ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Issues in the Estimation of Mis-Specified Models of Fractionally Integrated Processes

K. Nadarajah, Gael M. Martin and D.S. Poskitt

June 2014

Working Paper 18/14

Issues in the Estimation of Mis-Specified Models of Fractionally Integrated Processes[†]

K. Nadarajah, Gael M. Martin[‡]& D. S. Poskitt
Department of Econometrics & Business Statistics, Monash University

June 19, 2014

Abstract

In this paper we quantify the impact of model mis-specification on the properties of parameter estimators applied to fractionally integrated processes. We demonstrate the asymptotic equivalence of four alternative parametric methods: frequency domain maximum likelihood, Whittle estimation, time domain maximum likelihood and conditional sum of squares. We show that all four estimators converge to the same pseudo-true value and provide an analytical representation of their (common) asymptotic distribution. As well as providing theoretical insights, we explore the finite sample properties of the alternative estimators when used to fit mis-specified models. In particular we demonstrate that when the difference between the true and pseudo-true values of the long memory parameter is sufficiently large, a clear distinction between the frequency domain and time domain estimators can be observed – in terms of the accuracy with which the finite sample distributions replicate the common asymptotic distribution – with the time domain estimators exhibiting a closer match overall. Simulation experiments also demonstrate that the two time-domain estimators have the smallest bias and mean squared error as estimators of the pseudo-true value of the long memory parameter, with conditional sum of squares being the most accurate estimator overall and having a relative efficiency that is approximately double that of frequency domain maximum likelihood, across a range of mis-specification designs.

Keywords and phrases: bias, conditional sum of squares, frequency domain, long memory models, maximum likelihood, mean squared error, pseudo true parameter, time domain, Whittle.

MSC2010 subject classifications: Primary 62M10, 62M15; Secondary 62G09
JEL classifications: C18, C22, C52

1 Introduction

This paper examines the properties of four alternative parametric techniques – frequency domain maximum likelihood (FML), Whittle, time domain maximum likelihood (TML) and conditional sum of squares (CSS) – when they are employed to estimate a mis-specified model applied to a true data generating process (TDGP) that exhibits long range dependence. These estimators have a long history in time series analysis, dating back to the pioneering work of Grenander and Rosenblatt (1957), Whittle (1962), Walker (1964), Box and Jenkins (1970) and Hannan (1973), and their properties in the context of weakly dependent processes are well known (see, for instance, Brockwell and Davis, 1991). Extension of these methods

[†]This research has been supported by Australian Research Council (ARC) Discovery Grant DP120102344 and ARC Future Fellowship FT0991045.

[‡]Corresponding author: Gael Martin, Department of Econometrics and Business Statistics, Monash University, Clayton, Victoria 3800, Australia. Tel.: +61-3-9905-1189; fax: +61-3-9905-5474; email: gael.martin@monash.edu.

to the analysis of strongly dependent processes has been examined in Fox and Taqqu (1986), Dahlhaus (1989), Sowell (1992), Beran (1995) and Robinson (2006), among others, but this literature presupposes that the structure of the TDGP is known apart from the values of a finite number of parameters that are to be estimated. Recognition that the true structure can only ever be approximated by the model being fitted has given rise to two responses: (i) the development of semi-parametric techniques such as those advanced by Geweke and Porter-Hudak (1983) and Robinson (1995a,b) for example; and (ii) the examination of the consequences of mis-specification.

Significant contributions to the issue of mis-specification in long memory models have been made by Yajima (1992) and Chen and Deo (2006). Specifically, Yajima investigates the asymptotic properties of the estimators of the parameters in an autoregressive moving average (ARMA) model under a long memory fractional noise TDGP; whilst Chen and Deo focus on the estimation of the parameters in an incorrectly specified fractionally integrated model. Both studies demonstrate that once model mis-specification is accommodated consistency for the true parameters no longer obtains, and that the properties of inferential methods become case-specific and dependent on the precise nature and degree of mis-specification. In particular, it is shown that the estimator of the (vector-valued) parameter of a mis-specified model converges, subject to regularity, to a 'pseudo-true' value that is different from the true value and that the estimator may or may not achieve the usual $\sqrt{n}$ rate of convergence and limiting Gaussianity, depending on the magnitude of the deviation between the true and pseudo-true parameters.

By definition, the pseudo-true parameter is the value which optimizes the limiting form of the objective function that defines an estimator. Chen and Deo (2006) derive the form of this limiting objective function for the FML estimator, and proceed to demonstrate that the asymptotic behaviour of the parametric estimator of the incorrectly specified model is dependent on whether the distance between the true and pseudo-true values of the long memory parameter, d, is less than, equal to, or in excess of 0.25. For specific models in the autoregressive fractionally integrated moving average (ARFIMA) class, this distance is then linked to respective values of the ARMA parameter(s) in the true and mis-specified models. The extent to which mis-specification of the short memory dynamics is still compatible with $\sqrt{n}$-consistency and asymptotic Gaussianity is then documented for these particular examples.

