0.0054	-0.0062	-0.0076	-0.0092	-0.0076	-0.0054	**0.0003**
	0	96	-0.0525	-0.0223	-0.0268	-0.0315	-0.0386	-0.0412	-0.0336	**-0.0081**	-0.0164
		576	-0.0121	-0.0040	-0.0055	-0.0078	-0.0081	-0.0089	-0.0062	**-0.0006**	-0.0033
	0.25	96	-0.0641	-0.0112	-0.0167	-0.0253	-0.0342	-0.0410	-0.0391	**-0.0062**	-0.0165
		576	-0.0182	-0.0026	-0.0049	-0.0058	-0.0076	-0.0083	-0.0070	**-0.0019**	-0.0045
	0.45	96	-0.0489	-0.0210	-0.0130	-0.0222	-0.0312	-0.0423	-0.0193	**-0.0040**	-0.0097
		576	-0.0055	-0.0016	-0.0022	-0.0044	-0.0062	-0.0082	-0.0044	-0.0087	**0.0008**

Table 5.7: RMSE estimates of the unadjusted LPR estimator, the optimal jackknife estimator based on 2,3,4,6,8 non-overlapping (NO) sub-samples, the feasible jackknife estimator based on 2 moving block (MB) sub-samples, both versions of the GS estimator, and the prefiltered sieve bootstrap estimator, for the DGP: ARFIMA$(1, d_0, 0)$. The estimates are obtained under Gaussian and Student t_5 innovations, with $\alpha = 0.65$.

ϕ_0	d_0	n	$\widehat{d}_n$	$\widehat{d}_{J,2}^{NO}$	$\widehat{d}_{J,3}^{NO}$	$\widehat{d}_{J,4}^{NO}$	$\widehat{d}_{J,6}^{NO}$	$\widehat{d}_{J,8}^{NO}$	$\widehat{d}_{J,2}^{MB}$	$\widehat{d}_1^{GS}$	$\widehat{d}^{PFSB}$
						Gaussian					
-0.9	-0.25	96	**1.0359**	1.2543	1.2498	1.2350	1.2201	1.2101	1.2345	1.3386	1.2885
		576	0.7398	0.7892	0.7804	0.7762	0.7683	0.7616	0.7761	0.7371	**0.7359**
	0	96	**1.1148**	1.1620	1.1542	1.1522	1.1436	1.1344	1.1543	1.1819	1.2167
		576	0.8288	0.8642	0.8849	0.8724	0.8613	0.8541	0.8595	0.8173	**0.8053**
	0.25	96	**1.1618**	1.2041	1.1933	1.1866	1.1727	1.1649	1.1867	1.1484	1.2299
		576	**0.9175**	0.9668	0.9537	0.9489	0.9422	0.9338	0.9518	1.1171	1.1130
	0.45	96	**1.1286**	1.2301	1.2286	1.2234	1.2154	1.2034	1.2351	1.4331	1.5385
		576	**0.9708**	1.0049	0.9972	0.9936	0.9861	0.9805	0.9952	1.1124	1.1647
-0.4	-0.25	96	**0.2568**	0.2928	0.2845	0.2777	0.2622	0.2581	0.2749	0.2594	0.3028
		576	**0.1098**	0.1368	0.1213	0.1195	0.1269	0.1371	0.1262	0.1118	0.1272
	0	96	**0.2498**	0.2836	0.2792	0.2713	0.2648	0.2589	0.2711	0.2560	0.2930
		576	**0.1069**	0.1353	0.1276	0.1194	0.1118	0.1182	0.1212	0.1104	0.1247
	0.25	96	**0.2490**	0.2926	0.2881	0.2764	0.2621	0.2515	0.3467	0.2580	0.2879
		576	**0.1079**	0.1442	0.1367	0.1210	0.1175	0.1116	0.1226	0.1115	0.1239
	0.45	96	**0.2506**	0.2992	0.2842	0.2761	0.2682	0.2605	0.2835	0.2616	0.2506
		576	**0.1115**	0.1511	0.1475	0.1389	0.1203	0.1147	0.1385	0.1143	0.1230
0.4	-0.25	96	**0.1917**	0.2454	0.2420	0.2346	0.2237	0.2276	0.2374	0.2212	0.2717
		576	**0.0919**	0.1296	0.1216	0.1191	0.1122	0.1055	0.1167	0.1081	0.1198
	0	96	**0.1946**	0.2369	0.2318	0.2216	0.2134	0.2083	0.2266	0.2203	0.2546
		576	**0.0920**	0.1327	0.1256	0.1227	0.1188	0.1112	0.1283	0.1073	0.1166
	0.25	96	**0.1960**	0.2338	0.2267	0.2395	0.2469	0.2302	0.2347	0.2209	0.2482
		576	**0.0922**	0.1219	0.1193	0.1104	0.1086	0.1025	0.1134	0.1076	0.1158
	0.45	96	**0.1955**	0.2441	0.2367	0.2248	0.2334	0.2240	0.2267	0.2218	0.2453
		576	**0.0926**	0.1357	0.1302	0.1213	0.1185	0.1065	0.1126	0.1089	0.1149
						Student t_5					
-0.9	-0.25	96	**1.0321**	1.2154	1.2036	1.1942	1.1833	1.1795	1.1836	1.1741	1.0570
		576	0.7408	0.7882	0.7815	0.7764	0.7703	0.7681	0.7792	0.7406	**0.7309**
	0	96	**1.1120**	1.1953	1.1842	1.1765	1.1681	1.1586	1.1688	1.2792	1.2542
		576	**0.8291**	0.8642	0.8571	0.8516	0.8486	0.8421	0.8436	0.8484	0.8367
	0.25	96	**1.1577**	1.1985	1.1876	1.1772	1.1626	1.1566	1.1833	1.2620	1.2967
		576	0.9173	0.9848	0.9758	0.9705	0.9671	0.9611	0.9637	0.9174	**0.9133**
	0.45	96	**1.1272**	1.1973	1.1862	1.1767	1.1706	1.1682	1.1791	1.2314	1.2848
		576	0.9720	1.0682	1.0197	0.9982	0.9844	0.9752	0.9869	**0.9643**	0.9755
-0.4	-0.25	96	**0.2562**	0.2997	0.2902	0.2883	0.2791	0.2656	0.2884	0.2587	0.3008
		576	**0.1096**	0.1385	0.1275	0.1243	0.1193	0.1150	0.1205	0.1109	0.1264
	0	96	**0.2492**	0.2879	0.2800	0.2795	0.2712	0.2631	0.2788	0.2552	0.2912
		576	**0.1069**	0.1370	0.1313	0.1295	0.1203	0.1151	0.1213	0.1095	0.1241
	0.25	96	**0.2487**	0.2823	0.2779	0.2723	0.2667	0.2545	0.2864	0.2567	0.2865
		576	**0.1078**	0.1388	0.1299	0.1215	0.1196	0.1117	0.1387	0.1106	0.1233
	0.45	96	**0.2509**	0.2901	0.2811	0.2729	0.2645	0.2574	0.2665	0.2610	0.2881
		576	**0.1115**	0.1391	0.1300	0.1226	0.1163	0.1125	0.1222	0.1137	0.1228
0.4	-0.25	96	**0.1907**	0.2326	0.2295	0.2206	0.2157	0.2078	0.2276	0.2202	0.2698
		576	**0.0915**	0.1151	0.1108	0.1097	0.1021	0.0982	0.1204	0.1076	0.1190
	0	96	**0.1930**	0.2289	0.2195	0.2142	0.2064	0.2000	0.2224	0.2181	0.2532
		576	**0.0915**	0.1274	0.1205	0.1134	0.1092	0.1001	0.1296	0.1076	0.1190
	0.25	96	**0.1977**	0.2316	0.2288	0.2234	0.2128	0.2071	0.2264	0.2193	0.2361
		576	**0.0927**	0.1210	0.1186	0.1138	0.1088	0.1029	0.1223	0.1072	0.1216
	0.45	96	**0.1942**	0.2224	0.2241	0.2363	0.2104	0.2032	0.2345	0.2201	0.2440
		576	**0.0924**	0.1284	0.1205	0.1154	0.1062	0.0990	0.1086	0.1082	0.1142

Table 5.8: Bias estimates of the unadjusted LPR estimator, the optimal jackknife estimator based on 2,3,4,6,8 non-overlapping (NO) sub-samples, the feasible jackknife estimator based on 2 moving block (MB) sub-samples, both versions of the GS estimator, and the prefiltered sieve bootstrap estimator, for the DGP: ARFIMA$(0, d_0, 1)$. The estimates are obtained under Gaussian and Student t_5 innovations, with $\alpha = 0.65$.

θ_0	d_0	n	$\widehat{d}_n$	$\widehat{d}_{J,2}^{NO}$	$\widehat{d}_{J,3}^{NO}$	$\widehat{d}_{J,4}^{NO}$	$\widehat{d}_{J,6}^{NO}$	$\widehat{d}_{J,8}^{NO}$	$\widehat{d}_{J,2}^{MB}$	$\widehat{d}_1^{GS}$	$\widehat{d}^{PFSB}$
						Gaussian					
-0.9	-0.25	96	-0.5671	-0.5761	-0.5622	-0.5690	-0.5781	-0.5833	-0.5862	**-0.5450**	-0.5466
		576	-0.4527	-0.4582	-0.4682	-0.4700	-0.4765	-0.4839	-0.4728	-0.4385	**-0.4285**
	0	96	-0.7042	-0.6892	-0.6921	-0.7070	-0.7158	-0.7249	-0.7037	**-0.6575**	-0.6664
		576	-0.5594	-0.5568	-0.5612	-0.5789	-0.5815	-0.5887	-0.5716	**-0.5256**	-0.5375
	0.25	96	-0.7763	-0.7624	-0.7716	-0.7789	-0.7826	-0.7899	-0.7724	**-0.7524**	-0.7661
		576	-0.5880	-0.5641	-0.5760	-0.5827	-0.5873	-0.5915	-0.5832	**-0.5473**	-0.5621
	0.45	96	-0.8004	-0.7862	-0.7924	-0.8005	-0.8062	-0.8142	-0.8136	**-0.7600**	-0.7854
		576	-0.5880	-0.5536	-0.5676	-0.5788	-0.5801	-0.5875	-0.5543	**-0.5351**	-0.5527
-0.4	-0.25	96	-0.1437	-0.1340	-0.1448	-0.1308	-0.1472	-0.1581	-0.1471	**-0.1120**	-0.1240
		576	-0.0476	-0.0430	-0.0367	-0.0226	-0.0399	-0.0446	-0.0517	**-0.0187**	-0.0271
	0	96	-0.1653	-0.1375	-0.1464	-0.1571	-0.1528	-0.1670	-0.1523	-0.1305	**-0.1248**
		576	-0.0560	-0.0315	-0.0416	-0.0552	-0.0681	-0.0681	-0.0403	**-0.0265**	-0.0307
	0.25	96	-0.1692	-0.1342	-0.1516	-0.1615	-0.1500	-0.1672	-0.1620	-0.1297	**-0.1200**
		576	-0.0552	**-0.0221**	-0.0436	-0.0566	-0.0622	-0.0685	-0.0558	-0.0243	-0.0287
	0.45	96	-0.1630	**-0.0924**	-0.1448	-0.1755	-0.1836	-0.1977	-0.1536	-0.1190	-0.1118
		576	-0.0493	-0.0234	-0.0341	-0.0456	-0.0516	-0.0578	-0.0427	**-0.0169**	-0.0244
0.4	-0.25	96	0.0637	**0.0105**	0.0564	0.0692	0.0778	0.0783	0.0546	0.0154	0.0651
		576	0.0175	0.0162	0.0186	0.0201	0.0246	0.0183	0.0154	**0.0049**	0.0132
	0	96	0.0525	0.0468	0.0487	0.0515	0.0432	0.0469	0.0441	**0.0081**	0.0603
		576	0.0125	0.0220	0.0325	0.0392	0.0387	0.0326	0.0156	**0.0006**	0.0100
	0.25	96	0.0504	0.0421	0.0516	0.0674	0.0692	0.0726	0.0432	**0.0110**	0.0574
		576	0.0136	0.0082	0.0096	0.0166	0.0189	0.0260	0.0086	**0.0031**	0.0108
	0.45	96	0.0549	0.0416	0.0497	0.0553	0.0762	0.0617	0.0497	**0.0204**	0.0570
		576	0.0192	0.0098	0.0100	**0.0085**	0.0101	0.0168	0.0176	0.0103	0.0132
						Student t_5					
-0.7	-0.25	96	-0.5754	-0.5513	-0.5624	-0.5682	-0.5705	-0.5782	-0.5681	**-0.5479**	-0.5553
		576	-0.4589	-0.4262	-0.4351	-0.4482	-0.4506	-0.4570	-0.4432	-0.4275	**-0.4103**
	0	96	-0.7073	-0.6612	-0.6748	-0.6792	-0.6814	-0.6865	-0.6791	**-0.6427**	-0.6638
		576	-0.5613	-0.5523	-0.5681	-0.5703	-0.5783	-0.5816	-0.5671	**-0.5366**	-0.5570
	0.25	96	-0.7814	-0.7542	-0.7695	-0.7715	-0.7762	-0.7855	-0.7642	**-0.7373**	-0.7477
		576	-0.5876	-0.5641	-0.5706	-0.5738	-0.5869	-0.5901	-0.5712	**-0.5478**	-0.5532
	0.45	96	-0.8032	-0.7878	-0.7927	-0.7994	-0.8025	-0.8080	-0.7923	**-0.6661**	-0.7880
		576	-0.5875	-0.5439	-0.5483	-0.5529	-0.5587	-0.5613	-0.5624	**-0.5364**	-0.5438
-0.4	-0.25	96	-0.1442	-0.1302	-0.1398	-0.1482	-0.1546	-0.1673	-0.1585	**-0.1117**	-0.1224
		576	-0.0477	-0.0515	-0.0382	-0.0475	-0.0538	-0.0661	-0.0500	**-0.0187**	-0.0208
	0	96	-0.1646	-0.1483	-0.1390	-0.1441	-0.1538	-0.1639	-0.1666	-0.1299	**-0.1259**
		576	-0.0559	-0.0574	-0.0490	-0.0391	-0.0420	-0.0502	-0.0592	-0.0265	**-0.0264**
	0.25	96	-0.1686	-0.1378	-0.1492	-0.1538	-0.1635	-0.1740	-0.1632	-0.1290	**-0.1211**
		576	-0.0548	-0.0274	-0.0394	-0.0437	-0.0583	-0.0503	-0.0434	**-0.0242**	-0.0248
	0.45	96	-0.1621	**-0.0782**	-0.0845	-0.0957	-0.0975	-0.1016	-0.0982	-0.1183	-0.1071
		576	-0.0492	-0.0229	-0.0384	-0.0493	-0.0528	-0.0663	-0.0376	**-0.0169**	-0.0209
0.4	-0.25	96	0.0648	**0.0090**	0.0128	0.0213	0.0346	0.0427	0.0428	0.0159	0.0187
		576	0.0179	0.0118	0.0194	0.0249	0.0358	0.0442	0.0250	**0.0051**	0.0074
	0	96	0.0529	0.0429	0.0556	0.0694	0.0624	0.0619	0.0582	**0.0084**	0.0145
		576	0.0122	0.0218	0.0104	0.0059	0.0138	0.0195	0.0258	**0.0008**	0.0038
	0.25	96	0.0505	0.0347	0.0247	0.0285	0.0342	0.0445	0.0476	**0.0116**	0.0151
		576	0.0140	0.0065	0.0100	0.0148	0.0196	0.0204	0.0114	**0.0033**	0.0053
	0.45	96	0.0561	0.0313	0.0378	0.0435	0.0527	0.0515	0.0420	0.0209	**0.0187**
		576	0.0194	0.0099	0.0120	0.0148	0.0179	0.0192	0.0146	0.0103	**0.0076**

Table 5.9: RMSE estimates of the unadjusted LPR estimator, the optimal jackknife estimator based on 2,3,4,6,8 non-overlapping (NO) sub-samples, the feasible jackknife estimator based on 2 moving block (MB) sub-samples, both versions of the GS estimator, and the prefiltered sieve bootstrap estimator, for the DGP: ARFIMA$(0, d_0, 1)$. The estimates are obtained under Gaussian and Student t_5 innovations, with $\alpha = 0.65$.