In this paper we extend the analysis of Chen and Deo (2006) in several directions. Firstly, we derive the limiting form of the objective function for the three other commonly-used parametric estimators – namely, Whittle, TML and CSS – and show that the FML, Whittle, TML and CSS estimators will converge to the same pseudo-true parameter value under common mis-specification. Secondly, we derive closed-form representations for the first-order conditions that define the pseudo-true parameter for *general* ARFIMA model structures. Thirdly, we extend the asymptotic theory established by Chen and Deo for the FML estimator to the other three estimators, and show that all four methods are asymptotically equivalent, in that they converge in distribution under common mis-specification. Fourthly, we demonstrate how to implement numerically the asymptotic distribution that obtains under the most extreme type of mis-specification, by using an appropriate method of truncating the series expansion in random variables that characterises 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, when the difference between the true and pseudo-true values of d is greater than or equal to 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 simulation experiment in which the relative finite sample performance of all four mis-specified estimators is assessed, with the CSS estimator exhibiting superior performance, in terms of bias and mean squared error, across a range of mis-specification settings.

The paper is organized as follows. In Section 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 FML estimator and the three above-mentioned alternative estimators are specified, and we demonstrate that all four estimates converge under common mis-specification. The limiting form of the criterion function for a mis-specified ARFIMA model is presented in Section 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 4. The finite sample performance of the four parametric estimators of d in the mis-specified model – with reference to estimating the pseudo-true value d_1 – is documented in Section 5. The form of the sampling distribution is recorded, as is the bias and mean squared error (MSE), under different degrees of mis-specification. Section 6 then concludes. The proofs of the results presented in the paper are assembled in Appendix A, which also presents a lemma required in the proofs. Appendix B contains certain technical derivations referenced in the text.

2 Estimation Under Misspecification

Assume that $\{y_t\}$ is generated from a TDGP that is a stationary Gaussian process with spectral density given by

$$f_0(\lambda) = \frac{\sigma_{\varepsilon 0}^2}{2\pi} g_0(\lambda)\left(2\sin(\lambda/2)\right)^{-2d_0}, \tag{1}$$

where $g_0(\lambda)$ is a real valued function of λ defined on $[0,\pi]$ that is bounded above and bounded away from zero. The model refers to a parametric specification for the spectral density of $\{y_t\}$ of the form

$$f_1(\boldsymbol{\psi},\lambda) = \frac{\sigma_{\varepsilon}^2}{2\pi} g_1(\boldsymbol{\beta},\lambda)\left(2\sin(\lambda/2)\right)^{-2d}, \tag{2}$$

that is to be estimated from the data, where $g_1(\boldsymbol{\beta},\lambda)$ is a real valued function of λ defined on $[0,\pi]$ that is bounded above and bounded away from zero. Let $\boldsymbol{\Psi} = \mathbb{R}^+ \times (0, 0.5) \times \boldsymbol{\Theta}$ and denote by $\boldsymbol{\psi} = (\sigma_\varepsilon^2, \boldsymbol{\eta}^T)^T \in \boldsymbol{\Psi}$ the parameter vector of the model where $\boldsymbol{\eta} = (d, \boldsymbol{\beta}^T)^T$ and $\boldsymbol{\beta} \in \boldsymbol{\Theta}$, with $\boldsymbol{\Theta} \subset \mathbb{R}^l$ an l-dimensional compact convex set. It will be assumed that:

$(A.1)$ $g_1(\boldsymbol{\beta},\lambda)$ is thrice differentiable with continuous third derivatives.

$(A.2)$ $\inf_{\boldsymbol{\beta}} \inf_{\lambda} g_1(\boldsymbol{\beta},\lambda) > 0$ and $\sup_{\boldsymbol{\beta}} \sup_{\lambda} g_1(\boldsymbol{\beta},\lambda) < \infty$.

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

$(A.4)$ $\sup_{\lambda} \sup_{\boldsymbol{\beta}} \left| \frac{\partial^2 g_1(\boldsymbol{\beta},\lambda)}{\partial \beta_i \partial \beta_j} \right| < \infty$, $\sup_{\lambda} \sup_{\boldsymbol{\beta}} \left| \frac{\partial^2 g_1(\boldsymbol{\beta},\lambda)}{\partial \beta_i \partial \lambda} \right| < \infty$, $1 \leqslant i, j \leqslant l$.

$(A.5)$ $\sup_{\lambda} \sup_{\boldsymbol{\beta}} \left| \frac{\partial^3 g_1(\boldsymbol{\beta},\lambda)}{\partial \beta_i \partial \beta_j \partial \beta_k} \right| < \infty$, $1 \leqslant i, j, k \leqslant l$.

$(A.6)$ $\int_{-\pi}^{\pi} \log g_1(\boldsymbol{\beta},\lambda) d\lambda = 0$ for all $\boldsymbol{\beta} \in \boldsymbol{\Theta}$.

If there exists a subset of $[0,\pi]$ with non-zero Lebesgue measure in which $g_1(\boldsymbol{\beta},\lambda) \neq g_0(\lambda)$ for all $\boldsymbol{\beta} \in \boldsymbol{\Theta}$ then the model will be referred to as a mis-specified model (MM).

An ARFIMA model for a time series $\{y_t\}$ may be defined as follows,

$$\phi(L)(1-L)^d\{y_t-\mu\}=\theta(L)\varepsilon_t, \tag{3}$$

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 $\sigma_\varepsilon^2>0$. 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) is correct, $\{y_t\}$ is defined as the limit in mean square of a covariance-stationary process. When $d\leq 0$ the process is weakly dependent and in this case the behaviour of the estimators is to a large degree already known. We will therefore assume that $0<d<0.5$. When $0<d<0.5$ neither the moving average coefficients nor the autocovariances of the process are absolutely summable, declining at a slow hyperbolic rate rather than the exponential rate typical of an ARMA process, with the term 'long memory' invoked accordingly. A detailed outline of the properties of ARFIMA processes is provided in Beran (1994). For an ARFIMA model we have $g_1(\boldsymbol{\beta},\lambda)=|\theta(e^{i\lambda})|^2/|\phi(e^{i\lambda})|^2$ where $\boldsymbol{\beta}=(\phi_1,\phi_2,...,\phi_p,\theta_1,\theta_2,...,\theta_q)^T$ and Assumptions $A.1-A.6$ are satisfied. 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.