θ_0	d_0	n	$\widehat{d}_n$	$\widehat{d}^{NO}_{J,2}$	$\widehat{d}^{NO}_{J,3}$	$\widehat{d}^{NO}_{J,4}$	$\widehat{d}^{NO}_{J,6}$	$\widehat{d}^{NO}_{J,8}$	$\widehat{d}^{MB}_{J,2}$	$\widehat{d}^{GS}_1$	$\widehat{d}^{PFSB}$
						Gaussian					
-0.9	-0.25	96	**0.6233**	0.6678	0.6607	0.6582	0.6523	0.6492	0.6725	0.6385	0.8247
		576	**0.4794**	0.5124	0.5052	0.5009	0.4942	0.4872	0.4832	0.4885	0.4977
	0	96	**0.7361**	0.8599	0.8537	0.8462	0.8429	0.8369	0.8261	0.8413	0.8510
		576	**0.5687**	0.6421	0.6318	0.6288	0.6281	0.6215	0.6342	0.5838	0.5942
	0.25	96	**0.7996**	0.8516	0.8439	0.8384	0.8342	0.8268	0.8314	0.8268	0.8430
		576	**0.5951**	0.6482	0.6382	0.6315	0.6294	0.6240	0.6344	0.6219	0.6590
	0.45	96	**0.8219**	0.8729	0.8647	0.8605	0.8542	0.8348	0.8426	0.8590	0.8327
		576	**0.5950**	0.6384	0.6279	0.6211	0.6184	0.6124	0.6589	0.6298	0.6487
-0.4	-0.25	96	**0.2376**	0.2775	0.2658	0.2589	0.2532	0.2487	0.2799	0.2488	0.3103
		576	**0.1037**	0.1412	0.1357	0.1324	0.1245	0.1175	0.1345	0.1098	0.1254
	0	96	**0.2497**	0.2826	0.2748	0.2687	0.2550	0.2563	0.2659	0.2559	0.2883
		576	**0.1070**	0.1474	0.1394	0.1264	0.1235	0.1183	0.1264	0.1105	0.1215
	0.25	96	**0.2527**	0.2815	0.2727	0.2649	0.2580	0.2626	0.2793	0.2560	0.2782
		576	**0.1068**	0.1473	0.1385	0.1264	0.1148	0.1262	0.1374	0.1103	0.1199
	0.45	96	**0.2496**	0.2873	0.2838	0.2758	0.2699	0.2538	0.2638	0.2518	0.2725
		576	**0.1047**	0.1492	0.1409	0.1394	0.1336	0.1294	0.1365	0.1098	0.1188
0.4	-0.25	96	**0.1982**	0.2568	0.2484	0.2369	0.2237	0.2125	0.2398	0.2212	0.2809
		576	**0.0932**	0.1104	0.1227	0.1356	0.1256	0.1135	0.1036	0.1078	0.1268
	0	96	**0.1944**	0.2479	0.2385	0.2353	0.2236	0.2173	0.2264	0.2203	0.2701
		576	**0.0919**	0.1290	0.1184	0.1135	0.1048	0.1026	0.1175	0.1072	0.1243
	0.25	96	**0.1947**	0.2363	0.2205	0.2137	0.2039	0.2058	0.2374	0.2213	0.2663
		576	**0.0925**	0.1135	0.1175	0.1210	0.1186	0.1074	0.1283	0.1077	0.1238
	0.45	96	**0.1964**	0.2336	0.2288	0.2176	0.2038	0.2001	0.2375	0.2223	0.2643
		576	**0.0943**	0.1235	0.1163	0.1135	0.1073	0.1056	0.1248	0.1090	0.1229
						Student t_5					
-0.9	-0.25	96	**0.6316**	0.6813	0.6806	0.6764	0.6662	0.6512	0.6641	0.6421	0.6643
		576	**0.4858**	0.5364	0.5284	0.5243	0.5190	0.5103	0.5638	0.5262	0.5985
	0	96	**0.7387**	0.7924	0.7869	0.7812	0.7729	0.7648	0.7826	0.7614	0.7848
		576	**0.5709**	0.6363	0.6345	0.6284	0.6207	0.6183	0.6381	0.6045	0.5838
	0.25	96	**0.8053**	0.8469	0.8438	0.8376	0.8264	0.8175	0.8515	0.8387	0.8414
		576	0.5948	0.6684	0.6574	0.6543	0.6428	0.6348	0.6719	**0.5124**	0.5694
	0.45	96	**0.8249**	0.8694	0.8649	0.8573	0.8516	0.8448	0.8910	0.8646	0.8333
		576	0.5948	0.6435	0.6523	0.6428	0.6347	0.6255	0.6452	0.6112	**0.5639**
-0.4	-0.25	96	**0.2377**	0.2816	0.2737	0.2684	0.2541	0.2453	0.2664	0.2484	0.3067
		576	**0.1036**	0.1478	0.1396	0.1336	0.1293	0.1136	0.1242	0.1091	0.1205
	0	96	**0.2483**	0.2855	0.2739	0.2649	0.2563	0.2543	0.2536	0.2549	0.2892
		576	**0.1064**	0.1544	0.1456	0.1384	0.1325	0.1204	0.1383	0.1095	0.1171
	0.25	96	**0.2510**	0.2835	0.2739	0.2690	0.2655	0.2603	0.2532	0.2543	0.2779
		576	**0.1063**	0.1474	0.1424	0.1400	0.1365	0.1249	0.1250	0.1093	0.1169
	0.45	96	**0.2477**	0.2863	0.2748	0.2651	0.2677	0.2546	0.2503	0.2499	0.2652
		576	**0.1047**	0.1468	0.1385	0.1305	0.1235	0.1138	0.1247	0.1092	0.1169
0.4	-0.25	96	**0.1970**	0.2338	0.2304	0.2246	0.2144	0.2083	0.2162	0.2202	0.2208
		576	**0.0927**	0.1146	0.1112	0.1030	0.1073	0.1058	0.1025	0.1069	0.1041
	0	96	**0.1936**	0.2275	0.2195	0.2004	0.1945	0.2006	0.2144	0.2193	0.2110
		576	**0.0918**	0.1192	0.1136	0.1094	0.1013	0.0963	0.1040	0.1062	0.1134
	0.25	96	**0.1935**	0.2228	0.2169	0.2127	0.2004	0.1947	0.2020	0.2196	0.2126
		576	**0.0920**	0.1214	0.1185	0.1146	0.1090	0.0993	0.1053	0.1067	1056
	0.45	96	**0.1962**	0.2266	0.2174	0.2038	0.2095	0.2012	0.2120	0.2211	0.2176
		576	**0.0943**	0.1246	0.1213	0.1146	0.1053	0.1095	0.1183	0.1084	0.1154

estimator that we construct from the optimally weighted average of LPR estimators also has proven optimality under this general form of process. In addition to proving the consistency of the optimal jackknife estimator, we have the important result that the asymptotic variance of the estimator is equivalent to that of the unadjusted LPR estimator. That is, bias-adjustment is effected without any associated increase in asymptotic variance.

Our Monte Carlo study shows that, overall, the optimal jackknife estimator based on a small number of non-overlapping sub-samples, outperforms both the pre-filtered sieve bootstrap estimator of Poskitt *et al.* (2016) and the weighted-average estimator of Guggenberger and Sun (2006), albeit in the somewhat artificial case in which the parameters of the DGP are correctly identified and known, for the purpose of computing optimal weights. In the realistic case in which these parameters are not known, we suggest an iterative procedure in which the weights are constructed using consistent estimates. In this case the method is not dominant overall, compared to alternative bias-corrected methods, but is still the least biased in some cases, in particular when the true short memory dynamics are not too severe.

Throughout the chapter we assume that the number of sub-samples is fixed. One may wish to allow the number of sub-samples to vary and explore the characteristics of the resultant bias-adjusted estimators in this case. Importantly, alternative methods of estimating the weights are to be investigated, including the possible use of a non-parametric estimate of the spectral density (see, Moulines and Soulier, 1999), rather than replacing the true values with their consistent estimates, or the use of an adaptive method in the spirit of that suggested by Guggenberger and Sun. We also intend to explore the impact of model mis-specification on the computation of the optional weights.

Finally, although we focus on the LPR estimator, the jackknife procedure can easily be applied to other estimators such as the local Whittle estimator of Künsch (1987), the local polynomial Whittle estimator of Andrews and Sun (2004) or even to the (already analytically) bias-reduced estimators of Andrews and Guggenberger (2003) and Guggenberger and Sun (2006). Another possible extension is to relax the assumption of stationarity of the process using the results Velasco (1999b), and to derive the properties the optimal jackknife estimators in the non-stationary setting.

5.A Appendix: Proofs of Theorems and Lemmas

Proof of Theorem 5.1. Under Assumptions $(A.1)-(A.4)$, the proof of the theorem follows immediately after applying the results of Corollary 3.1 of Chapter 3 to Lemmas, 2, 5, 6 and 7 of Hurvich *et al.* (1998). Hence we omit the proof. ■

Prior to providing the proofs of the other theorems and lemmas, we will introduce the following definition, and its properties, to be used hereinafter.

Define $\Delta^{(T)}(\lambda)=\sum_{t=1}^{T}\exp(-\imath\lambda t)$. Then,

$$\begin{aligned}\Delta^{(T)}(\lambda) &= \exp\left(-\imath\frac{\lambda}{2}(T+1)\right)\frac{\sin\left(\frac{\lambda T}{2}\right)}{\sin\left(\frac{\lambda}{2}\right)} \\ &= \left\{\begin{array}{lll} 0 & if & \lambda\not\equiv 0\,(\mathrm{mod}\ \pi) \\ T & if & \lambda\equiv 0\,(\mathrm{mod}\ 2\pi) \\ 0\ or\ T & if & \lambda=\pm\pi,\pm 3\pi,\ldots \end{array}\right. ,\end{aligned} \tag{5.32}$$

where, $a\equiv b\,(\mathrm{mod}\ \alpha)$ means that the difference $(a-b)$ is an integral multiple of α for $\alpha,x,y\in\mathbb{R}$.

Consider

$$\sum_{t=-T}^{T}\exp\{-\imath\lambda t\}=1+\sum_{t=1}^{T}\exp\{-\imath\lambda t\}+\sum_{t=1}^{T}\exp\{-\imath(-\lambda)t\}=1+2\Delta^{(T)}(\lambda),\ \text{using (5.32).}$$

This immediately gives that

$$\lim_{T\to\infty} \frac{1}{2\pi} \sum_{t=-T}^{T} \exp\left\{-\imath\lambda t\right\} = \eta\left(\lambda\right). \tag{5.33}$$

We will derive the following two properties of $\Delta^{(T)}\left(\lambda\right)$.

1. Sum:

$$\begin{aligned} \lim_{T\to\infty} \left[\Delta^{(T)}\left(\lambda\right) + \Delta^{(T)}\left(-\lambda\right)\right] &= \lim_{T\to\infty} \Big(\sum_{t=-T}^{T} \exp\left\{\imath\lambda t\right\} - 1 \Big) \\ &= 2\pi\eta\left(\lambda\right) - 1, \text{ by (5.33).} \end{aligned} \tag{5.34}$$

2. Product:

$$\begin{aligned} T^{-2}\Delta^{(T)}\left(-\lambda\right)\Delta^{(T)}\left(\lambda\right) &= T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T} \exp\left\{-\imath\lambda\left(t-s\right)\right\} \\ &= T^{-2} \sum_{t=-(T-1)}^{T-1} \left(T-|t|\right)\exp\left\{-\imath\lambda t\right\} \\ &= T^{-1} \sum_{t=-(T-1)}^{T-1} \exp\left\{-\imath\lambda t\right\} - \sum_{t=-(T-1)}^{T-1} \frac{|t|}{T^2}\exp\left\{-\imath\lambda t\right\}. \end{aligned} \tag{5.35}$$

Consider the second term in the above expression,

$$\Big| \sum_{t=-(T-1)}^{T-1} \frac{|t|}{T^2} \exp\left\{-\imath\lambda t\right\} \Big| \leq \Big| \sum_{t=-(T-1)}^{T-1} \frac{|t|}{T^2} \Big| \to 0 \text{ as } T \to \infty.$$

Hence the expression in (5.35) is given by,

$$T^{-2}\Delta^{(T)}\left(-\lambda\right)\Delta^{(T)}\left(\lambda\right) = T^{-1}2\pi\eta\left(\lambda\right) + o(1). \tag{5.36}$$

Lemma 5.1 *Let $\mathbf{W}_t$ be a stationary h vector-valued time series with n observations satisfying the spectral density given in (5.1). Suppose Assumptions $(A.1)-(A.3)$ hold. The k^{th}-order cumulant of* *the multivariate series,* $\kappa\left\{D^{(n)}_{W_{a_1}}(\lambda_1),...,D^{(n)}_{W_{a_k}}(\lambda_k)\right\}$ *is given by,*

$$n^{-\frac{k}{2}}(2\pi)^{\frac{k}{2}-1}\Delta^{(n)}\left(\sum_{j=1}^{k}\lambda_j\right)f_{W_{a_1}...W_{a_k}}(\lambda_1,...,\lambda_{k-1})+o\left(n^{1-2d-\frac{k}{2}}\right), \tag{5.37}$$

where $f_{W_{a_1}...W_{a_k}}(\lambda_1,...,\lambda_{k-1})$ *is the* k^{th}*-order spectrum of the series* $\mathbf{W}_t$*, and* $a_1,\ldots,a_k=1,2,\ldots,h$*, and* $k=1,2,\ldots$*.*

Proof. As Lemma P4.2 of Brillinger (1981), the cumulant, $\kappa\{D^{(n)}_{W_{a_1}}(\lambda_1),...,D^{(n)}_{W_{a_k}}(\lambda_k)\}$ has the form

$$\sum_{t_1=-\infty}^{\infty}...\sum_{t_k=-\infty}^{\infty}\exp\left(-\imath\sum_{j=1}^{k}\lambda_j t_j\right)\kappa_{W_{a_1}...W_{a_k}}(t_1-t_k,...,t_{k-1}-t_k)$$

Using the substitution, $u_j=t_j-t$ where $t=t_k$, and $-S\le u_j\le S$, for $j=1,\ldots,k-1$ and denoting $S=2(n-1)$ we have that

$$\begin{aligned}
&\kappa\{D^{(n)}_{W_{a_1}}(\lambda_1),D^{(n)}_{W_{a_2}}(\lambda_2),...,D^{(n)}_{W_{a_k}}(\lambda_k)\}\\
=\;&(2\pi n)^{-\frac{k}{2}}\sum_{t=-\infty}^{\infty}\sum_{u_1=-S}^{S}\cdots\sum_{u_k=-S}^{S}\exp\left(-\imath\sum_{j=1}^{k}\lambda_j(u_j+t)\right)\kappa_{W_{a_1}...W_{a_k}}(u_1,...,u_{k-1})\\
=\;&(2\pi n)^{-\frac{k}{2}}\sum_{u_1=-S}^{S}\cdots\sum_{u_k=-S}^{S}\exp\left(-\imath\sum_{j=1}^{k-1}\lambda_j u_j\right)\kappa_{W_{a_1}...W_{a_k}}(u_1,...,u_{k-1})\sum_{t=-\infty}^{\infty}\exp\left(-\imath\sum_{j=1}^{k}\lambda_j t\right)\\
=\;&(2\pi)^{-\frac{k}{2}+1}n^{-\frac{k}{2}}\Delta^{(n)}\left(\textstyle\sum_{j=1}^{k}\lambda_j\right)\sum_{u_1=-S}^{S}\cdots\sum_{u_k=-S}^{S}\exp\left(-\imath\sum_{j=1}^{k-1}\lambda_j u_j\right)\kappa_{W_{a_1}...W_{a_k}}(u_1,...,u_{k-1}).
\end{aligned}$$

Now, let us measure the rapidity of $\sum_{u_1=-S}^{S}\cdots\sum_{u_k=-S}^{S}\exp\left(-\imath\sum_{j=1}^{k-1}\lambda_j u_j\right)\kappa_{W_{a_1}...W_{a_k}}(u_1,...,u_{k-1})$ to $f_{W_{a_1}...W_{a_k}}(\lambda_1,\ldots,\lambda_{k-1})$ as $n\to\infty$.

$$\begin{aligned}
&\left|\sum_{u_1=-S}^{S}\cdots\sum_{u_k=-S}^{S}\exp\left(-\imath\sum_{j=1}^{k-1}\lambda_j u_j\right)\kappa_{W_{a_1}...W_{a_k}}(u_1,...,u_{k-1})-f_{W_{a_1}...W_{a_k}}(\lambda_1,\ldots,\lambda_{k-1})\right|\\
=\;&\left|\sum_{|u_1|>S}\cdots\sum_{|u_k|>S}\exp\left(-\imath\sum_{j=1}^{k-1}\lambda_j u_j\right)\kappa_{W_{a_1}...W_{a_k}}(u_1,...,u_{k-1})\right|\\
\le\;&\sum_{|u_1|>S}\cdots\sum_{|u_k|>S}\left|\kappa_{W_{a_1}...W_{a_k}}(u_1,...,u_{k-1})\right|\\
\le\;&n^{-1+2d}\sum_{|u_1|>S}\cdots\sum_{|u_k|>S}\left(\left|\frac{u_1}{n}\right|^{1-2d}+\cdots+\left|\frac{u_{k-1}}{n}\right|^{1-2d}\right)\left|\kappa_{W_{a_1}...W_{a_k}}(u_1,...,u_{k-1})\right|.
\end{aligned}$$

Hence the proof is completed since Assumption $(A.1)$ holds and $n^{-1+2d}\left(|u_1| + \cdots + |u_{k-1}|\right) \to 0$ as $n \to \infty$. ■

The above Lemma shows that when the DFTs correspond to multivariate time series with the same number of observations in their sample, the k^{th}-order cumulant of the multivariate series can be approximated with the expression given in (5.37). The only difference between this Lemma and Proposition 5.1 is that, the proposition deals with different sample sizes for the time series in the multivariate set-up.

Proof of Proposition 5.1. The proof of the proposition can be established in a similar fashion to the above proof. Hence, we omit the proof here. ■

Proof of Theorem 5.2. The expectation of the DFT of the full sample or the sub-sample is

$$
\begin{aligned}
E\left(D_{X_{a_i}}^{(L_i)}(\lambda)\right) &= E\left(\frac{1}{\sqrt{2\pi L_i}}\sum_{t=1}^{L_i} y_t \exp(-\imath\lambda t)\right) \\
&= \frac{\mu_Y}{\sqrt{2\pi L_i}}\Delta^{(L_i)}(\lambda) \\
&= \begin{cases} 0 & if \quad \lambda \not\equiv 0\,(\text{mod } 2\pi) \\ \sqrt{\frac{L_i}{2\pi}}\mu_Y & if \quad \lambda \equiv \pi\,(\text{mod } 2\pi) \\ 0 \; or \; \sqrt{\frac{L_i}{2\pi}}\mu_Y & if \quad \lambda = \pm\pi, \pm 3\pi, \ldots \end{cases},
\end{aligned}
$$

where $E(y_t) = \mu_Y$. Therefore, $D_{X_{a_i}}^{(L_i)}(\lambda)$ behaves in the manner required by the theorem as the first-order cumulant provides the mean of the random variable of interest.

The covariance between $D_{X_{a_i}}^{(L_i)}(\lambda)$ and $D_{X_{a_j}}^{(L_j)}(\mu)$ is measured by the second-order cumulant and Proposition 5.1 gives that

$$
Cov\left(D_{X_{a_i}}^{(L_i)}(\lambda), D_{X_{a_j}}^{(L_j)}(\mu)\right) = \tfrac{1}{L}\Delta^{(L)}(\lambda+\mu) f_{X_{a_i},X_{a_j}}(\lambda) + o\left(L^{-2d}\right),
$$

where $L = \min\left(L_i, L_j\right)$. Thus, the covariance between the DFTs of the full sample and the sub-sample tends to 0 as $n \to \infty$. ■

Proof of Theorem 5.3. The covariance between $I_{X_{a_i}}^{(L_i)}(\lambda)$ and $I_{X_{a_j}}^{(L_j)}(\mu)$ is given by,

$$\begin{aligned}
Cov\left(I_{X_{a_i}}^{(L_i)}(\lambda), I_{X_{a_j}}^{(L_j)}(\mu)\right) &= E\left(I_{X_{a_i}}^{(L_i)}(\lambda) I_{X_{a_j}}^{(L_j)}(\mu)\right) - E\left(I_{X_{a_i}}^{(L_i)}(\lambda)\right) E\left(I_{X_{a_j}}^{(L_j)}(\mu)\right) \\
&= E\left(D_{X_{a_i}}^{(L_i)}(\lambda) D_{X_{a_i}}^{(L_i)}(-\lambda) D_{X_{a_j}}^{(L_j)}(\mu) D_{X_{a_j}}^{(L_j)}(-\mu)\right) \\
&\quad -E\left(D_{X_{a_i}}^{(L_i)}(\lambda) D_{X_{a_i}}^{(L_i)}(-\lambda)\right) E\left(D_{X_{a_j}}^{(L_j)}(\mu) D_{X_{a_j}}^{(L_j)}(-\mu)\right).
\end{aligned}$$

Since the expectations can be expressed in terms of cumulants (see Appendix 5.B for more details), we may express the covariance term as follows,

$$\begin{aligned}
Cov\left(I_{X_{a_i}}^{(L_i)}(\lambda), I_{X_{a_j}}^{(L_j)}(\mu)\right) &= \kappa\left(D_{X_{a_i}}^{(L_i)}(\lambda), D_{X_{a_i}}^{(L_i)}(-\lambda), D_{X_{a_j}}^{(L_j)}(\mu), D_{X_{a_j}}^{(L_j)}(-\mu)\right) \\
&\quad +\kappa\left(D_{X_{a_i}}^{(L_i)}(-\lambda), D_{X_{a_j}}^{(L_j)}(\mu)\right) \kappa\left(D_{X_{a_i}}^{(L_i)}(\lambda), D_{X_{a_j}}^{(L_j)}(-\mu)\right) \\
&\quad +\kappa\left(D_{X_{a_i}}^{(L_i)}(\lambda), D_{X_{a_j}}^{(L_j)}(\mu)\right) \kappa\left(D_{X_{a_i}}^{(L_i)}(-\lambda), D_{X_{a_j}}^{(L_j)}(-\mu)\right).
\end{aligned}$$

Then Proposition 5.1 gives us that,

$$\begin{aligned}
Cov\left(I_{X_{a_i}}^{(L_i)}(\lambda), I_{X_{a_j}}^{(L_j)}(\mu)\right) &= L^{-2}(2\pi)\Delta^{(L)}(\lambda+\mu-\lambda-\mu) f_{X_{a_i}X_{a_i}X_{a_j}X_{a_j}}(\lambda,-\lambda,\mu) + o\left(L^{-1-2d}\right) \\
&\quad + \left(L^{-1}\Delta^{(L)}(-\lambda+\mu) f_{X_{a_i}X_{a_j}}(-\lambda) + o\left(L^{-2d}\right)\right) \\
&\quad \times \left(L^{-1}\Delta^{(L)}(\lambda-\mu) f_{X_{a_i}X_{a_j}}(\lambda) + o\left(L^{-2d}\right)\right) \\
&\quad + \left(L^{-1}\Delta^{(L)}(\lambda+\mu) f_{X_{a_i}X_{a_j}}(\lambda) + o\left(L^{-2d}\right)\right) \\
&\quad \times \left(L^{-1}\Delta^{(L)}(-\lambda-\mu) f_{X_{a_i}X_{a_j}}(-\lambda) + o\left(L^{-2d}\right)\right) \\
&= L^{-2}(2\pi)\Delta^{(L)}(0) f_{X_{a_i}X_{a_i}X_{a_j}X_{a_j}}(\lambda,-\lambda,\mu) + o\left(L^{-1-2d}\right) \\
&\quad +L^{-2}\Delta^{(L)}(-\lambda+\mu)\Delta^{(L)}(\lambda-\mu)\left(f_{X_{a_i}X_{a_j}}(\lambda)\right)^2 \\
&\quad +L^{-1}\left(\Delta^{(L)}(-\lambda+\mu)+\Delta^{(L)}(\lambda-\mu)\right) f_{X_{a_i}X_{a_j}}(\lambda)\, o\left(L^{-2d}\right) \\
&\quad +L^{-2}\Delta^{(L)}(\lambda+\mu)\Delta^{(L)}(-\lambda-\mu)\left(f_{X_{a_i}X_{a_j}}(\lambda)\right)^2 \\
&\quad +L^{-1}\Delta^{(L)}(\lambda+\mu) f_{X_{a_i}X_{a_j}}(-\lambda) + \Delta^{(L)}(-\lambda-\mu) f_{X_{a_i}X_{a_j}}(-\lambda)\, o\left(L^{-2d}\right)
\end{aligned}$$

$$
\begin{aligned}
&= L^{-1}(2\pi) f_{X_{a_i}X_{a_i}X_{a_j}X_{a_j}}(\lambda,-\lambda,\mu) + L^{-2}\left[\Delta^{(L)}(-\lambda+\mu)\Delta^{(L)}(\lambda-\mu)\right. \\
&\quad \left. + \Delta^{(L)}(\lambda+\mu)\Delta^{(L)}(-\lambda-\mu)\right]\left(f_{X_{a_i}X_{a_j}}(\lambda)\right)^2 + \left[\Delta^{(L)}(-\lambda+\mu)\right. \\
&\quad \left. + \Delta^{(L)}(\lambda-\mu) + \Delta^{(L)}(\lambda+\mu) + \Delta^{(L)}(-\lambda-\mu)\right] f_{X_{a_i}X_{a_j}}(\lambda)\, o\left(L^{-2d}\right) \\
&\quad + o\left(L^{-1-2d}\right) + o\left(L^{-4d}\right). \qquad (5.38)
\end{aligned}
$$

Using the two properties in (5.34) and (5.36), the covariance in (5.38) is simplified further as follows,

$$
\begin{aligned}
Cov\left(I^{(L_i)}_{X_{a_i}}(\lambda), I^{(L_j)}_{X_{a_j}}(\mu)\right) &= \frac{2\pi}{L}\left[\eta(\lambda-\mu)+\eta(\lambda+\mu)\right]\left\{f_{X_{a_i}X_{a_j}}(\lambda)\right\}^2 + \frac{2\pi}{l^{\dagger}} f_{X_{a_i}X_{a_i}X_{a_j}X_{a_j}}(\lambda,-\lambda,\mu) \\
&\quad + 2\pi\left[\eta(\lambda-\mu)+\eta(\lambda+\mu)\right] f_{X_{a_i}X_{a_j}}(\lambda)\, o\left(l^{\dagger^{-2d}}\right) + o\left(L^{-1-2d}\right).
\end{aligned}
$$

Now let us consider the asymptotic distribution of $I^{(L_i)}_{X_{a_i}}(\lambda)$. We may re-write the periodogram as follows,

$$
I^{(L_i)}_{X_{a_i}}(\lambda) = \left[\operatorname{Re} D^{(L_i)}_{X_{a_i}}(\lambda)\right]^2 + \left[\operatorname{Im} D^{(L_i)}_{X_{a_i}}(\lambda)\right]^2,
$$

where

$$
\operatorname{Re} D^{(L_i)}_{X_{a_i}}(\lambda) = \frac{1}{\sqrt{2\pi L_i}}\sum_{t=1}^{L_i} y_t \cos(\lambda t), \text{ and, } \operatorname{Im} D^{(L_i)}_{X_{a_i}}(\lambda) = \frac{1}{\sqrt{2\pi L_i}}\sum_{t=1}^{L_i} y_t \sin(\lambda t).
$$

Following Theorem 2.1 of Lahiri (2003), we have that

$$
\begin{bmatrix}
\dfrac{\operatorname{Re} D^{(L_i)}_{X_{a_i}}(\lambda) - E\left(\operatorname{Re} D^{(L_i)}_{X_{a_i}}(\lambda)\right)}{\sqrt{L_i f_{X_{a_i}X_{a_i}}(\lambda)}} \\
\dfrac{\operatorname{Im} D^{(L_i)}_{X_{a_i}}(\lambda) - E\left(\operatorname{Im} D^{(L_i)}_{X_{a_i}}(\lambda)\right)}{\sqrt{L_i f_{X_{a_i}X_{a_i}}(\lambda)}}
\end{bmatrix} \to^{D} N(\mathbf{0}, \mathbf{I}_2).
$$

Hence the result. ■

Proof of Theorem 5.4. Recall that $x_j = \ln(2\sin(\lambda_j/2))$, $a_j = x_j - \overline{x}$ and $S_{xx} = \sum_{j=1}^{N_n}(X_j - \overline{X})^2$. From Hurvich *et al.* (1998) we have that $S_{xx} = N_n(1+o(1))$ and $a_j = \log j - \log N_n + 1 +$

$o(1) + o\left(\frac{N_n^2}{n^2}\right)$, $j = 1, \ldots, N_n$. Thus, $\sup_j |a_j| = 1 + o(1) + O\left(\frac{N_n^2}{n^2}\right)$. Using Appendix 5.B we have that

$$\begin{aligned} Cov\left(\log I_{X_{a_i}}^{(L_i)}(\lambda_j), \log I_{X_{a_j}}^{(L_j)}(\mu_k)\right) &= (1-\rho^2)^{\frac{1}{2}} \sum_{k=1}^{\infty} \left(\Psi\left(\frac{1}{2}+k\right) + \Psi\left(\frac{1}{2}\right)\right)^2 \frac{\Gamma\left(\frac{1}{2}+k\right)}{\Gamma\left(\frac{1}{2}\right)} \frac{\left(\rho^2\right)^k}{k!} \\ &\quad - (1-\rho^2)\left(\sum_{k=1}^{\infty} \left(\Psi\left(\frac{1}{2}+k\right) + \Psi\left(\frac{1}{2}\right)\right) \frac{\Gamma\left(\frac{1}{2}+k\right)}{\Gamma\left(\frac{1}{2}\right)} \frac{\left(\rho^2\right)^k}{k!}\right)^2 \\ &\leq (1-\rho^2)^{\frac{1}{2}} \sum_{k=1}^{\infty} \left(\Psi\left(\frac{1}{2}+k\right) + \Psi\left(\frac{1}{2}\right)\right)^2 \frac{\Gamma\left(\frac{1}{2}+k\right)}{\Gamma\left(\frac{1}{2}\right)} \frac{\left(\rho^2\right)^k}{k!}, \end{aligned}$$

where $\rho = Corr\left(I_{X_{a_i}}^{(L_i)}(\lambda_j), I_{X_{a_j}}^{(L_j)}(\mu_k)\right) = o(n^{-1})$ by Theorem 5.3. Thus,

$$Cov\Big(\log I_{X_{a_i}}^{(L_i)}(\lambda_j), \log I_{X_{a_j}}^{(L_j)}(\mu_k)\Big) = o\left(n^{-1}\right).$$

This leads to

$$\begin{aligned} Cov\left(\widehat{d}_n, \widehat{d}_i\right) &= \frac{1}{4S_{xx}} \frac{1}{S'_{xx}} \sum_{j=1}^{N_n} \sum_{k=1}^{N_l} a_j a_k^{(i)} Cov\left(\log I_{X_{a_i}}^{(L_i)}(\lambda_j), \log I_{X_{a_j}}^{(L_j)}(\mu_k)\right) \\ &\leq \sup_{j,k} \frac{1}{4S_{xx}} \frac{1}{S'_{xx}} N_n N_l \left| a_j a_k^{(i)} Cov\left(\log I_{X_{a_i}}^{(L_i)}(\lambda_j), \log I_{X_{a_j}}^{(L_j)}(\mu_k)\right)\right| \\ &= \frac{(1+o(1))^{-2}}{4} \sup_{j,k} |a_j| \left|a_k^{(i)}\right| \left|Cov\left(\log I_{X_{a_i}}^{(L_i)}(\lambda_j), \log I_{X_{a_j}}^{(L_j)}(\mu_k)\right)\right| \\ &= \frac{(1+o(1))^{-2}}{4} \left(1 + o(1) + O\left(\frac{N_n^2}{n^2}\right)\right)^2 \sup_{j,k} \left|Cov\left(\log I_{X_{a_i}}^{(L_i)}(\lambda_j), \log I_{X_{a_j}}^{(L_j)}(\mu_k)\right)\right| \\ &= o\left(n^{-1}\right). \end{aligned}$$

Similarly, we can prove that $Cov\left(\widehat{d}_i, \widehat{d}_j\right) = o(n^{-1})$. Hence the result. ■

Proof of Theorem 5.5. Consider,

$$\left(\widehat{d}_{J,m}^{Opt} - d_0\right) = w_n^* \left(\widehat{d}_n - d_0\right) - \sum_{i=1}^{m} w_i^* \left(\widehat{d}_{i,m} - d_0\right). \tag{5.39}$$

Recall that $w_n^* = \left[1 - \left(\frac{1}{m}\frac{N_n}{n}\frac{l}{N_l}\right)^2\right]^{-1}$ and $\sum_{i=1}^m w_i^* = w_n^* - 1$; for $i = 1, \ldots, m$. Let us firstly consider w_n^*. For fixed m and for the choice of N_n such that $N_n \log N_n / n \to 0$,

$$w_n^* = \frac{1}{1 - \left(n^{-1} l n^{-1+\alpha} l^{1-\alpha}\right)^2} = 1 + o(1), \tag{5.40}$$

and hence

$$\sum_{i=1}^{m} w_i^* = o\left(1\right), \tag{5.41}$$

with $w_i^* \to 0$ as $n \to \infty$ (see the proof of Theorem 5.4).

By virtue of the consistency of $\widehat{d}_n$, we have that $w_n^*\left(\widehat{d}_n - d\right) = o_p\left(1\right)$ using (5.40).