The estimators to be considered (denoted generally by $\widehat{\boldsymbol{\psi}}$) are all to be obtained by minimizing an objective function, $Q_n(\boldsymbol{\psi})$, say, and under mis-specification the estimator $\widehat{\boldsymbol{\psi}}_1$ is obtained by minimizing $Q_n(\boldsymbol{\psi})$ on the assumption that $\{y_t\}$ follows the MM.[1] For any given $Q_n(\boldsymbol{\psi})$, there exists a non-stochastic limiting objective function $Q(\boldsymbol{\psi})$, that is independent of the sample size n, such that $|Q_n(\boldsymbol{\psi})-Q(\boldsymbol{\psi})|\to^p 0$ for all $\boldsymbol{\psi}\in\boldsymbol{\Phi}$, and provided certain conditions hold, $Q_n(\widehat{\boldsymbol{\psi}}_1)$ will converge to $Q(\boldsymbol{\psi}_1)$ where $\boldsymbol{\psi}_1$ is the minimizer of $Q(\boldsymbol{\psi})$ and $\widehat{\boldsymbol{\psi}}_1\to^p \boldsymbol{\psi}_1$ as a consequence. In Subsection 2.1 we specify the form of $Q_n(\boldsymbol{\psi})$ associated with the FML estimator, $\widehat{\boldsymbol{\psi}}_1^{(1)}$ hereafter, and outline the asymptotic results derived in Chen and Deo (2006) pertaining to the convergence of $\widehat{\boldsymbol{\psi}}_1^{(1)}$ to $\boldsymbol{\psi}_1$. In Subsection 2.2 the equivalence of the values that minimize the limiting criterion functions of the three alternative estimators to the value that minimizes the limiting criterion function of the FML estimator is demonstrated and, hence, the asymptotic convergence of these four estimators established.

2.1 Frequency domain maximum likelihood estimation

Chen and Deo (2006) focus on the FML estimator of $\boldsymbol{\eta}=(d,\boldsymbol{\beta}^T)^T$, $\widehat{\boldsymbol{\eta}}_1$, defined as the value of $\boldsymbol{\eta}\in(0,0.5)\times\boldsymbol{\Theta}$ that minimizes the objective function

$$Q_n(\boldsymbol{\eta})=\frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)}, \tag{4}$$

where $I(\lambda_j)$ is the periodogram, defined as $I(\lambda)=\frac{1}{2\pi n}|\sum_{t=1}^n y_t\exp(-i\lambda t)|^2$ 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, and, with a slight abuse of notation, $f_1(\boldsymbol{\eta},\lambda_j)=g_1(\boldsymbol{\beta},\lambda_j)\,(2\sin(\lambda_j/2))^{-2d}$. The objective function in (4) is a frequency domain approximation to the negative of the Gaussian log-likelihood (See Brockwell and Davis, 1991, §10.8, for example.). Indeed, one of the alternative estimators that we consider (TML) is the minimizer of the exact version of this negative log-likelihood function.

[1] We follow the usual convention by denoting the estimator obtained under mis-specification as $\widehat{\boldsymbol{\psi}}_1$ rather than simply by $\widehat{\boldsymbol{\psi}}$, 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 $\widehat{\boldsymbol{\psi}}_0$.

Let

$$Q(\boldsymbol{\eta}) = \lim_{n\to\infty} E_0\left[Q_n(\boldsymbol{\eta})\right] = \int_0^{\pi} \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta},\lambda)} d\lambda\,, \tag{5}$$

where here, and in what follows, the zero subscript denotes that the moments are defined with respect to the TDGP. From Lemma 2 of Chen and Deo (2006) it follows that under Assumptions $A.1 - A.3$,

$$\sup_{\boldsymbol{\eta}\in(0,0.5)\times\boldsymbol{\Theta}} \left| \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)} - Q(\boldsymbol{\eta}) \right| \to^p 0\,. \tag{6}$$

The limiting objective function $Q(\boldsymbol{\eta})$, in turn, defines the pseudo-true parameter $\boldsymbol{\eta}_1$ to which $\widehat{\boldsymbol{\eta}}_1$ will converge under the assumed regularity. This follows from (6) and the additional assumption:

$(A.7)$ There exists a unique vector $\boldsymbol{\eta}_1 = (d_1, \boldsymbol{\beta}_1^T)^T \in (0, 0.5) \times \boldsymbol{\Theta}$, with $\boldsymbol{\beta}_1 = (\beta_{11}, ..., \beta_{l1})^T$, which satisfies $\boldsymbol{\eta}_1 = \arg\min_{\boldsymbol{\eta}} Q(\boldsymbol{\eta})$.

On application of a standard argument for M-estimators, (6) and $(A.7)$ imply that $\text{plim}\, \widehat{\boldsymbol{\eta}}_1 = \boldsymbol{\eta}_1$ (see Chen and Deo, 2006, Corollary 1).