Now, we show that the second term in (5.39) is $o_p\left(1\right)$.

$$\begin{aligned}
\lim_{n\to\infty} \Pr\left[\left|\sum_{i=1}^{m} w_i^*\left(\widehat{d}_i - d_0\right)\right| \geq \varepsilon\right] &\leq \lim_{n\to\infty} \frac{E\left(\sum_{i=1}^{m} w_i^*\left(\widehat{d}_i - d_0\right)\right)^2}{\varepsilon^2} \\
&= \lim_{n\to\infty} \frac{Var\left(\widehat{d}_i\right)}{\varepsilon^2}\sum_{i=1}^{m}\left(w_i^*\right)^2 \\
&\quad + \frac{2}{\varepsilon^2}\lim_{n\to\infty}\sum_{i=1}^{m}\sum_{j=i+1}^{m} w_i^* w_j^* Cov\left(\widehat{d}_i, \widehat{d}_j\right) \\
&= 0,
\end{aligned}$$

since $\lim_{n\to\infty} Var\left(\widehat{d}_i\right) = 0$ from (5.7), $\lim_{n\to\infty} Cov\left(\widehat{d}_i, \widehat{d}_j\right) = 0$ directly from Theorem 5.2 and the limit of $\sum_{i=1}^{m} w_i^*$ given in (5.41). This completes the proof of consistency.

The proof of asymptotic normality of the optimal jackknife statistic depends on the joint convergence of $\widehat{d}_n$ and $\widehat{d}_{i,m}$. Firstly, let us consider the following standardized optimal jackknife estimator,

$$\sqrt{N_n}\left(\widehat{d}_{J,m}^{Opt} - d_0\right) = w_n^*\sqrt{N_n}\left(\widehat{d}_n - d_0\right) - \sum_{i=1}^{m} w_i^*\sqrt{N_n}\left(\widehat{d}_i - d_0\right). \tag{5.42}$$

Using Theorem 5.1 we have that $\sqrt{N_n}\left(\widehat{d}_n - d_0\right) \to^D N\left(0, \frac{\pi^2}{24}\right)$. Therefore, regarding the first component in (5.42), it immediately follows that

$$w_n^*\sqrt{N_n}\left(\widehat{d}_n - d_0\right) \to^D N\left(0, \frac{\pi^2}{24}\right), \text{ using (5.40).}$$

Now, let us consider the second term in (5.42):

$$\lim_{n\to\infty} \Pr\left[\left|\sum_{i=1}^{m} w_i^*\sqrt{N_n}\left(\widehat{d}_i - d_0\right)\right| \geq \varepsilon\right] \leq \lim_{n\to\infty} \frac{E\left(\sum_{i=1}^{m} w_i^*\left(\widehat{d}_i - d_0\right)\right)^2}{\varepsilon^2} N_n$$

$$= \lim_{n\to\infty} \frac{Var\left(\widehat{d}_i\right)}{\varepsilon^2} N_n \sum_{i=1}^{m} \left(w_i^*\right)^2 + \lim_{n\to\infty} \frac{2N_n}{\varepsilon^2} \sum_{i=1}^{m} \sum_{j=i+1}^{m} w_i^* w_j^* Cov\left(\widehat{d}_i, \widehat{d}_j\right). \quad (5.43)$$

By considering the first term in (5.43), for fixed m we have that

$$\lim_{n\to\infty} \frac{Var\left(\widehat{d}_i\right)}{\varepsilon^2} N_n \sum_{i=1}^{m} \left(w_i^*\right)^2 = \lim_{n\to\infty} \frac{\sum_{i=1}^{m} \left(w_i^*\right)^2}{\varepsilon^2} \left[\frac{\pi^2}{24} + o\left(1\right)\right] = 0,$$

using (5.7) and (5.40). The second term in (5.43) would give us that,

$$\lim_{n\to\infty} \frac{2N_n}{\varepsilon^2} \sum_{i=1}^{m} \sum_{j=i+1}^{m} w_i^* w_j^* Cov\left(\widehat{d}_i, \widehat{d}_j\right) = 0,$$

straightforwardly from (5.40). Therefore, $\Pr\left[\left|\sum_{i=1}^{m} w_i^* \sqrt{N_l}\left(\widehat{d}_i - d_0\right)\right| \geq \varepsilon\right] \to 0$ as $n \to \infty$. Hence the result. ■

5.B Appendix: Additional technical results

5.B.1 Evaluation of the covariance terms in (5.15) and (5.16)

The main purpose of this exercise is to calculate the covariances between the full-sample and sub-sample LPR estimators (refer to (5.15)) and the covariance between two distinct sub-sample LPR estimators (refer to (5.16))). These covariance terms depend on the covariance between the log-periodograms associated with either the full sample and a given sub-sample or two different sub-samples.

To obtain the covariance between the log-periodograms associated with the full sample and a given sub-sample, or between sub-samples, we follow the method stated below.

Step 1: Write down the joint distribution of the periodograms $\left(I_{X_{a_i}}^{(L_i)}\left(\lambda\right), I_{X_{a_j}}^{(L_j)}\left(\mu\right)\right)$.

Step 2: Write down the joint distribution of the log transformed periodograms $(\log I_{X_{a_i}}^{(L_i)}(\lambda), \log I_{X_{a_j}}^{(L_j)}(\mu))$ using the expression of the covariance between the two different periodograms,

Step 3: Find the expression for the covariance between the above mentioned log-periodograms, $Cov(\log I_{X_{a_i}}^{(L_i)}(\lambda), \log I_{X_{a_j}}^{(L_j)}(\mu))$, using the moment generating function.

In relation to Step 1: Using the results of Theorem 5.3, we can say that the periodograms associated with the full sample and the sub-sample have a limiting distribution of the form $f_{X_1X_1}(\lambda)\, \chi^2_{(2)}\big/ 2$. For notational convenience, let us denote by (U, V) the bivariate χ^2_k random variables, $(I_{X_{a_i}}^{(L_i)}(\lambda), I_{X_{a_j}}^{(L_j)}(\mu))$. Although $k = 2$, we use the generic notation for the degrees of freedom, k. Note that we ignore the constant term $f_{X_1X_1}(\lambda)\big/ 2$ for convenience, as these terms will disappear in the calculation of the covariance between two different LPR estimators (either the full- and sub-sample LPR estimators or two distinct sub-sample LPR estimators).

The joint probability density function (pdf), $f_{U,V}(u,v)$, is defined by (see, Krishnaiah *et al.*, 1963),

$$f_{U,V}(u,v) = \left(1-\rho^2\right)^{\frac{k-1}{2}} \sum_{i=0}^{\infty} \frac{\Gamma\left(\frac{k-1}{2}+i\right)\rho^{2i}(uv)^{\frac{k-3+2i}{2}} \exp\left[-\frac{u+v}{2(1-\rho^2)}\right]}{\Gamma\left(\frac{k-1}{2}\right) i! \left[2^{\frac{k-1}{2}+i}\Gamma\left(\frac{k-1}{2}+i\right)\left(1-\rho^2\right)^{\frac{k-1}{2}+i}\right]^2},$$

where $\rho = \dfrac{\sigma_{uv}}{\sigma_u \sigma_v}$. Here, $\sigma_{uv} = cov(U, V)$. Then, the marginal densities of U and V, $f_U(u)$ and $f_V(v)$, are respectively given by,

$$f_U(u) = \frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} u^{\frac{k}{2}} \exp\left\{-\frac{u}{2}\right\}, \text{ and, } f_V(v) = \frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} v^{\frac{k}{2}} \exp\left\{-\frac{v}{2}\right\}.$$

In relation to Step 2: Let $W = \log U = \log I_{X_{a_i}}^{(L_i)}(\lambda)$ and $Z = \log V = \log I_{X_{a_j}}^{(L_j)}(\mu)$. Then, the joint pdf of W and Z is given by,

$$\begin{aligned}
f_{W,Z}(w,z) &= f_{U,V}(\exp w, \exp z)\left|\begin{array}{cc} \frac{\partial \exp w}{\partial w} & \frac{\partial \exp w}{\partial z} \\ \frac{\partial \exp z}{\partial w} & \frac{\partial \exp z}{\partial z} \end{array}\right| \\
&= \left(1-\rho^2\right)^{\frac{k-1}{2}} \sum_{i=0}^{\infty} \frac{\Gamma\left(\frac{k-1}{2}+i\right)\rho^{2i}\left(\exp w \exp z\right)^{\frac{k-3+2i}{2}} \exp\left[-\frac{\exp w+\exp z}{2(1-\rho^2)}\right]}{\Gamma\left(\frac{k-1}{2}\right) i! \left[2^{\frac{k-1}{2}+i}\Gamma\left(\frac{k-1}{2}+i\right)\left(1-\rho^2\right)^{\frac{k-1}{2}+i}\right]^2} \exp w \exp z \\
&= \left(1-\rho^2\right)^{\frac{k-1}{2}} \sum_{i=0}^{\infty} \frac{\Gamma\left(\frac{k-1}{2}+i\right)\rho^{2i} \exp\left(\frac{k-1}{2}+i\right)(w+z) \exp\left[-\frac{\exp w+\exp z}{2(1-\rho^2)}\right]}{\Gamma\left(\frac{k-1}{2}\right) i! \left[2^{\frac{k-1}{2}+i}\Gamma\left(\frac{k-1}{2}+i\right)\left(1-\rho^2\right)^{\frac{k-1}{2}+i}\right]^2}.
\end{aligned}$$

In relation to Step 3: The moment generating function (MGF) of (W, Z) is given by,

$$\begin{aligned}
M_{W,Z}(t_1,t_2) &= E\left(\exp\left(t_1 W + t_2 Z\right)\right) = \int_0^\infty \int_0^\infty \exp\left(t_1 w + t_2 z\right) f_{W,Z}(w,z)\, dw dz \\
&= \left(1-\rho^2\right)^{\frac{k-1}{2}} \sum_{i=0}^{\infty} \frac{\Gamma\left(\frac{k-1}{2}+i\right)\rho^{2i}}{\Gamma\left(\frac{k-1}{2}\right) i! \left[2^{\frac{k-1}{2}+i}\Gamma\left(\frac{k-1}{2}+i\right)\left(1-\rho^2\right)^{\frac{k-1}{2}+i}\right]^2} \\
&\quad \times \int_0^\infty \int_0^\infty \exp\left(t_1 w + t_2 z\right) \exp\left(\frac{k-1}{2}+i\right)(w+z) \exp\left[-\frac{\exp w + \exp z}{2\left(1-\rho^2\right)}\right] dw dz \\
&= \left(1-\rho^2\right)^{\frac{k-1}{2}} \sum_{i=0}^{\infty} \frac{\Gamma\left(\frac{k-1}{2}+i\right)\rho^{2i}}{\Gamma\left(\frac{k-1}{2}\right) i! \left[2^{\frac{k-1}{2}+i}\Gamma\left(\frac{k-1}{2}+i\right)\left(1-\rho^2\right)^{\frac{k-1}{2}+i}\right]^2} \\
&\quad \times \int_0^\infty \exp\left(\frac{k-1}{2}+t_1+i\right) w \exp\left[-\frac{\exp w}{2\left(1-\rho^2\right)}\right] dw \\
&\quad \times \int_0^\infty \exp\left(\frac{k-1}{2}+t_2+i\right) z \exp\left[-\frac{\exp z}{2\left(1-\rho^2\right)}\right] dz. \qquad (5.44)
\end{aligned}$$

Now let us consider the form of the last expression in (5.44). Let $\alpha_1 = \dfrac{k-1}{2} + t_2 + i$ and $\alpha_2 = \dfrac{1}{2\left(1-\rho^2\right)}$. Then, substituting $x = \exp z$ would give us that

$$\int_0^\infty \exp \alpha_1 z \exp\left[-\alpha_2 \exp z\right] dz = \int_0^\infty x^{\alpha_1 - 1} \exp\left[-\alpha_2 x\right] dx = \frac{\Gamma\left(\alpha_1\right)}{\alpha_2^{\alpha_1}}. \qquad (5.45)$$

Therefore, using (5.45), the MGF given in (5.44) may be re-arranged as follows,

$$M_{W,Z}(t_1,t_2) = \left[2\left(1-\rho^2\right)\right]^{t_1+t_2}\left(1-\rho^2\right)^{\frac{k-1}{2}} \sum_{i=0}^{\infty} \frac{\Gamma\left(\frac{k-1}{2}+i\right)\rho^{2i}\Gamma\left(\frac{k-1}{2}+t_2+i\right)\Gamma\left(\frac{k-1}{2}+t_1+i\right)}{i!\Gamma\left(\frac{k-1}{2}\right)\left[\Gamma\left(\frac{k-1}{2}+i\right)\right]^2}$$

$$
\begin{aligned}
&= \left[2\left(1-\rho^2\right)\right]^{t_1+t_2}\left(1-\rho^2\right)^{\frac{k-1}{2}}\frac{\Gamma\left(\frac{k-1}{2}+t_1\right)\Gamma\left(\frac{k-1}{2}+t_2\right)}{\left[\Gamma\left(\frac{k-1}{2}\right)\right]^2} \\
&\quad\times\sum_{i=0}^{\infty}\frac{\Gamma\left(\frac{k-1}{2}+t_1+i\right)\Gamma\left(\frac{k-1}{2}+t_2+i\right)\Gamma\left(\frac{k-1}{2}\right)}{\Gamma\left(\frac{k-1}{2}+t_1\right)\Gamma\left(\frac{k-1}{2}+t_2\right)\Gamma\left(\frac{k-1}{2}+i\right)}\frac{\left(\rho^2\right)^i}{i!} \\
&= \left[2\left(1-\rho^2\right)\right]^{t_1+t_2}\left(1-\rho^2\right)^{\frac{k-1}{2}}\frac{\Gamma\left(\frac{k-1}{2}+t_1\right)\Gamma\left(\frac{k-1}{2}+t_2\right)}{\left[\Gamma\left(\frac{k-1}{2}\right)\right]^2} \\
&\quad\times{}_2F_1\left(\frac{k-1}{2}+t_1,\frac{k-1}{2}+t_2;\frac{k-1}{2};\rho^2\right).
\end{aligned}
$$

Setting $k=2$ gives,

$$
M_{W,Z}\left(t_1,t_2\right)=\left[2\left(1-\rho^2\right)\right]^{t_1+t_2}\left(1-\rho^2\right)^{\frac{1}{2}}\frac{\Gamma\left(\frac{1}{2}+t_1\right)\Gamma\left(\frac{1}{2}+t_2\right)}{\left[\Gamma\left(\frac{1}{2}\right)\right]^2}\,{}_2F_1\left(\frac{1}{2}+t_1,\frac{1}{2}+t_2;\frac{1}{2};\rho^2\right).
$$

Therefore the cumulant generating function is given by,

$$
\begin{aligned}
K\left(t_1,t_2\right) &= \log M_{W,Z}\left(t_1,t_2\right) \\
&= \left(t_1+t_2\right)\log\left[2\left(1-\rho^2\right)\right]+\frac{1}{2}\log\left(1-\rho^2\right)+\log\Gamma\left(\frac{1}{2}+t_1\right) \\
&\quad+\log\Gamma\left(\frac{1}{2}+t_2\right)-2\log\left[\Gamma\left(\frac{1}{2}\right)\right]+\log_2 F_1\left(\frac{1}{2}+t_1,\frac{1}{2}+t_2;\frac{1}{2};\rho^2\right).
\end{aligned}
$$

The covariance between W and Z when $k=2$, is given by, $cov\left(W,Z\right)=\left.\frac{\partial^2K(t_1,t_2)}{\partial t_1\partial t_2}\right|_{t_1=0,t_2=0}$.

Therefore, let us firstly evaluate $\partial K\left(t_1,t_2\right)/\,\partial t_1$, as

$$
\begin{aligned}
\frac{\partial K\left(t_1,t_2\right)}{\partial t_1} &= \log\left[2\left(1-\rho^2\right)\right]+\Psi\left(\frac{1}{2}+t_1\right) \\
&\quad+\left({}_2F_1\left(\frac{1}{2}+t_1,\frac{1}{2}+t_2;\frac{1}{2};\rho^2\right)\right)^{-1}\frac{\partial{}_2F_1\left(\frac{1}{2}+t_1,\frac{1}{2}+t_2;\frac{1}{2};\rho^2\right)}{\partial t_1},
\end{aligned}
\tag{5.46}
$$

where $\Psi\left(.\right)$ is the digamma function and where $\partial{}_2F_1\left(\frac{1}{2}+t_1,\frac{1}{2}+t_2;\frac{1}{2};\rho^2\right)/\,\partial t_1$ is given by,

$$
\begin{aligned}
&\sum_{i=1}^{\infty}\frac{\partial\,\Gamma\left(\frac{1}{2}+t_1+i\right)/\,\Gamma\left(\frac{1}{2}+t_1\right)}{\partial t_1}\frac{\Gamma\left(\frac{1}{2}+t_2+i\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}+t_2\right)\Gamma\left(\frac{1}{2}+i\right)}\frac{\left(\rho^2\right)^i}{i!} \\
=\;&\sum_{i=1}^{\infty}\left(\frac{\Gamma\left(\frac{1}{2}+t_1\right)\Gamma\left(\frac{1}{2}+t_1+i\right)\Psi\left(\frac{1}{2}+t_1+i\right)}{\left(\Gamma\left(\frac{1}{2}+t_1\right)\right)^2}+\frac{\Gamma\left(\frac{1}{2}+t_1+i\right)\Psi\left(\frac{1}{2}+t_1\right)\Gamma\left(\frac{1}{2}+t_1\right)}{\left(\Gamma\left(\frac{1}{2}+t_1\right)\right)^2}\right)
\end{aligned}
$$

$$
\begin{aligned}
&\times\frac{\Gamma\left(\frac{1}{2}+t_2+i\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}+t_2\right)\Gamma\left(\frac{1}{2}+i\right)}\frac{\left(\rho^2\right)^i}{i!}\\
=\;& \sum_{i=1}^{\infty}\left(\frac{\Gamma\left(\frac{1}{2}+t_1+i\right)\Psi\left(\frac{1}{2}+t_1+i\right)+\Gamma\left(\frac{1}{2}+t_1+i\right)\Psi\left(\frac{1}{2}+t_1\right)}{\Gamma\left(\frac{1}{2}+t_1\right)}\right)\frac{\Gamma\left(\frac{1}{2}+t_2+i\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}+t_2\right)\Gamma\left(\frac{1}{2}+i\right)}\frac{\left(\rho^2\right)^i}{i!}.
\end{aligned}
$$

This leads to,

$$
\left.\frac{\partial{}_2F_1\left(\frac{1}{2}+t_1,\frac{1}{2}+t_2;\frac{1}{2};\rho^2\right)}{\partial t_1}\right|_{t_1=0,t_2=0}=\sum_{i=1}^{\infty}\left(\Psi\left(\frac{1}{2}+i\right)+\Psi\left(\frac{1}{2}\right)\right)\frac{\Gamma\left(\frac{1}{2}+i\right)}{\Gamma\left(\frac{1}{2}\right)}\frac{\left(\rho^2\right)^i}{i!}.
$$

Now let us evaluate the second order derivative of $K\left(t_1,t_2\right)$,

$$
\begin{aligned}
\frac{\partial^2 K\left(t_1,t_2\right)}{\partial t_1\partial t_2} &= \frac{\partial\left({}_2F_1\left(\frac{1}{2}+t_1,\frac{1}{2}+t_2;\frac{1}{2};\rho^2\right)\right)^{-1}\frac{\partial{}_2F_1\left(\frac{1}{2}+t_1,\frac{1}{2}+t_2;\frac{1}{2};\rho^2\right)}{\partial t_1}}{\partial t_2}\\
&= \left({}_2F_1\left(\frac{1}{2}+t_1,\frac{1}{2}+t_2;\frac{1}{2};\rho^2\right)\right)^{-1}\frac{\partial^2\,{}_2F_1\left(\frac{1}{2}+t_1,\frac{1}{2}+t_2;\frac{1}{2};\rho^2\right)}{\partial t_1\partial t_2}\\
&\quad-\left({}_2F_1\left(\frac{1}{2}+t_1,\frac{1}{2}+t_2;\frac{1}{2};\rho^2\right)\right)^{-2}\frac{\partial{}_2F_1\left(\frac{1}{2}+t_1,\frac{1}{2}+t_2;\frac{1}{2};\rho^2\right)}{\partial t_2}\\
&\quad\times\frac{\partial{}_2F_1\left(\frac{1}{2}+t_1,\frac{1}{2}+t_2;\frac{1}{2};\rho^2\right)}{\partial t_1},
\end{aligned}
$$

where $\partial^2_2F_1\left(\frac{1}{2}+t_1,\frac{1}{2}+t_2;\frac{1}{2};\rho^2\right)/\partial t_1\partial t_2$ is given by,

$$
\begin{aligned}
&\sum_{i=1}^{\infty}\left(\frac{\Gamma\left(\frac{1}{2}+t_1+i\right)\Psi\left(\frac{1}{2}+t_1+i\right)}{\Gamma\left(\frac{1}{2}+t_1\right)}+\frac{\Gamma\left(\frac{1}{2}+t_1+i\right)\Psi\left(\frac{1}{2}+t_1\right)}{\Gamma\left(\frac{1}{2}+t_1\right)}\right)\frac{\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}+i\right)}\frac{\left(\rho^2\right)^i}{i!}\\
&\times\left(\frac{\Gamma\left(\frac{1}{2}+t_2+i\right)\Psi\left(\frac{1}{2}+t_2+i\right)}{\Gamma\left(\frac{1}{2}+t_2\right)}+\frac{\Gamma\left(\frac{1}{2}+t_2+i\right)\Psi\left(\frac{1}{2}+t_2\right)}{\Gamma\left(\frac{1}{2}+t_2\right)}\right).
\end{aligned}
$$

Thus,

$$
\left.\frac{\partial^2\,{}_2F_1\left(\frac{1}{2}+t_1,\frac{1}{2}+t_2;\frac{1}{2};\rho^2\right)}{\partial t_1\partial t_2}\right|_{t_1=0,t_2=0}=\sum_{i=1}^{\infty}\left(\Psi\left(\frac{1}{2}+i\right)+\Psi\left(\frac{1}{2}\right)\right)^2\frac{\Gamma\left(\frac{1}{2}+i\right)}{\Gamma\left(\frac{1}{2}\right)}\frac{\left(\rho^2\right)^i}{i!}.
$$

Hence $cov\left(W,Z\right)=\left.\frac{\partial^2 K\left(t_1,t_2\right)}{\partial t_1\partial t_2}\right|_{t_1=0,t_2=0}$ is given by,

$$
\begin{aligned}
&\left(1-\rho^2\right)^{\frac{1}{2}}\left.\frac{\partial^2\,{}_2F_1\left(\frac{1}{2}+t_1,\frac{1}{2}+t_2;\frac{1}{2};\rho^2\right)}{\partial t_1\partial t_2}\right|_{t_1=0,t_2=0}\\
&-\left(1-\rho^2\right)\left.\frac{\partial{}_2F_1\left(\frac{1}{2}+t_1,\frac{1}{2}+t_2;\frac{1}{2};\rho^2\right)}{\partial t_2}\frac{\partial{}_2F_1\left(\frac{1}{2}+t_1,\frac{1}{2}+t_2;\frac{1}{2};\rho^2\right)}{\partial t_1}\right|_{t_1=0,t_2=0}
\end{aligned}
$$

$$\begin{aligned}
&= \left(1-\rho^2\right)^{\frac{1}{2}} \sum_{i=1}^{\infty} \left(\Psi\left(\frac{1}{2}+i\right)+\Psi\left(\frac{1}{2}\right)\right)^2 \frac{\Gamma\left(\frac{1}{2}+i\right)}{\Gamma\left(\frac{1}{2}\right)} \frac{\left(\rho^2\right)^i}{i!} \\
&\quad - \left(1-\rho^2\right) \left(\sum_{i=1}^{\infty} \left(\Psi\left(\frac{1}{2}+i\right)+\Psi\left(\frac{1}{2}\right)\right) \frac{\Gamma\left(\frac{1}{2}+i\right)}{\Gamma\left(\frac{1}{2}\right)} \frac{\left(\rho^2\right)^i}{i!}\right)^2, \qquad (5.47)
\end{aligned}$$

using the fact ${}_1F_0\left(a;;z\right) = \left(1-z\right)^{-a}$.

Let us now provide the expression for ρ in (5.47). For example, consider calculating the correlation between the full- and sub-sample periodograms. Using the similar arguments, the correlation between two sub-samples periodograms can be derived,

$$\rho = corr\left(I_Y^{(n)}\left(\lambda\right), I_{Y_i}^{(l)}\left(\mu\right)\right) = \frac{Cov\left(I_Y^{(n)}\left(\lambda\right), I_{Y_i}^{(l)}\left(\mu\right)\right)}{\sqrt{Var\left(I_Y^{(n)}\left(\lambda\right)\right)}\sqrt{Var\left(I_{Y_i}^{(l)}\left(\mu\right)\right)}}, \qquad (5.48)$$

where,

$$\begin{aligned}
Cov\left(I_Y^{(n)}\left(\lambda\right), I_{Y_i}^{(l)}\left(\mu\right)\right) &\approx \frac{2\pi}{l} f_{YYY_iY_i}\left(\lambda,-\lambda,\mu\right) + l^{-2}\left[\Delta^{(l)}\left(-\lambda+\mu\right)\Delta^{(l)}\left(\lambda-\mu\right)\right. \\
&\quad \left. + \Delta^{(l)}\left(\lambda+\mu\right)\Delta^{(l)}\left(-\lambda-\mu\right)\right] \left|f_{YY_i}\left(\lambda\right)\right|^2, \qquad (5.49)
\end{aligned}$$

and $Var\left(I_Y^{(n)}\left(\lambda\right)\right)$ and $Var\left(I_{Y_i}^{(l)}\left(\mu\right)\right)$ can be calculated from the above given covariance formula. The covariance and variance terms rely upon certain joint spectral densities. Those spectral densities can be expressed in closed form as follows. Let us firstly consider the cross spectrum corresponding to the full sample and *jth* sub-sample, $f_{YY_j}\left(\lambda\right)$. Suppose we consider the jackknife approach using non-overlapping subsamples. Then, the general definition of spectral density gives that

$$\begin{aligned}
f_{YY_j}\left(\lambda\right) &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \exp\left(-ik\lambda\right) \kappa\left(Y_{t+k}, Y_{t+(j-1)l}\right) \\
&= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \exp\left(-ik\lambda\right) \gamma\left(k-\left(j-1\right)l\right) \\
&= \frac{\exp\left(-i\left(j-1\right)l\lambda\right)}{2\pi} \sum_{k=-\infty}^{\infty} \exp\left(-i\left(k-\left(j-1\right)l\right)\lambda\right) \gamma\left(k-\left(j-1\right)l\right)
\end{aligned}$$

$$= \exp\left(-i\left(j-1\right)l\lambda\right) f_{YY}\left(\lambda\right).$$

Similarly, for moving-block subsamples we have the relationship $f_{YY_j}\left(\lambda\right) = \exp\left(-i\left(j+l-1\right)\lambda\right) f_{YY}\left(\lambda\right)$ and $f_{Y_jY_k}\left(\lambda\right) = \exp\left(-i\left(j-k\right)l\lambda\right) f_{YY}\left(\lambda\right)$.

Lemma 2 of Yajima (1989) immediately gives that,

$$f_{YYYY}\left(\lambda,-\lambda,\mu\right) = \frac{1}{\left(2\pi\right)^3} b\left(\lambda\right) b\left(-\lambda\right) b\left(\mu\right) b\left(-\mu\right) f_{\varepsilon\varepsilon\varepsilon\varepsilon}\left(\lambda,-\lambda,\mu\right), \tag{5.50}$$

where $b\left(\lambda\right) = \sum_{j=0}^{\infty} b_j \exp\left(\imath j\omega\right)$ with

$$b_j = \sum_{r=0}^{j} \frac{k\left(j-r\right)\Gamma\left(r+d\right)}{\Gamma\left(r+1\right)\Gamma\left(d\right)}, \tag{5.51}$$

and $k\left(z\right)$ is the transfer function of a stable and invertible autoregressive moving average (ARMA) process such that $\sum_{j=0}^{\infty}\left|k\left(j\right)\right| < \infty$. Here,

$$f_{\varepsilon\varepsilon\varepsilon\varepsilon}\left(\lambda,-\lambda,\mu\right) = \sum_{u_1=-\infty}^{\infty}\sum_{u_2=-\infty}^{\infty}\sum_{u_3=-\infty}^{\infty} \exp\left(-i\left(\lambda u_1 - \lambda u_2 + \mu u_3\right)\right) \kappa_{\varepsilon\varepsilon\varepsilon\varepsilon}\left(u_1,u_2,u_3\right),$$

where

$$\begin{aligned}
\kappa_{\varepsilon\varepsilon\varepsilon\varepsilon}\left(u_1,u_2,u_3\right) &= \kappa\left(\varepsilon_{t+u_1},\varepsilon_{t+u_2},\varepsilon_{t+u_3},\varepsilon_t\right) \\
&= E\left(\varepsilon_{t+u_1}\varepsilon_{t+u_2}\varepsilon_{t+u_3}\varepsilon_t\right) - E\left(\varepsilon_{t+u_1}\varepsilon_{t+u_2}\right)E\left(\varepsilon_{t+u_3}\varepsilon_t\right) \\
&\quad - E\left(\varepsilon_{t+u_2}\varepsilon_{t+u_3}\right)E\left(\varepsilon_{t+u_1}\varepsilon_t\right) - E\left(\varepsilon_{t+u_1}\varepsilon_{t+u_3}\right)E\left(\varepsilon_{t+u_2}\varepsilon_t\right).
\end{aligned}$$

Suppose the errors are *i.i.d* normal random variables with zero mean and a constant variance σ^2,

$$\begin{aligned}
\kappa_{\varepsilon\varepsilon\varepsilon\varepsilon}\left(u_1,u_2,u_3\right) &= \begin{cases} E\left(\varepsilon_t^4\right) - 3\left(E\left(\varepsilon_t^2\right)\right)^2 & if\ u_1 = u_2 = u_3 = 0 \\ 0 & otherwise \end{cases} \\
&= \begin{cases} 3\sigma^4 & if\ u_1 = u_2 = u_3 = 0 \\ 0 & otherwise \end{cases}.
\end{aligned}$$

Then $f_{YYYY}(\lambda,-\lambda,\mu)$ simplified as follows.

$$f_{YYYY}(\lambda,-\lambda,\mu)=\frac{3\sigma^4}{(2\pi)^3}b(-\lambda)\,b(\lambda)\,b(\mu)\,b(-\mu)=\frac{3}{2\pi}f_{YY}(\lambda)\,f_{YY}(\mu)\,, \tag{5.52}$$

since $f_{YY}(\lambda)=\dfrac{\sigma^2}{2\pi}b(\lambda)\,b(-\lambda)\,.$

Now let us consider $f_{YYY_jY_j}(\lambda,-\lambda,\mu)$,

$$\begin{aligned}
f_{YYY_jY_j}(\lambda,-\lambda,\mu) &= \frac{1}{(2\pi)^3}\sum_{u_1=-\infty}^{\infty}\sum_{u_2=-\infty}^{\infty}\sum_{u_3=-\infty}^{\infty}\exp\left(-\imath\left(\lambda u_1-\lambda u_2+\mu u_3\right)\right)\\
&\quad\times\kappa\left(Y_{t+u_1},Y_{t+u_2},Y_{t+(j-1)l+u_3},Y_{t+(j-1)l}\right)\\
&= \frac{1}{(2\pi)^3}\sum_{u_1=-\infty}^{\infty}\sum_{u_2=-\infty}^{\infty}\sum_{u_3=-\infty}^{\infty}\exp\left(-\imath\left(\lambda\left(u_1-(j-1)\,l\right)-\lambda\left(u_2-(j-1)\,l\right)+\mu u_3\right)\right)\\
&\quad\times\kappa\left(Y_{t-(j-1)l+u_1},Y_{t-(j-1)l+u_2},Y_{t+u_3},Y_t\right)\\
&= f_{YYYY}(\lambda,-\lambda,\mu)\,.
\end{aligned}$$

The covariance and variance terms in (5.49) can thus be simplified as follows.

$$\begin{aligned}
Cov\left(I_Y^{(n)}(\lambda),I_{Y_i}^{(l)}(\mu)\right) &\approx \frac{3}{l}f_{YY}(\lambda)\,f_{YY}(\mu)+\frac{1}{l^2}\Big[\Delta^{(l)}(-\lambda+\mu)\,\Delta^{(l)}(\lambda-\mu)\\
&\quad+\Delta^{(l)}(\lambda+\mu)\,\Delta^{(l)}(-\lambda-\mu)\Big]\left(f_{YY}(\lambda)\right)^2,\\
Var\left(I_Y^{(n)}(\lambda)\right) &\approx \left[1+\frac{3}{l}+\frac{1}{l^2}\Delta^{(l)}(2\lambda)\,\Delta^{(l)}(-2\lambda)\right]\left(f_{YY}(\lambda)\right)^2.
\end{aligned}$$

Hence, the correlation given in (5.48) can be simplified as,

$$\rho\approx\frac{\frac{3}{l}+\frac{1}{l^2}\left[\Delta^{(l)}(-\lambda+\mu)\,\Delta^{(l)}(\lambda-\mu)+\Delta^{(l)}(\lambda+\mu)\,\Delta^{(l)}(-\lambda-\mu)\right]\frac{f_{YY}(\lambda)}{f_{YY}(\mu)}}{\sqrt{\left(1+\frac{3}{l}+\frac{1}{l^2}\Delta^{(l)}(2\lambda)\,\Delta^{(l)}(-2\lambda)\right)}\sqrt{\left(1+\frac{3}{l}+\frac{1}{l^2}\Delta^{(l)}(2\mu)\,\Delta^{(l)}(-2\mu)\right)}}.$$

5.B.2 Positiveness of the principle minors of the bordered Hessian matrix

Here we show that for every $m\in\mathbb{N}$, $\left|\mathbf{H}^B_{(m+3)\times(m+3)}\right|>0$ using mathematical induction. For our convenience, we assume that

$$\varphi_{\min}\left(\mathbf{H}^B_{(m+3)\times(m+3)}\right)>(m+3)^2\,\frac{12N_l}{\pi^2}, \tag{5.53}$$

where $\varphi_{\min}(\mathbf{A})$ is the minimum eigenvalue corresponding to the matrix $\mathbf{A}$.