2.2 Alternative Estimators

Index by $i = 1, 2, 3$ and 4 respectively, the criterion function associated with the FML estimator, the Whittle estimator, the TML estimator and the CSS estimator, each viewed as a function of $\boldsymbol{\psi}$ or $\boldsymbol{\eta}$, that is $Q_n^{(i)}(\cdot)$, $i = 1, 2, 3, 4$. The criterion function of the FML estimator is given in (4). The criterion functions of the three alternative estimators are defined as follows:

- The objective function for the Whittle estimator as considered in Beran (1994) is

$$Q_n^{(2)}(\boldsymbol{\psi}) = \frac{4}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \log f_1(\boldsymbol{\psi},\lambda_j) + \frac{4}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{f_1(\boldsymbol{\psi},\lambda_j)}\,, \tag{7}$$

  where $\boldsymbol{\psi} = (\sigma_\varepsilon^2, \boldsymbol{\eta}^T)^T$, which when re-expressed as an explicit function of σ_ε^2 and $\boldsymbol{\eta}$ gives

$$Q_n^{(2)}(\sigma_\varepsilon^2, \boldsymbol{\eta}) = \frac{4}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \log\left[\frac{\sigma_\varepsilon^2}{2\pi} f_1(\boldsymbol{\eta},\lambda_j)\right] + \frac{8\pi}{\sigma_\varepsilon^2 n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)}.$$

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

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

  The Gaussian log-likelihood function for the TML estimator is

$$-\frac{1}{2}\left(n \log(2\pi\sigma_\varepsilon^2) + \log|\boldsymbol{\Sigma}_\eta| + \frac{1}{\sigma_\varepsilon^2}\mathbf{Y}^T \boldsymbol{\Sigma}_\eta^{-1}\mathbf{Y}\right)\,, \tag{8}$$

  and maximizing (8) with respect to $\boldsymbol{\psi}$ is equivalent to minimizing the criterion function

$$Q_n^{(3)}(\sigma_\varepsilon^2, \boldsymbol{\eta}) = \log\sigma_\varepsilon^2 + \frac{1}{n}\log|\boldsymbol{\Sigma}_\eta| + \frac{1}{n\sigma_\varepsilon^2}\mathbf{Y}^T \boldsymbol{\Sigma}_\eta^{-1}\mathbf{Y}\,. \tag{9}$$

- To construct the CSS estimator note that we can expand $(1-z)^d$ in a binomial expansion as
$$(1-z)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} z^j \,, \tag{10}$$
where $\Gamma(\cdot)$ is the gamma function. Furthermore, since $g_1(\boldsymbol{\beta}, \lambda)$ is bounded, by Assumption $(A.2)$, we can employ the method of Whittle (Whittle, 1984, §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
$$Q_n^{(4)}(\boldsymbol{\eta}) = \frac{1}{n} \sum_{t=1}^{n} e_t^2 \,, \tag{11}$$
where
$$e_t = \sum_{i=0}^{t-1} \tau_i(\boldsymbol{\eta}) y_{t-i} \tag{12}$$
and the coefficients $\tau_j(\boldsymbol{\eta})$, $j = 0, 1, 2, \ldots$, are given by $\tau_0(\boldsymbol{\eta}) = 1$ and
$$\tau_j(\boldsymbol{\eta}) = \sum_{s=0}^{j} \frac{\alpha_{j-s}(\boldsymbol{\beta})\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} \,, \quad j = 1, 2, \ldots \,. \tag{13}$$

In Appendix A we prove that for $i = 1, 2, 3$ 4, we have $Q_n^{(i)}(\cdot) \to^p \mathcal{Q}^{(i)}(\sigma_\varepsilon^2, Q(\boldsymbol{\eta}))$, where the minimum of the function $\mathcal{Q}^{(i)}(\sigma_\varepsilon^2, Q(\boldsymbol{\eta}))$ occurs at $\sigma_\varepsilon^2 = 2Q(\boldsymbol{\eta}_1)$ for all i, and each $\mathcal{Q}^{(i)}$, when concentrated with respect to σ_ε^2, is a monotonically increasing function of $Q(\boldsymbol{\eta})$, with $Q(\boldsymbol{\eta})$ as defined in (5). Hence, with $\boldsymbol{\eta}$ being the (vector-valued) parameter of interest, we can state the following proposition:

Proposition 1 *Suppose that the TDGP of* $\{y_t\}$ *is a Gaussian process with a spectral density as given in* (1) *and that the MM satisfies Assumptions* $A.1 - A.7$. *Let* $\widehat{\boldsymbol{\eta}}_1^{(i)}$, $i = 1, 2, 3, 4$, *denote, respectively, the FML, Whittle, TML and CSS estimators of the parameter vector* $\boldsymbol{\eta} = (d, \boldsymbol{\beta}^T)^T$ *of the MM. Then* $\|\widehat{\boldsymbol{\eta}}_1^{(i)} - \widehat{\boldsymbol{\eta}}_1^{(j)}\| \to_P 0$ *for all* $i, j = 1, 2, 3, 4$ *and the common probability limit of* $\widehat{\boldsymbol{\eta}}_1^{(i)}$, $i = 1, 2, 3, 4$, *is* $\boldsymbol{\eta}_1 = \arg\min_{\boldsymbol{\eta}} Q(\boldsymbol{\eta})$.

Note that if the MM were used to construct a one-step-ahead prediction, the mean squared prediction error would be
$$\sigma_\varepsilon^2 = 2Q(\boldsymbol{\eta}) = \int_{-\pi}^{\pi} \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta}, \lambda)} d\lambda \geq \sigma_{\varepsilon 0}^2 \,,$$
where $\sigma_{\varepsilon 0}^2$ is the mean squared prediction error of the minimum mean squared error predictor of the TDGP, (Brockwell and Davis, 1991, Proposition 10.8.1). The implication of Assumption $A.7$ is that among all spectral densities within the mis-specified family the member characterised by the parameter value $\boldsymbol{\eta}_1$ is closest to the true spectral density $f_0(\lambda)$. Evidently it is $\boldsymbol{\eta}_1$ that the estimators should be trying to target as this will give fitted parameter values that yield the predictor from within the MM class whose mean squared prediction error is closest to that of the optimal predictor. Having established that the four parametric estimators converge towards $\boldsymbol{\eta}_1$ under mis-specification, 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 4 by establishing that all four alternative parametric estimators converge in distribution. Prior to doing this, however, we indicate the precise form of the limiting objective function $Q(\boldsymbol{\eta})$, and the associated first-order conditions that define the (common) pseudo-true value $\boldsymbol{\eta}_1$ of the four estimation procedures, in the ARFIMA case. As well as being relevant for all four estimation methods, these derivations apply in complete generality with respect to the models that specify both the TDGP and the MM. Hence, in this sense also the results represent a substantive extension of the corresponding results in Chen and Deo (2006).

3 Pseudo-True Parameters Under ARFIMA Mis-Specification

Under Assumptions $A.1 - A.7$ $\boldsymbol{\eta}_1 = \arg\min_{\boldsymbol{\eta}} Q(\boldsymbol{\eta})$ 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 MM is an ARFIMA $(0, d, 0)$. They then cite (without providing detailed derivations) certain results that obtain when the MM 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

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

and that of the MM, an *ARFIMA* (p, d, q) model, by

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

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

$$Q(\boldsymbol{\psi}) = \int_0^\pi \frac{f_0(\lambda)}{f_1(\boldsymbol{\psi}, \lambda)} d\lambda = \frac{\sigma_{\varepsilon 0}^2}{\sigma_{\varepsilon}^2} \int_0^\pi \frac{|A_\beta(e^{i\lambda})|^2}{|B_\beta(e^{i\lambda})|^2} |2\sin(\lambda/2)|^{-2(d_0 - d)} d\lambda, \tag{14}$$

where

$$A_\beta(z) = \sum_{j=0}^{\underline{q}} a_j z^j = \theta_0(z)\phi(z) = \left(1 + \theta_{10} z + \dots + \theta_{q_0 0} z^{q_0}\right)\left(1 + \phi_1 z + \dots + \phi_p z^p\right) \tag{15}$$

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

$$B_\beta(z) = \sum_{j=0}^{\underline{p}} b_j z^j = \phi_0(z)\theta(z) = \left(1 + \phi_{10} z + \dots + \phi_{p_0 0} z^{p_0}\right)\left(1 + \theta_1 z + \dots + \theta_q z^q\right) \tag{16}$$

with $\underline{p} = p_0 + q$. The expression for $Q(\boldsymbol{\psi})$ in (14) 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 $Q(\boldsymbol{\psi})$ 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 (15) and (16) respectively. Then (14) can be expanded to give

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

Using standard results for the integral $\int_0^{\pi} (\sin x)^{v-1} \cos(ax) dx$ from Gradshteyn and Ryzhik (2007, p 397) yields, after some algebraic manipulation,

$$Q(\boldsymbol{\psi}) = \frac{\pi}{(1-2(d_0-d))} \frac{\sigma_{\varepsilon 0}^2}{\sigma_\varepsilon^2} \left[\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{c_j c_k \cos\left((j-k)\pi\right)}{\mathcal{B}\left(1-(d_0-d)+(j-k), 1-(d_0-d)-(j-k)\right)} \right],$$

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

$$Q(\boldsymbol{\psi}) = \{\pi \frac{\sigma_{\varepsilon 0}^2}{\sigma_\varepsilon^2} \frac{\Gamma(1-2(d_0-d))}{\Gamma^2(1-(d_0-d))}\} K(\boldsymbol{\eta}), \tag{17}$$

where

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

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

$$\frac{\partial Q(\boldsymbol{\psi})}{\partial \beta_r} = \{\pi \frac{\sigma_{\varepsilon 0}^2}{\sigma_\varepsilon^2} \frac{\Gamma(1-2(d_0-d))}{\Gamma^2(1-(d_0-d))}\} \frac{\partial K(\boldsymbol{\eta})}{\partial \beta_r}, \quad r = 1, 2, ..., l,$$

where

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

and

$$\frac{\partial Q(\boldsymbol{\psi})}{\partial d} = \{\pi \frac{\sigma_{\varepsilon 0}^2}{\sigma_\varepsilon^2} \frac{\Gamma(1-2(d_0-d))}{\Gamma^2(1-(d_0-d))}\} \left\{ 2\left(\Psi[1-2(d_0-d)] - \Psi[1-(d_0-d)]\right) K(\boldsymbol{\eta}) + \frac{\partial K(\boldsymbol{\eta})}{\partial d} \right\},$$

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

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

Eliminating the common (non-zero) factor $\{\pi \frac{\sigma_{\varepsilon 0}^2}{\sigma_\varepsilon^2} \frac{\Gamma(1-2(d_0-d))}{\Gamma^2(1-(d_0-d))}\}$ from both $\partial Q(\boldsymbol{\psi})/\partial \boldsymbol{\beta}$ and $\partial Q(\boldsymbol{\psi})/\partial d$, it follows that the pseudo-true parameter values of the *ARFIMA* (p, d, q) MM can be obtained by solving

$$\frac{\partial K(\boldsymbol{\eta})}{\partial \beta_r} = 0, \quad r = 1, 2, ..., l, \tag{18}$$

and

$$2(\Psi[1-2(d_0-d)]-\Psi[1-(d_0-d)])K(\boldsymbol{\eta})+\frac{\partial K(\boldsymbol{\eta})}{\partial d}=0 \tag{19}$$

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 (18) and (19) and solving (numerically) for ϕ_{i1},, $i=1,2,...,p$, θ_{j1}, $j=1,2,...,q$, and d_1.