Let us start with $m = 1$. Then the determinant of the first minor of the bordered Hessian matrix, $\mathbf{H}^B_{4\times 4}$, is,

$$\begin{aligned}
\left|\mathbf{H}^B_{4\times 4}\right| &= \begin{vmatrix} 0 & 0 & -m^2\frac{N_l^2}{l^2} \\ 1 & \frac{N_n^2}{n^2} & -2c^*_{n,1} \\ -1 & -m^2\frac{N_l^2}{l^2} & \frac{\pi^2}{12N_l} \end{vmatrix} + \begin{vmatrix} 0 & 0 & \frac{N_n^2}{n^2} \\ 1 & \frac{N_n^2}{n^2} & \frac{\pi^2}{12N_n} \\ -1 & -m^2\frac{N_l^2}{l^2} & -2c^*_{n,1} \end{vmatrix} \\
&= -m^2\frac{N_l^2}{l^2}\left(-m^2\frac{N_l^2}{l^2} + \frac{N_n^2}{n^2}\right) + \frac{N_n^2}{n^2}\left(-m^2\frac{N_l^2}{l^2} + \frac{N_n^2}{n^2}\right) \\
&= \left(\frac{N_n^2}{n^2} - m^2\frac{N_l^2}{l^2}\right)^2 > 0.
\end{aligned}$$

That is, $\left|\mathbf{H}^B_{(m+3)\times(m+3)}\right| > 0$ for $m = 1$.

Suppose that $\left|\mathbf{H}^B_{(m+3)\times(m+3)}\right| > 0$ is true for $m = k$, then we need to show that it is true for $m = k+1$. To do so, we consider the partition of $\mathbf{H}^B_{(k+4)\times(k+4)}$ is as follows;

$$\mathbf{H}^B_{(k+4)\times(k+4)} = \begin{pmatrix} \mathbf{H}^B_{(k+3)\times(k+3)} & \mathbf{U} \\ \mathbf{U}^T & \frac{\pi^2}{12N_l} \end{pmatrix},$$

where $\mathbf{U} = \begin{bmatrix} -1 & -(k+1)^2\frac{N_l^2}{l^2} & -2c^*_{n,k+1} & 2c^\dagger_{1,k+1} & \dots & 2c^\dagger_{k,k+1} \end{bmatrix}^\top$. The determinant of $\mathbf{H}^B_{(k+4)\times(k+4)}$ is,

$$\left|\mathbf{H}^B_{(k+4)\times(k+4)}\right| = \left|\mathbf{H}^B_{(k+3)\times(k+3)}\right|\left(\frac{\pi^2}{12N_l} - \mathbf{U}^\top\left(\mathbf{H}^B_{(k+3)\times(k+3)}\right)^{-1}\mathbf{U}\right) \tag{5.54}$$

Since $\left|\mathbf{H}^B_{(k+3)\times(k+3)}\right| > 0$, we have that $\mathbf{U}^\top\left(\mathbf{H}^B_{(k+3)\times(k+3)}\right)^{-1}\mathbf{U} > 0$. Therefore,

$$\mathbf{U}^\top\left(\mathbf{H}^B_{(k+3)\times(k+3)}\right)^{-1}\mathbf{U} \leq \frac{1}{\varphi_{\min}\left(\mathbf{H}^B_{(k+3)\times(k+3)}\right)}\max_{\mathbf{U}\in\mathbb{R}^{k+3}\setminus\{\mathbf{0}\}}\mathbf{U}^\top\mathbf{U} < \frac{\pi^2}{12N_l}.$$

as $\max_{\mathbf{U}\in\mathbb{R}^{k+3}\setminus\{\mathbf{0}\}}\mathbf{U}^\top\mathbf{U} = 1$. Hence this completes the proof.

5.C Appendix: Additional simulation results

Table 5.10: Bias estimates of the unadjusted LPR estimator, the optimal jackknife estimator based on 2,3,4,6,8 non-overlapping (NO) sub-samples, the optimal jackknife estimator based on 2 moving block (MB) sub-samples, both versions of the GS estimator, and the prefiltered sieve bootstrap estimator, for the DGP: ARFIMA$(1, d_0, 0)$. The estimates are obtained under Gaussian innovations, with $\alpha = 0.5$.

ϕ_0	d_0	n	$\widehat{d}_n$	$\widehat{d}_{J,2}^{Opt-NO}$	$\widehat{d}_{J,3}^{Opt-NO}$	$\widehat{d}_{J,4}^{Opt-NO}$	$\widehat{d}_{J,6}^{Opt-NO}$	$\widehat{d}_{J,8}^{Opt-NO}$	$\widehat{d}_{J,2}^{Opt-MB}$	$\widehat{d}_1^{GS}$	$\widehat{d}_1^{Opt-GS}$	$\widehat{d}^{PFBS}$
-0.9	-0.25	96	0.5870	**0.5019**	0.5103	0.5246	0.5489	0.5681	0.5222	0.5322	0.5154	0.5256
		576	0.2746	**0.2216**	0.2636	0.2715	0.2738	0.2744	0.2254	0.2394	0.2267	0.2225
	0	96	0.5701	**0.4944**	0.5035	0.5174	0.5238	0.5422	0.5823	0.5261	0.5129	0.5181
		576	0.2692	**0.2101**	0.2196	0.2239	0.2311	0.2401	0.2701	0.2335	0.2248	0.2182
	0.25	96	0.5805	**0.4760**	0.4744	0.4776	0.4861	0.4937	0.4837	0.5288	0.5173	0.5167
		576	0.2725	**0.2168**	0.2252	0.2352	0.2472	0.2545	0.2745	0.2374	0.2264	0.2181
	0.45	96	0.5813	**0.5206**	0.5309	0.5459	0.5583	0.5653	0.5853	0.5336	0.5218	0.5302
		576	0.2808	**0.2127**	0.2217	0.2301	0.2478	0.2537	0.2837	0.2468	0.2236	0.2194
-0.4	-0.25	96	0.2768	**0.0295**	0.0673	0.0730	0.0765	0.0767	0.0297	0.0468	0.0314	0.0359
		576	0.0183	**0.0009**	0.0010	0.0116	0.0161	0.0175	0.0175	0.0104	0.0025	0.0018
	0	96	0.0673	0.0284	0.0377	0.0431	0.0688	0.0782	0.0682	0.0016	**0.0008**	0.0360
		576	0.0117	0.0008	0.0037	0.0087	0.0121	0.0119	0.0119	0.0002	**0.0001**	-0.0012
	0.25	96	0.0699	**0.0334**	0.0411	0.0557	0.0633	0.0719	0.0719	0.0420	0.0354	0.0378
		576	0.0150	**0.0005**	0.0013	0.0024	0.0062	0.0083	0.0163	0.0035	0.0011	0.0008
	0.45	96	0.0782	**0.0395**	0.0265	0.0322	0.0532	0.0611	0.0811	0.0540	0.0421	0.0440
		576	0.0241	**0.0018**	0.0068	0.0039	0.0016	0.0026	0.0263	-0.0135	0.0095	0.0024
0.4	-0.25	96	-0.0037	**-0.0013**	-0.0071	-0.0051	0.0089	0.0023	-0.0023	-0.0433	-0.0121	-0.0095
		576	-0.0078	**-0.0007**	-0.0013	-0.0021	0.0051	-0.0067	-0.0067	-0.0100	-0.0022	-0.0009
	0	96	-0.0146	-0.0063	-0.0125	-0.0143	-0.0153	-0.0160	-0.0150	-0.0342	**-0.0042**	-0.0075
		576	-0.0019	**-0.0008**	-0.0023	-0.0021	-0.0020	-0.0020	-0.0020	-0.0028	-0.0009	-0.0040
	0.25	96	-0.0126	**-0.0008**	-0.0043	-0.0053	-0.0113	-0.0120	-0.0120	-0.0385	-0.0024	-0.0009
		576	0.0013	**-0.0002**	-0.0032	-0.0091	-0.0040	-0.0023	-0.0023	-0.0066	-0.0011	-0.0020
	0.45	96	-0.0022	-0.0012	-0.0032	-0.0046	-0.0007	-0.0007	**-0.0006**	-0.0489	-0.0015	-0.0066
		576	0.0105	-0.0004	-0.0018	-0.0027	-0.0055	-0.0047	-0.0124	-0.0166	-0.0007	-0.0007

Table 5.11: RMSE estimates of the unadjusted LPR estimator, the optimal jackknife estimator based on 2,3,4,6,8 non-overlapping (NO) sub-samples, the optimal jackknife estimator based on 2 moving block (MB) sub-samples, both versions of the GS estimator, and the prefiltered sieve bootstrap estimator, for the DGP: ARFIMA($1, d_0, 0$). The estimates are obtained under Gaussian innovations, with $\alpha = 0.5$.

ϕ_0	d_0	n	$\widehat{d}_n$	$\widehat{d}_{J,2}^{Opt-NO}$	$\widehat{d}_{J,3}^{Opt-NO}$	$\widehat{d}_{J,4}^{Opt-NO}$	$\widehat{d}_{J,6}^{Opt-NO}$	$\widehat{d}_{J,8}^{Opt-NO}$	$\widehat{d}_{J,2}^{Opt-MB}$	$\widehat{d}_1^{GS}$	$\widehat{d}_1^{Opt-GS}$	$\widehat{d}^{PFBS}$
-0.9	-0.25	96	1.7257	1.7330	1.7259	1.7157	1.7078	**1.7069**	1.7269	1.7375	1.7171	1.7896
		576	0.9774	0.9811	0.9715	0.9688	0.9579	**0.9477**	0.9777	0.9954	0.9594	0.9963
	0	96	1.7217	1.7533	1.7557	1.7383	1.7251	**1.7226**	1.7236	1.7350	1.7283	1.7748
		576	0.9750	0.9233	0.9876	0.9796	0.9766	**0.9713**	0.9756	0.9941	0.9813	0.9916
	0.25	96	1.7225	1.7147	1.7709	1.7459	1.7271	**1.7207**	1.7251	1.7368	1.7151	1.7775
		576	0.9768	0.9737	0.9803	0.9850	0.9796	**0.9700**	0.9779	0.9954	0.9842	0.9978
	0.45	96	1.7230	1.7653	1.7825	1.7515	1.7285	**1.7260**	1.7261	1.7391	1.7267	1.7063
		576	0.9809	0.9809	0.9959	0.9929	0.9850	**0.9716**	0.9825	0.9979	0.9829	0.9946
-0.4	-0.25	96	0.3243	0.3874	0.3368	0.3297	0.3253	**0.3229**	0.3249	0.3723	0.3599	0.4553
		576	0.1623	0.1673	0.1703	0.1646	0.1630	**0.1626**	0.1626	0.1926	0.1751	0.2030
	0	96	0.3217	0.3293	0.3288	0.3194	0.3131	**0.3219**	0.3225	0.3708	0.3466	0.4321
		576	0.1611	0.1602	0.1573	0.1541	0.1520	**0.1514**	0.1614	0.1912	0.1782	0.1965
	0.25	96	0.3238	0.3104	0.3068	0.3042	0.2957	**0.2829**	0.3248	0.3718	0.3434	0.4210
		576	0.1619	0.1556	0.1549	0.1510	0.1462	**0.1424**	0.1624	0.1917	0.1729	0.1931
	0.45	96	0.3257	0.4680	0.4551	0.4489	0.4281	**0.3167**	0.3270	0.3731	0.3391	0.4226
		576	0.1637	0.1608	0.1538	0.1497	0.1456	**0.1445**	0.1645	0.1930	0.1656	0.1890
0.4	-0.25	96	0.3155	0.3235	0.3208	0.3220	0.3166	**0.3112**	0.3161	0.3741	0.3215	0.4624
		576	0.1620	0.1573	0.1512	0.1447	0.1428	**0.1423**	0.1623	0.1921	0.1762	0.2061
	0	96	0.3148	0.3067	0.3000	0.2914	0.2859	**0.2854**	0.3154	0.3725	0.3174	0.4284
		576	0.1608	0.1596	0.1519	0.1437	0.1417	**0.1411**	0.1611	0.1911	0.1717	0.1963
	0.25	96	0.3157	0.3133	0.3110	0.3023	0.2968	**0.2963**	0.3163	0.3745	0.3254	0.4103
		576	0.1611	0.1670	0.1619	0.1544	0.1522	**0.1515**	0.1615	0.1919	0.1770	0.1919
	0.45	96	0.3167	0.3150	0.3140	0.3111	0.3079	**0.3023**	0.3173	0.3755	0.3315	0.4062
		576	0.1625	0.1580	0.1508	0.1471	0.1438	**0.1430**	0.1630	0.1931	0.1543	0.1886

Table 5.12: Bias estimates of the unadjusted LPR estimator, the optimal jackknife estimator based on 2,3,4,6,8 non-overlapping (NO) sub-samples, the optimal jackknife estimator based on 2 moving block (MB) sub-samples, both versions of the GS estimator, and the prefiltered sieve bootstrap estimator, for the DGP: ARFIMA$(0, d_0, 1)$. The estimates are obtained under Gaussian innovations, with $\alpha = 0.5$.

θ_0	d_0	n	$\widehat{d}_n$	$\widehat{d}_{J,2}^{Opt-NO}$	$\widehat{d}_{J,3}^{Opt-NO}$	$\widehat{d}_{J,4}^{Opt-NO}$	$\widehat{d}_{J,6}^{Opt-NO}$	$\widehat{d}_{J,8}^{Opt-NO}$	$\widehat{d}_{J,2}^{Opt-MB}$	$\widehat{d}_1^{GS}$	$\widehat{d}_1^{Opt-GS}$	$\widehat{d}^{PFBS}$
-0.9	-0.25	96	-0.3908	**-0.3222**	-0.3353	-0.3431	-0.3555	-0.3635	-0.3735	-0.3429	-0.3341	-0.3239
		576	-0.2356	**-0.1749**	-0.1868	-0.1866	-0.1996	-0.2074	-0.2153	-0.2013	-0.1882	-0.2276
	0	96	-0.3647	**-0.3215**	-0.3356	-0.3405	-0.3681	-0.3666	-0.3147	-0.3413	-0.3334	-0.3435
		576	-0.2276	**-0.1678**	-0.1777	-0.1721	-0.1795	-0.1885	-0.1927	-0.2025	-0.1759	-0.2014
	0.25	96	-0.3801	**-0.3130**	-0.3228	-0.3373	-0.3419	-0.3812	-0.3314	-0.3458	-0.3242	-0.3295
		576	-0.2665	**-0.2149**	-0.2548	-0.2632	-0.2657	-0.2663	-0.2315	-0.2307	-0.2169	-0.2382
	0.45	96	-0.3710	**-0.3066**	-0.3169	-0.23202	-0.3413	-0.3713	-0.3245	-0.3543	-0.3315	-0.3116
		576	-0.2565	**-0.2140**	-0.2263	-0.2365	-0.2434	-0.2554	-0.2260	-0.2398	-0.2214	-0.2323
-0.4	-0.25	96	-0.0426	**-0.0164**	-0.0211	-0.0270	-0.0358	-0.0445	-0.0344	-0.0299	-0.0185	-0.0456
		576	-0.0003	**-0.0014**	-0.0018	-0.0020	-0.0036	-0.0008	-0.0008	0.0099	-0.0047	-0.0017
	0	96	-0.0663	**-0.0324**	-0.0466	-0.0520	-0.0679	-0.0673	-0.0573	-0.0441	-0.0354	-0.0426
		576	-0.0112	**-0.0066**	-0.0131	-0.0182	-0.0216	-0.0114	-0.0094	-0.0126	-0.0086	-0.0061
	0.25	96	-0.0653	**-0.0148**	-0.0551	-0.0611	-0.0650	-0.0652	-0.0451	-0.0314	-0.0223	-0.0335
		576	-0.0081	**-0.0010**	-0.0012	-0.0020	-0.0057	-0.0072	-0.0062	-0.0041	-0.0025	-0.0039
	0.45	96	-0.0542	**-0.0155**	-0.0164	-0.0215	-0.0362	-0.0532	-0.0332	-0.0285	-0.0186	-0.0247
		576	0.0012	**-0.0007**	-0.0024	-0.0049	-0.0059	0.0030	0.0010	0.0110	-0.0074	-0.0008
0.4	-0.25	96	0.0260	**0.0044**	0.0091	0.0182	0.0248	0.0254	0.0154	0.0120	0.0084	0.0094
		576	0.0096	**0.0006**	0.0011	0.0021	0.0070	0.0086	0.0076	0.0082	0.0015	0.0009
	0	96	0.0151	**0.0026**	0.0118	0.0138	0.0158	0.0156	0.0106	0.0076	0.0035	0.0127
		576	0.0024	**0.0002**	0.0009	0.0025	0.0026	0.0025	0.0025	0.0023	0.0014	0.0039
	0.25	96	0.0177	**0.0017**	0.0115	0.0195	0.0203	0.0192	0.0092	0.0057	0.0026	0.0177
		576	0.0057	**0.0008**	0.0024	0.0039	0.0046	0.0068	0.0038	0.0041	0.0011	0.0022
	0.45	96	0.0274	**0.0085**	0.0685	0.0476	0.0316	0.0298	0.0198	0.0173	0.0086	0.0230
		576	0.0149	**0.0006**	0.0036	0.0022	0.0021	0.0170	0.0070	0.0142	0.0014	0.0034

Table 5.13: RMSE estimates of the unadjusted LPR estimator, the optimal jackknife estimator based on 2,3,4,6,8 non-overlapping (NO) sub-samples, the optimal jackknife estimator based on 2 moving block (MB) sub-samples, both versions of the GS estimator, and the prefiltered sieve bootstrap estimator, for the DGP: ARFIMA$(0, d_0, 1)$. The estimates are obtained under Gaussian innovations, with $\alpha = 0.5$.