Theorem 2 *Set* $C_N(z)=\sum_{j=0}^{N}c_jz^j$ *and let* $Q_N(\boldsymbol{\psi})=\left(\sigma_{\varepsilon0}^2/\sigma_\varepsilon^2\right)I_N$ *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*

$$Q(\boldsymbol{\psi})=Q_N(\boldsymbol{\psi})+R_N=\left\{\pi\frac{\sigma_{\varepsilon0}^2}{\sigma_\varepsilon^2}\frac{\Gamma(1-2(d_0-d))}{\Gamma^2(1-(d_0-d))}\right\}K_N(\boldsymbol{\eta})+R_N$$

where

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

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

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, $C(z)=1+\sum_{j=1}^{\underline{q}}c_jz^j$ 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:

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

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

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

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

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

for the required derivatives. The pseudo-true values ϕ_{r1}, $r=1,\ldots,p$, and d_1 can now be obtained by solving (18) and (19) 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.

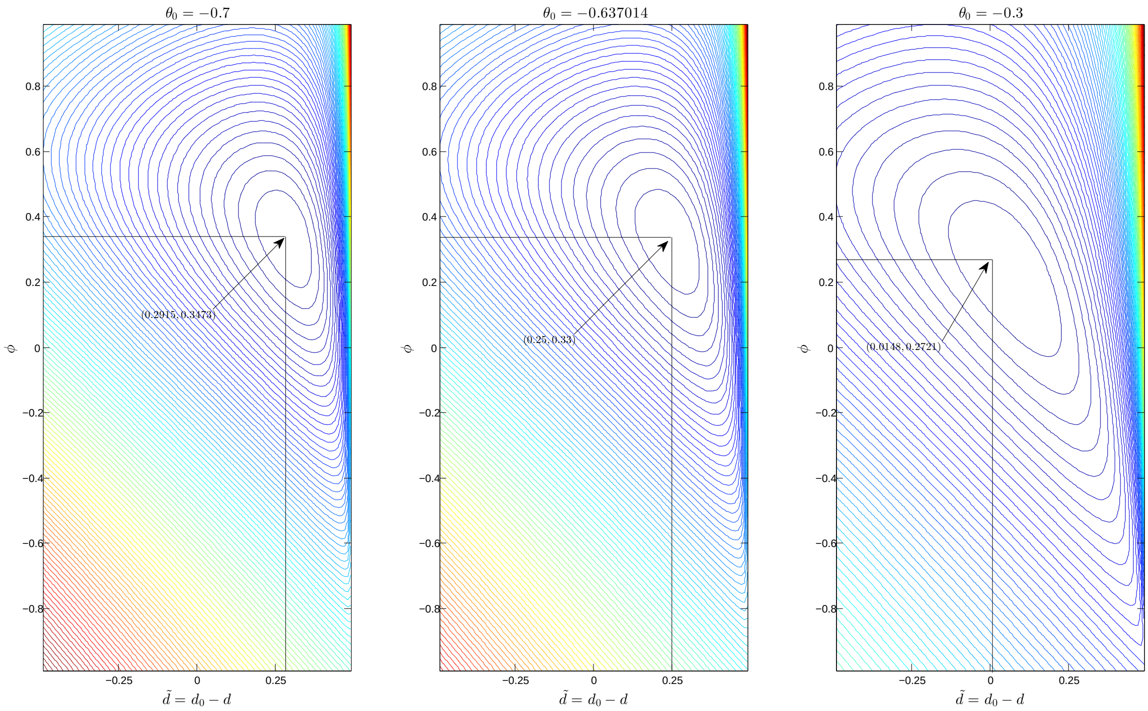
Let us further highlight some features of this special case by focussing on the case where the TDGP is an ARFIMA$(0,d_0,1)$ and the MM an ARFIMA$(1,d,0)$. In this example $\underline{q}=2$

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 (17) is now

$$K(d,\phi) = 1 + (\theta_0 + \phi)^2 + (\theta_0\phi)^2 + \frac{2\left[\theta_0\phi(d_0 - d + 1) - (1 + \theta_0\phi)(\theta_0 + \phi)(d_0 - d - 2)\right](d_0 - d)}{(d_0 - d - 1)(d_0 - d - 2)} . \quad (20)$$

The derivatives $\partial K(d,\phi)/\partial\phi$ and $\partial K(d,\phi)/\partial d$ can be readily determined from (20) and hence the pseudo-true values d_1 and ϕ_1 evaluated.

It is clear from (20) 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 $Q\,(d,\phi) = (\sigma_\varepsilon^2/\sigma_{\varepsilon 0}^2) Q\,(\boldsymbol{\psi})$ similarly. Figure 1 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 $\sigma_\varepsilon^2 = \sigma_{\varepsilon 0}^2$. 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 1 show that



(a) $\theta_0 = -0.7$. (b) $\theta_0 = -0.637014$. (c) $\theta_0 = -0.3$.

Figure 1: 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$(1, d, 0)$ MM; $\widetilde{d} \in (-0.5, 0.5)$, $\phi \in (-1, 1)$. Pseudo-true coordinates $(d_0 - d_1, \phi_1)$ are (a) $(0.2915, 0.3473)$, (b) $(0.25, 0.33)$ and (c) $(0.0148, 0.2721)$.

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 $d^* = d_0 - d_1$.[2] In Figure 1a 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 1b when $d^* = 0.25$. The contours of $Q(d,\phi)$ in Figure 1c, 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

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

least mis-specified cases, correspond to the three different forms of asymptotic distribution presented in the following section.

4 Asymptotic Distributions

In this section we show that the key theoretical results derived in Chen and Deo (2006) pertaining to the asymptotic distribution of the FML estimator are also applicable to the Whittle, TML and CSS estimators. Writing $\widehat{\boldsymbol{\eta}}_1$ for any one of these estimators, the critical feature is that the rate of convergence and the nature of the asymptotic distribution of $\widehat{\boldsymbol{\eta}}_1$ is determined by the deviation of the pseudo-true value of d, d_1, from the true value, d_0; in Theorem 3 we summarize these different properties as they relate to three ranges of values for $d^* = d_0 - d_1 : d^* > 0.25$, $d^* = 0.25$ and $d^* < 0.25$.

Theorem 3 *Suppose that the TDGP of* $\{y_t\}$ *is a Gaussian process with a spectral density as given in* (1) *and that the MM satisfies Assumptions* $A.1 - A.7$. *Let*

$$\mathbf{B} = -2\int_{-\pi}^{\pi} \frac{f_0(\lambda)}{f_1^3(\boldsymbol{\eta}_1,\lambda)} \frac{\partial f_1(\boldsymbol{\eta}_1,\lambda)}{\partial \boldsymbol{\eta}} \frac{\partial f_1(\boldsymbol{\eta}_1,\lambda)}{\partial \boldsymbol{\eta}^T} d\lambda + \int_{-\pi}^{\pi} \frac{f_0(\lambda)}{f_1^2(\boldsymbol{\eta}_1,\lambda)} \frac{\partial^2 f_1(\boldsymbol{\eta}_1,\lambda)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T} d\lambda\,, \tag{21}$$

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* $\widehat{\boldsymbol{\eta}}_1$.[3] *Then the limiting distribution of the estimator is as follows:*

Case 1: When $d^* = d_0 - d_1 > 0.25$,

$$\frac{n^{1-2d^*}}{\log n}\left(\widehat{\boldsymbol{\eta}}_1 - \boldsymbol{\eta}_1 - \boldsymbol{\mu}_n\right) \to^D \mathbf{B}^{-1}\left(\sum_{j=1}^{\infty} W_j, 0, ...0\right)^T, \tag{22}$$

where $\sum_{j=1}^{\infty} W_j$ *is defined as the mean-square limit of the random sequence* $\sum_{j=1}^{s} W_j$ *as* $s \to \infty$, *wherein*

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

and $\{U_j\}$ *and* $\{V_k\}$ *denote sequences of Gaussian random variables with zero mean and covariances* $Cov_0\left(U_j, U_k\right) = Cov_0\left(U_j, V_k\right) = Cov_0\left(V_j, V_k\right)$ *with*

$$Cov_0\left(U_j, V_k\right) = \iint_{[0,1]^2} \{\sin(2\pi jx)\sin(2\pi ky) + \sin(2\pi kx)\sin(2\pi jy)\}\,|x-y|^{2d_0-1}\,dxdy\,.$$

Case 2: When $d^* = d_0 - d_1 = 0.25$,

$$n^{1/2}\left[\overline{\Lambda}_{dd}\right]^{-1/2}\left(\widehat{\boldsymbol{\eta}}_1 - \boldsymbol{\eta}_1\right) \to^D \mathbf{B}^{-1}\left(Z, 0, ..., 0\right)^T, \tag{23}$$

where

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

and Z *is a standard normal random variable.*

[3] Heuristically, $\boldsymbol{\mu}_n$ measures the bias associated with the estimator $\widehat{\boldsymbol{\eta}}_1$. That is, $\boldsymbol{\mu}_n \approx E_0\left(\widehat{\boldsymbol{\eta}}_1\right) - \boldsymbol{\eta}_1$. Note that the expression for $\boldsymbol{\mu}_n$ given in Chen and Deo (2006, p 263) is incorrect. The derivation of $\boldsymbol{\mu}_n$ for all four estimation methods considered in the paper is provided in Appendix B.