θ_0	d_0	n	$\widehat{d}_n$	$\widehat{d}_{J,2}^{Opt-NO}$	$\widehat{d}_{J,3}^{Opt-NO}$	$\widehat{d}_{J,4}^{Opt-NO}$	$\widehat{d}_{J,6}^{Opt-NO}$	$\widehat{d}_{J,8}^{Opt-NO}$	$\widehat{d}_{J,2}^{Opt-MB}$	$\widehat{d}_1^{GS}$	$\widehat{d}_1^{Opt-GS}$	$\widehat{d}^{PFBS}$
-0.9	-0.25	96	0.8762	0.8973	0.8927	0.8836	0.8796	**0.8681**	0.8815	0.8819	0.8757	0.8830
		576	0.6692	0.6536	0.6456	0.6447	0.6311	**0.6300**	0.6649	0.6954	0.6513	0.6485
	0	96	0.8517	0.8446	0.8430	0.8267	0.8147	**0.8134**	0.8482	0.8474	0.8341	0.8303
		576	0.6744	0.6521	0.6466	0.6388	0.6260	**0.6150**	0.6274	0.6539	0.6215	0.6223
	0.25	96	0.9215	0.9473	0.9212	0.9112	0.9036	**0.9027**	0.9153	0.9346	0.9154	0.9869
		576	0.6744	0.6709	0.6690	0.6560	0.6550	**0.6447**	0.6678	0.6942	0.6597	0.6955
	0.45	96	0.9165	0.9355	0.9257	0.9212	0.9176	**0.9052**	0.9060	0.9298	0.9125	0.9628
		576	0.6719	0.6618	0.6536	0.6518	0.6498	**0.6419**	0.6451	0.6938	0.6550	0.6913
-0.4	-0.25	96	0.3192	0.3127	0.3086	0.3078	0.3007	**0.2901**	0.3201	0.3728	0.3284	0.4744
		576	0.1622	0.1565	0.1525	0.1453	0.1331	**0.1325**	0.1625	0.1929	0.1659	0.2080
	0	96	0.3214	0.3179	0.3083	0.2990	0.2828	**0.2722**	0.3222	0.3725	0.3418	0.4328
		576	0.1612	0.1703	0.1704	0.1642	0.1622	**0.1622**	0.1616	0.1913	0.1706	0.1960
	0.25	96	0.3221	0.3485	0.3354	0.3279	0.3232	**0.3215**	0.3227	0.3731	0.3583	0.4101
		576	0.1613	0.1520	0.1508	0.1442	0.1322	**0.1217**	0.1617	0.1916	0.1734	0.1913
	0.45	96	0.3212	0.3157	0.3033	0.2962	0.2920	**0.2817**	0.3217	0.3732	0.3569	0.4013
		576	0.1622	0.1497	0.1461	0.1359	0.1233	**0.1126**	0.1626	0.1929	0.1594	0.1880
0.4	-0.25	96	0.3164	0.3488	0.3304	0.3223	0.3174	**0.3169**	0.3170	0.3721	0.3458	0.4578
		576	0.1617	0.1512	0.1407	0.1344	0.1225	**0.1120**	0.1620	0.1921	0.1641	0.2066
	0	96	0.3148	0.3072	0.3001	0.2915	0.2859	**0.2754**	0.3154	0.3708	0.3257	0.4298
		576	0.1607	0.1595	0.1497	0.1436	0.1316	**0.1210**	0.1610	0.1911	0.1549	0.1992
	0.25	96	0.3164	0.3037	0.2940	0.2840	0.2777	**0.2671**	0.3171	0.3722	0.3338	0.4153
		576	0.1613	0.1597	0.1428	0.1349	0.1224	**0.1118**	0.1618	0.1918	0.1644	0.1956
	0.45	96	0.3175	0.3018	0.2990	0.2867	0.2791	**0.2683**	0.3183	0.3728	0.3048	0.4123
		576	0.1627	0.1421	0.1302	0.1278	0.1243	**0.1133**	0.1633	0.1930	0.1591	0.1917

Chapter 6

Conclusion

6.1 Introduction

The statistical analysis of strongly dependent, or long range dependent, data is at the core of the economics and finance disciplines, amongst others. Such time series cannot be modelled by the conventional autoregressive moving average [ARMA] or autoregressive integrated moving average [ARIMA] models, as the autocovariances tend to zero like a power function and slowly enough for their sum to diverge. The statistical model that plays the central role in these studies is the *fractionally integrated autoregressive moving average* [ARFIMA] model, which was introduced by Granger and Joyeux (1980) and Hosking (1981). The model is a natural extension of the autoregressive integrated moving average [ARIMA] model by permitting the order of integration to be any real number. The ARFIMA(p, d, q) model takes the form of $\phi(L)(1-L)^d(y_t - \mu) = \theta(L)\varepsilon_t$, where y_t is the t^{th} observation of the time series, μ is the process mean, $\phi(L)$ and $\theta(L)$ are the autoregressive and moving average components of orders p and q respectively, and $\{\varepsilon_t\}$ is a white noise process with zero mean and constant variance. Here, d is the order of integration that will be known as the fractional differencing parameter hereinafter. Throughout the thesis we impose the restriction the true value of d lies in the range, $d \in (-0.5, 0.5)$, and that the characteristic roots of the components $\phi(L)$ and $\theta(L)$ lie within the unit circle. Under this setting, the process $\{y_t\}$ is said to be is stationary. Interest in fractionally integrated processes stems from the fact that they allow for the full continuum of memory properties, including the long memory that is the primary focus here. For reviews of early work on the use of fractional processes see Beran (1994) and Baillie (1996), with more recent developments being discussed in Beran *et al.* (2013), Pipiras and Taqqu (2017) and Hassler (2018).

The thesis aims to make methodological and theoretical contributions in the context of the stationary class of ARFIMA models. Statistical inference based on parametric estimation techniques typically relies on the assumption that the assumed model is correctly specified. In this case, parametric estimators of the dynamics parameters are $\sqrt{n}-$ consistent for the true values, and asymptotic normality is achieved (see, Fox and Taqqu, 1986, Dahlhaus, 1989, Hualde and Robinson, 2011 and Lieberman *et al.*, 2012, among others). This assumption of correct specification of the model for a given data generating process [DGP] is violated in practice, as the true values of p and q – the number of AR and MA components – are not known. Assessing the impact of incorrect specification of the model is a non-trivial task. The thesis presents a complete analysis of the repercussions of mis-specifying the model, both asymptotically and in finite samples.

While the convention in the area has been to adopt complete parametric specifications for the dynamics in the time series, semi-parametric approaches have featured of late; for example, see Geweke and Porter-Hudak (1983) and Robinson (1995a) among others. These methods are most widely used, as they are not influenced by mis-specification of the short memory dynamics. However, they exhibit large finite sample bias and the bias-corrected semi-parametric estimators are inefficient compared to their (correctly-specified) parametric counterparts due to the trade-off between bias-reduction and variance. Therefore, the other main contribution of the thesis is to develop an optimally bias-corrected semi-parametric estimator, by controlling both bias-reduction and the increase in variance.

6.2 Summary and discussion

Chapter 1 introduces the ARFIMA model and provides an outline of the thesis. A review of the class of ARFIMA models and the existing parametric and semi-parametric estimation methods is given in Chapter 2. In what follows we shall summarize the contribution of the three main chapters.

Chapter 3:

This chapter quantifies the impact of mis-specification of short memory dynamics in stationary fractionally integrated models on four alternative parametric estimators, namely, frequency domain maximum likelihood [FML], discrete Whittle [DWH], time domain maximum likelihood [TML] and conditional sum of squares [CSS]. Under common mis-specification, we show that all four parametric estimators converge to the same pseudo-true value, which is different from the true value of the vector parameter. A closed-form expression for the first order conditions that define the pseudo-true value is provided. The rate of convergence and the limiting distribution are case-specific, depending on the degree of mis-specification measured by the difference between the true value of the fractional differencing parameter (d_0) and its pseudo-true value (d_1), denoted as $d^* = d_0 - d_1$. If $0.25 < d^* < 0.5$, the limiting distribution is non-Gaussian and the rate of convergence is slower than $\sqrt{n}$, and depends on d^*, when n is sample size. If $d^* = 0.25$, asymptotic normality is achieved with a rate of convergence that is a function of the true spectral density and the spectral density of the fitted model and the sample size. Whenever $-1 < d^* < 0.25$, both asymptotic normality and a $\sqrt{n}$

rate of convergence hold. Under mis-specification, all four parametric estimation methods are asymptotically equivalent. The above mentioned theoretical results are established under the assumption that the process mean is known, but without assuming Gaussianity for the true process.

A simulation exercise is used to investigate the finite sample performance of the alternative methods in terms of estimating the pseudo-true parameter d_1. Provided that the process mean is known, the CSS estimator uniformly outperforms the other three estimators, while the FML estimator is the least efficient estimator. When the mean is estimated by the sample mean, the DWH estimator is preferred among the four parametric estimators.

Chapter 4:

This chapter investigates the repercussions of mis-specifying the short memory dynamics in stationary fractionally integrated models when the process mean is not known, by extending the theory established in Chapter 3. We establish the theory only for the EWH, TML and CSS estimation methods, as FML and DWH are invariant to the mean. We show that the three estimators converge to the same pseudo-true value that is identified under the known mean case. The limiting distribution is identical to that of FML and DWH estimators. Two examples of mean estimators, namely sample mean and the best linear unbiased estimator [BLUE] are considered. Although the sample mean is unaffected by the specification of the model, the BLUE is. However, BLUE is consistent for the true mean regardless of the model specification. In a Monte-Carlo experiment with the sample mean and the BLUE being the estimators of mean, it is observed that DWH performs best in estimating d_1, and this conclusion holds for both the Gaussian and Chi-squared error processes.

Chapter 5:

This chapter develops a new bias-correction method for the log-periodogram regression [LPR] estimator by focusing on bias reduction while minimizing the concurrent increase in variance. We adopt a jackknife technique for such bias-adjustment. The resultant jackknife based bias-corrected LPR estimator is a linear combination of the full sample and sub-sample LPR estimators, where sub-samples are drawn using either the non-overlapping or moving-block method. The weights in the linear combination are 'optimal' in the sense of producing bias reduction with the minimum increase in variance, and the associated jackknife estimator referred to as 'optimal' accordingly. The optimal weights are functions of two types of covariance terms: (i) the covariance between the full- and sub-sample log-periodogram ordinates, and, (ii) the covariance between distinct sub-sample log-periodogram ordinates. We derive these terms as follows. Firstly, we derive the cumulants of the discrete Fourier transforms [DFT] associated with the full sample and the sub-samples, and derive the covariances between the periodograms using the cumulants. Under certain regularity conditions on the spectral density of the underlying process, we show that the periodograms associated with the full sample and the sub-samples are asymptotically independent Chi-squared random variables. Using the distributional results, we find the joint probability distribution of the log-periodogram associated with the full sample and any sub-sample (or for two distinct sub-samples). Using the joint distribution of the log-periodograms, we can derive the moment generating function of the joint distribution. This leads to the derivation of the above mentioned covariance terms. We show that the optimally bias-corrected jackknife LPR estimator is consistent for d_0 and asymptotically normal, with an asymptotic variance that is same as that of the original LPR estimator. That is, there is no loss in asymptotic efficiency compared to the LPR estimator.

In a simulation study, the optimally bias-corrected jackknife LPR estimator performs better than the alternative bias-reduced LPR estimators of Guggenberger and Sun (2006) and Poskitt *et al.* (2016), in terms of bias and root mean squared error [RMSE]. This result holds when the true values of all the dynamic parameters are used in evaluating the optimal weights. In practice the true values are not known and hence the optimal jackknife method is infeasible. To this end, we introduce a feasible version of the estimator by implementing the jackknife estimator using an iterative procedure. The feasible version of the jackknife estimator does not consistently outperform the feasible version of Guggenberger and Sun or Poskitt *et al.*, in terms of bias or RMSE, but still remains competitive and is sometimes the least biased estimator of all.

6.3 Future directions

In this section we briefly discuss some new research ideas that have emerged while investigating the research issues presented in the thesis. These ideas are not pursued here, but, as discussed below, we expect that they would produce some challenging topics for future research.

In Chapters 3 and 4, we developed the asymptotic theory for parametric estimators when fitting a mis-specified ARFIMA model with an incorrect choice of the number of short run parameters. The technical results are established under the assumption that both the true DGP and the fitted (mis-specified) model are stationary, where an ARFIMA model is stationary only if $d < 0.5$ (and the short memory dynamics are in the stationary and invertible region) and is non-stationary if $d \geq 0.5$. However, in practice this strong assumption is often violated and non-stationarity has been documented in many financial time series such as treasury bills and interest rates (see Škare and Stjepanović, 2013).[1]. Besides, a non-stationary ARFIMA process can be easily mis-specified as a non-stationary ARIMA model, and vice versa, due to the slow decay observed in the autocovariances of both models. Hence, relaxing the assumption of stationarity in the true DGP and/or the fitted model is an important avenue to explore. Doing so would lead to several fundamental questions: (i) how is the degree of mis-specification measured?, (ii) do any of the parametric estimators still converge? If so, do they converge to the same pseudo-true value that occurs in the stationary case or to a different value?, (iii) can the $\sqrt{n}$– rate of convergence shall be achieved under some conditions, and (iv) is the limiting distribution different from normal? This extension is non-trivial and the limiting criterion function,

$$Q(\boldsymbol{\eta}) = \frac{\sigma_0^2}{2\pi} \int_0^{\pi} \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta},\lambda)} d\lambda \,, \tag{6.1}$$

that defines the pseudo-true value, is not well-defined whenever the values of d_0 and d are greater than or equal to 0.5. Hence the remaining technical details need to be derived rigorously. The topic would provide an interesting introduction to other challenging topics for future research along this line.

Chapters 3 and 4 use artificial data to illustrate the in-sample finite sample performance of five parametric estimators (FML, DWH, EWH, TML and CSS) of the pseudo-true value of the fractional differencing parameter. We intend to extend this analysis to forecasting performance, documenting the performance of both point and distributional forecasts under:

[1] In the literature, there has been some theory focussed on estimation of non-stationary fractionally integrated processes under correct specification of the model (see, Tanaka, 1999, Velasco, 1999a,b and Robinson, 2005). For example, Robinson shows that the ordinary least squares estimator converges to a normal distribution, provided that certain strong regularity conditions are satisfied.

different degrees of mis-specification, for different forecast horizons, and using different (mis-specified) estimators. Results of this analysis will then inform an empirical investigation, in which mis-specification is an unavoidable feature. We expect that these explorations will reveal the importance of mis-specification for a practitioner and, ideally, provide useful guidelines to follow.

Although the theory developed in Chapters 3 and 4 is restricted to linear ARFIMA models, the main ideas are likely to carry over to some time-varying and/or nonlinear time series models as well. For example, we consider extending the issues related to mis-specification in generalized autoregressive conditional heteroskedastic (GARCH) models of Bollerslev (1986). The extension to such models is technically challenging as the process $\{\varepsilon_t^2\}$ is autocorrelated.

In Chapter 5, the proposed optimal bias correction is used to reduce the dominant bias term exhibited by the LPR estimator without inflating the variance. Whilst this technique has been shown to be very successful, as mentioned previously, the resultant estimator is infeasible in practice, as the optimal weights depend on the true, and unknown, values of the dynamic parameters. The suggested feasible version of the jackknife estimator does not show significant reduction in bias compared with the feasible estimators of Guggenberger and Sun (2006) and Poskitt *et al.* (2016), and we need to explore more effective ways of estimating the parameters that enter the weight functions. One possible approach is to replace the spectral density with its estimate, as suggested by Moulines and Soulier (1999). We expect that this may provide better finite sample performance, but is something left for future research.

An alternative semi-parametric estimator, the local Whittle [LW] estimator of Robinson (1995a) also encounters the same challenge as that of the LPR estimator in terms of exhibiting large finite sample bias. Therefore the insights developed in this chapter, including the detailed technical results, will certainly be useful for extending the jackknife-based bias reduction methodology to the LW estimator.

Bibliography

Adenstedt, R. K. (1974). On large-sample estimation for the mean of a stationary random sequence. *The Annals of Statistics*, **2**, 1095–1107.

Agiakloglou, C., Newbold, P., and Wohar, M. (1993). Bias in an estimator of the fractional difference parameter. *Journal of Time Series Analysis*, **14**, 235–246.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, **71**, 579–625.

Andrews, D. W. K. and Guggenberger, P. (2003). A bias-reduced log-periodogram regression estimator for the long-memory parameter. *Econometrica*, **71**, 675–712.

Andrews, D. W. K. and Sun, Y. (2004). Adaptive local polynomial Whittle estimation of long-range dependence. *Econometrica*, **72**, 569–614.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, **73**, 5–59.

Beran, J. (1991). M estimators of location for Gaussian and related processes with slowly decaying serial correlations. *Journal of the American Statistical Association*, **86**, 704–708.

Beran, J. (1994). *Statistics for long-memory processes*. Chapman and Hall, New York, 1st edition.

Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *Journal of the Royal Statistical Society*, **57**, 659–672.

Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2013). *Long-memory processes: Probabilistic properties and statistical methods*. Springer: New York.

Berk, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *The Annals of Mathematical Statistics*, **37**, 51–58.

Berk, R. H. (1970). Consistency a posteriori. *The Annals of Mathematical Statistics*, **41**, 894–904.

Bhansali, R. J., Giraitis, L., and Kokoszka, P. S. (2006). Estimation of the memory parameter by fitting fractionally differenced autoregressive models. *Journal of Multivariate Analysis*, **97**, 2101–2130.

Boes, D. C., Davis, R. A., and Gupta, S. N. (1989). Parameter estimation in low order fractionally differenced ARMA processes. *Stochastic Hydrology and Hydraulics*, **3**, 97–110.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**, 307 – 327.

Bollerslev, T. and Mikkelsen, H. O. (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, **73**, 151–184.

Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco.

Brillinger, D. (1981). *Time series: Data Analysis and Theory*. Holden Day, San Francisco, 2nd edition.

Brockwell, P. J. and Davis, R. A. (1991). *Time series: Theory and Methods*. Springer, New York, 2nd edition.

Cavaliere, G., Nielsen, M. O., and Taylor, M. R. (2017). Quasi-maximum likelihood estimation and bootstrap inference in fractional time series models with heteroskedasticity of unknown form. *Journal of Econometrics*, **198**, 165–188.

Chambers, M. J. (2013). Jackknife estimation of stationary autoregressive models. *Journal of Econometrics*, **172**, 142–157.

Chen, G., Abraham, B., and Peiris, S. (1994). Lag window estimation of the degree of differencing in fractionally integrated time series models. *Journal of Time Series Analysis*, **15**, 473–487.

Chen, W. W. and Deo, R. S. (2006). Estimation of mis-specified long memory models. *Journal of Econometrics*, **134**, 257–281.

Chen, Y. and Yu, J. (2015). Optimal jackknife for unit root models. *Statistics and Probability Letters*, **99**, 135–142.

Cheung, Y. W. (2016). Long memory in foreign-exchange rates. *Journal of Business & Economic Statistics*, **11**, 93–101.

Cheung, Y. W. and Diebold, F. X. (1994). On maximum likelihood estimation of the differencing parameter of fractionally-integrated noise with unknown mean. *Journal of Econometrics*, **62**, 301–316.

Chiang, A. C. and Wainwright, K. (2005). *Fundamental methods of mathematical economics*. McGraw-Hill, New York.

Chung, C. F. and Baillie, R. T. (1993). Small sample bias in conditional sum-of-squares estimators of fractionally integrated ARMA model. *Empirical Economics*, **18**, 791–806.

Contreras-reyes, J. E. and Palma, W. (2013). Statistical analysis of autoregressive fractionally integrated moving average models in R. *Computational Statistics*, **28**, 2309–2331.

Cox, D. R. (1984). Long-range dependence: A review. In H. A. David and H. T. David, editors, *Statistics: An Appraisal*, pages 55–74. The Iowa State University Press, Ames, Iowa.

Crato, N. and Ray, B. K. (1996). Model selection and forecasting for long-range dependent processes. *Journal of Forecasting*, **15**, 107–125.

Crato, N. and Taylor, H. M. (1996). Stationary persistent time series misspecified as nonstationary ARIMA. *Statistical Papers*, **37**, 215–266.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, **17**, 1749–1766.

Delgado, M. A. and Robinson, P. M. (1994). New methods for the analysis of long memory time series-Application to Spanish inflation. *Journal of Forecasting*, **13**, 97–107.

Diebold, F. X. and Rudebusch, G. D. (1989). Long memory and persistence in aggregate output. *Journal of Monetary Economics*, **24**, 189–209.

Diebold, F. X., Husted, S., and Rush, M. (1991). Real exchange rates under the gold standard. *Journal of Political Economy*, **99**, 1252–1271.

Doornik, J. A. and Ooms, M. (2003). Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. *Computational Statistics and Data Analysis*, **42**, 333–348.

Doukhan, P., Oppenheim, G., and Taqqu, S. M. (2003). *Theory and applications of long-range dependence*. Springer Science & Business Media.

Ekonomi, L. and Butka, A. (2011). Jackknife and bootstrap with cycling blocks for the estimation of fractional parameter in ARFIMA model. *Turkish Journal of Mathematics*, **35**, 151–158.

Fox, R. and Taqqu, S. M. (1986). Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *The Annals of Statistics*, **14**, 517–532.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, **4**, 221–238.

Gil-Alana, L. A. (2012). U.K. rainfall data: A long-term persistence approach. *Journal of Applied Meteorology and Climatology*, **51**, 1904–1913.

Gil-Alana, L. A. and Robinson, P. M. (1997). Testing of unit root and other nonstationary hypotheses in macroeconomic time series. *Journal of Econometrics*, **80**, 241–268.

Giraitis, L. and Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probability Theory and Related Fields*, **86**, 87–104.

Giraitis, L., Koul, H. L., and Surgailis, D. (2011). *Large sample inference for long memory processes*. World Scientific Publishing Company.

Gradshtein, I. S. and Ryzhik, I. M. (2007). *Tables of integrals, series and products*. Academic Press, Sydney.

Granger, C. W. J. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P500 absolute stock returns. *Journal of Empirical Finance*, **11**, 399–421.

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, **1**, 15–29.

Grenander, U. (1954). On the estimation of regression coefficients in the case of an autocorrelated disturbance. *Annals of Mathematical Statistics*, **25**, 252–272.

Grenander, U. and Szego, G. (1958). *Toeplitz forms and their application*. University of California Press, Berkeley.

Guggenberger, P. and Sun, Y. (2006). Bias-reduced log-periodogram and Whittle estimation of the long-memory Parameter without variance inflation. *Econometric Theory*, **22**, 863–912.

Hahn, J. and Newey, W. K. (2013). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, **72**, 1295–1319.

Hannan, E. J. (1973). The asymptotic theory of linear time-series models. *Journal of Applied Probability*, **10**, 130–145.

Hassler, U. (1993). Regression of spectral estimators with fractionally integrated time series. *Journal of Time Series Analysis*, **14**, 369–380.

Hassler, U. (1994). (Mis)specification of long memory in seasonal time series. *Journal of Time Series Analysis*, **14**, 19–30.

Hassler, U. (2018). *Time series analysis with long memory in view*. Wiley.

Hassler, U. and Wolters, J. (1995). Long memory in inflation rates: International evidence. *Journal of Business & Economic Statistics*, **13**, 37–45.

Heyde, C. C. and Yang, Y. (1997). Defining and measuring long-range dependence. *Journal of Applied Probability*, **34**, 939–944.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, **68**, 165–175.

Hosking, J. R. M. (1984). Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. Technical report, 2752, Mathematics Research Center, University of Wisconsin, Madison, WI.

Hosking, J. R. M. (1996). Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. *Journal of Econometrics*, **73**, 261–284.

Hualde, J. and Robinson, P. M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models. *The Annals of Statistics*, **39**, 3152–3181.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley symposium*, pages 221–233.

Hurst, H. E. (1951). Methods of using long-term storage in reservoirs. *Transactions of American Civil Engineers*, **33**, 619–626.

Hurvich, C. M. and Brodsky, J. (2001). Broadband semiparametric estimation of the memory parameter of a long-memory time series using fractional exponential models. *Journal of Time Series Analysis*, **22**, 221–249.

Hurvich, C. M., Deo, R. S., and Brodsky, J. (1998). The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series. *Journal of Time Series Analysis*, **19**, 19–46.

Ibragimov, I. A. and Linnik, I. V. (1971). *Independent and stationary sequences of random variables*. Wolters-Noordhoff, Groningen.

Karagiannis, T., Molle, M., and Faloutsos, M. (2004). Long-range dependence-Ten years of internet traffic modeling. *IEEE Internet Computing*, **8**, 1–9.

Kim, C. and Phillips, P. (2006). Log periodogram regression: the nonstationary case. *Cowles Foundation Discussion Papers*.

Krishnaiah, P. R., Haigs, P. J., and Steinberg, L. (1963). A note on the bivariate chi distribution. *SIAM Review*, **5**, 140–144.

Künsch, H. R. (1987). Statistical aspects of self-similar processes. In *Proceedings of the First World Congress of the Bernoulli Society*, pages 67–74, Utrecht. VNU Science Press.

Lahiri, S. N. (2003). A necessary and sufficient condition for asymptotic independence of discrete Fourier transforms under short- and long-range dependence. *Annals of Statistics*, **31**, 613–641.

Lawrance, A. J. and Kottegoda, N. T. (1977). Stochastic modelling of riverflow time series. *Journal of the Royal Statistical Society. Series A (General)*, **140**, 1–47.

Leland, W. E., Taqqu, M. S., Willinger, W., and Wilson, D. V. (1993). On the self-similar nature of ethernet traffic (extended version). *IEEE/ACM Transactions on Networking (ToN)*, **2**, 1–15.

Lieberman, O. (2001). The exact bias of the log-periodogram regression estimator. *Econometric Reviews*, **20**, 369–383.

Lieberman, O., Rosemarin, R., and Rousseau, J. (2010). Asymptotic theory for maximum likelihood estimation of the memory parameter in stationary Gaussian processes. *Working paper*.

Lieberman, O., Rosemarin, R., and Rousseau, J. (2012). Asymptotic theory for maximum likelihood estimation of the memory parameter in stationary Gaussian processes. *Econometric Theory*, **28**, 457–470.

Lundahl, T., Ohley, W. J., Kay, S. M., and Siffert, R. (1986). Fractional brownian motion: A maximum likelihood estimator and its application to image texture. *IEEE Transactions on Medical Imaging*, **5**, 152–161.

Martin, G. M., Nadarajah, K., and Poskitt, D. S. (2018). Issues in the estimation of mis-specified models of fractionally integrated processes. https://www.monash.edu/business/ebs/research/publications/ebs/wp18-2018.pdf.

Moulines, E. and Soulier, P. (1999). Broadband log-periodogram regression of time series with long-range dependence. *Annals of Statistics*, **27**, 1415–1439.

Nielsen, M. O. and Frederiksen, P. H. (2005). Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. *Econometric Reviews*, **24**, 405–443.

Ooms, M. and Franses, P. H. (2001). A seasonal periodic long memory model for monthly river flows. *Environmental Modelling and Software*, **16**, 559–569.

Palma, W. (2010). On the sample mean of locally stationary long-memory processes. *Journal of Statistical Planning and Inference*, **140**, 3764–3774.

Phillips, P. C. B. and Shimotsu, K. (2004). Local whittle estimation in nonstationary and unit root cases. *Annals of Statistics*, **32**, 656–692.

Pipiras, V. and Taqqu, M. S. (2017). *Long-range dependence and self-similarity*. Cambridge university press.

Poskitt, D. S. (2008). Properties of the sieve bootstrap for fractionally integrated and non-invertible processes. *Journal of Time Series Analysis*, **29**, 224–250.

Poskitt, D. S., Martin, G. M., and Grose, S. D. (2016). Bias correction of semiparametric long memory parameter estimators via the prefiltered sieve bootstrap. *Econometric Theory*, **1**, 1–32.

Reisen, V. A. (1994). Estimation of the fractional difference parameter in the ARIMA(p,d,q) model using the smoothed periodogram. *Journal of Time Series Analysis*, **15**, 335–350.

Reisen, V. A. and Lopes, S. (1999). Some simulations and applications of forecasting long-memory time-series models. *Journal of Statistical Planning and Inference*, **80**, 269–287.

Robinson, P. M. (1995a). Gaussian semiparametric estimation of the long range dependence. *The Annals of Statistics*, **23**, 1630–1661.

Robinson, P. M. (1995b). Log-periodogram regression of time series with long range dependence. *Annals of Statistics*, **23**, 1048–1072.

Robinson, P. M. (2004). *Time series with long memory*. Oxford University Press, New York.

Robinson, P. M. (2005). Efficiency improvements in inference on stationary and nonstationary fractional time series. *The Annals of Statistics*, **33**, 1800–1842.

Robinson, P. M. (2006). Conditional-sum-of-squares estimation of models for stationary time series with long memory. *IMS Lecture Notes-Monograph Series Time Series and Related Topics*, **52**, 130–137.

Robinson, P. M. (2014). The estimation of misspecified long memory models. *Journal of Econometrics*, **178**, 225–230.

Robinson, P. M. and Henry, M. (2003). Higher-order kernel semiparametric M-estimation of long memory. *Journal of Econometrics*, **114**, 1–27.

Robinson, Y. K. and Kaufmann, H. (2015). Bias-corrected estimation in mildly explosive autoregressions. In *In Annual Conference 2015 (Muenster): Economic Development-Theory and Policy (No. 112897)*. Verein für Socialpolitik/German Economic Association.

Samarov, A. and Taqqu, S. M. (1988). On the efficiency of the sample mean in long-memory noise. *Journal of Time Series Analysis*, **9**, 191–210.

Scherrer, A., Larrieu, N., Owezarski, P., Borgnat, P., and Abry, P. (2007). Non-gaussian and long memory statistical characterizations for internet traffic with anomalies. *IEEE Transactions on Dependable and Secure Computing*, **4**, 56–70.

Shimotsu, K. and Phillips, P. C. B. (2002). Pooled log periodogram regression. *Journal of Time Series Analysis*, **23**, 57–93.

Shimotsu, K. and Phillips, P. C. B. (2005). Exact local Whittle estimation of fractional integration. *The Annals of Statistics*, **33**, 1890–1933.

Škare, M. and Stjepanović, S. (2013). A fractionally integrated model for the Croatian aggregate output (GDP) series. *Economic Research-Ekonomska Istraživanja*, **26**, 1–34.

Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, **53**, 165–188.

Tanaka, K. (1999). The nonstationary fractional unit root. *Econometric Theory*, **15**, 549–582.

Varneskov, R. T. and Perron, P. (2018). Combining long memory and level shifts in modelling and forecasting the volatility of asset returns. *Quantitative Finance*, **18**, 371–393.

Velasco, C. (1999a). Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis*, **20**, 87–127.

Velasco, C. (1999b). Non-stationary log-periodogram regression. *Journal of Econometrics*, **91**, 325–371.

Velasco, C. (2000). Non-Gaussian log-periodogram regression. *Econometric Theory*, **16**, 44–79.

Velasco, C. and Robinson, P. M. (2000). Whittle pseudo-maximum likelihood estimation for nonstationary time series. *Journal of the American Statistical Association*, **95**, 1229–1243.

Wagenmakers, E. J., Farrell, S., and Ratcliff, R. (2004). Estimation and interpretation of $1/f\alpha$ noise in human cognition. *Psychonomic Bulletin and Review*, **11**, 579–615.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–25.

Whittle, P. (1952). Estimation and information in stationary time series. *ARKIV FOR MATEMATIK*, **2**, 423–434.

Whittle, P. (1957). Curve and periodogram smoothing. *Journal of the Royal Statistical Society*, **19**, 38–63.

Whittle, P. (1962). Gaussian estimation in stationary time series. *Bulletin of the International Statistical Institute*, **39**, 105–129.

Whittle, P. (1984). *Prediction and regulation by linear least squares methods*. Basil Blackwell.

Yajima, Y. (1989). A central limit theorem of Fourier transforms of strongly dependent stationary processes. *Journal of Time Series Analysis*, **10**, 375–383.

Yajima, Y. (1993). Asymptotic properties of estimates in incorrect ARMA models for long memory time series. In D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt, and M. Taqqu, editors, *In New directions in time series analysis, Part II.*, pages 375–382. Springer, New York.