Case 3: When $d^* = d_0 - d_1 < 0.25$,

$$\sqrt{n}\,(\widehat{\boldsymbol{\eta}}_1 - \boldsymbol{\eta}_1) \to^D N(0, \boldsymbol{\Xi}), \tag{24}$$

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

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

We refer to Chen and Deo (2006, Theorems 1, 3 and 2) for details of the proof of Theorem 3 in the case of the FML estimator $\widehat{\boldsymbol{\eta}}_1^{(1)}$. For the Whittle, TML and CSS estimators we will 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

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

and

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

where $\|\boldsymbol{\eta}_1 - \dot{\boldsymbol{\eta}}_1^{(\cdot)}\| \leq \|\boldsymbol{\eta}_1 - \widehat{\boldsymbol{\eta}}_1^{(\cdot)}\|$. Since $\operatorname{plim} \widehat{\boldsymbol{\eta}}_1^{(\cdot)} = \boldsymbol{\eta}_1$ it is therefore sufficient to show that there exists a constant $\mathcal{C}$ independent of $\boldsymbol{\eta}$ such that

$$\frac{\partial^2 \{\mathcal{C} \cdot Q_n^{(i)}(\boldsymbol{\eta}_1) - Q_n^{(j)}(\boldsymbol{\eta}_1)\}}{\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}'} = o_p(1) \tag{25}$$

and

$$R_n\mathcal{C} \cdot \frac{\partial Q_n^{(i)}(\boldsymbol{\eta}_1)}{\partial \boldsymbol{\eta}} \to^D R_n \frac{\partial Q_n^{(j)}(\boldsymbol{\eta}_1)}{\partial \boldsymbol{\eta}}\,. \tag{26}$$

The condition in (25) is established by showing that for each $i = 1, 2, 3$ and 4 the Hessian $\partial^2\{Q_n^{(i)}(\boldsymbol{\eta}_1)\}/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'$ converges in probability to a matrix proportional to $\mathbf{B}$, as defined in (21). This result parallels the convergence of $Q_n^{(1)}(\boldsymbol{\eta})$ itself to the limiting objective function seen in (6), following the replacement of $f_1(\boldsymbol{\eta}_1, \lambda)^{-1}$ by $\partial^2\{f_1(\boldsymbol{\eta}_1, \lambda)^{-1}\}/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'$ and $Q(\boldsymbol{\eta})$ by $\mathbf{B}$. The proof that the Hessians so converge uses arguments similar those employed in the proof of Proposition 1, the details are therefore omitted. The proof of (26) is more involved because of the presence of the scaling factor R_n. In Appendix A we present the steps necessary to prove (26) for each estimator.

A key point to note from the three cases delineated in Theorem 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 $d^* = 0.25$).

5 Finite Sample Performance of the Mis-Specified Parametric Estimators of the Pseudo-True Parameter

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 refer to these estimators as $\widehat{d}_1^{(1)}$ (FML), $\widehat{d}_1^{(2)}$ (Whittle), $\widehat{d}_1^{(3)}$ (TML) and $\widehat{d}_1^{(4)}$ (CSS). 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. 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, MSE and relative efficiency 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^*.

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 (2001), 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.[4] However, we report selected results (only) from the full set due to space constraints. The bias, MSE and relative efficiency 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 1 *: An ARFIMA$(0, d_0, 1)$ TDGP, with parameter values $d_0 = \{0.2, 0.4\}$ and $\theta_0 = \{-0.7, -0.444978, -0.3\}$; and an ARFIMA$(0, d, 0)$ MM. 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 2 *: An ARFIMA$(0, d_0, 1)$ TDGP, with parameter values $d_0 = \{0.2, 0.4\}$ and $\theta_0 = \{-0.7, -0.637014, -0.3\}$; and an ARFIMA$(1, d, 0)$ MM. 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 Subsection 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 for $d_0 = 0.2$ only. In Subsection 5.3 we 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 both values of d_0.

5.2 Finite sample distributions

In this section we consider in turn the three cases listed under Theorem 3. For notational ease and clarity we use $\widehat{d}_1$ to denote the (generic) estimator obtained under mis-specification,

[4]Note that the scope of the experimental design is constrained by the restriction that the pseudo-true value d_1 implied by any choice of parameter values should lie in the interval $(0, 0.5)$.

remembering that this estimator may be produced by any one of the four estimation methods. 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.

5.2.1 Case 1: $d^* > 0.25$

The limiting distribution for $\widehat{d}_1$ in this case is

$$\frac{n^{1-2d^*}}{\log n}\left(\widehat{d}_1 - d_1 - \mu_n\right) \to^D b^{-1}\sum_{j=1}^{\infty} W_j\,, \tag{27}$$

where $\mu_n = b^{-1}E_0\left(\frac{\partial Q_n(\boldsymbol{\eta}_1)}{\partial d}\right)$,

$$\begin{aligned} b &= -2\int_{-\pi}^{\pi}\frac{f_0(\lambda)}{f_1^3(\boldsymbol{\eta}_1,\lambda)}\left(\frac{\partial f_1(\boldsymbol{\eta}_1,\lambda)}{\partial d}\right)^2 d\lambda + \int_{-\pi}^{\pi}\frac{f_0(\lambda)}{f_1^2(\boldsymbol{\eta}_1,\lambda)}\frac{\partial^2 f_1(\boldsymbol{\eta}_1,\lambda)}{\partial d^2}d\lambda \\ &= -2\int_0^{\pi}(1+\theta_0^2+2\theta_0\cos(\lambda))(2\sin(\lambda/2))^{-2d^*}(2\log(2\sin(\lambda/2)))^2 d\lambda\,, \end{aligned} \tag{28}$$

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. (With reference to Theorem 3, both $\mathbf{B}$ and $\boldsymbol{\mu}_n$ in (22) are here scalars since in Example 1 there is only one parameter to estimate under the MM, namely d. Hence the obvious changes made to notation. All other notation is as defined in the theorem.)

Given that the distribution in (27) 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 in probability 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 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 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

$$S_n = \widehat{Var}_0\left[\frac{n^{1-2d^*}}{\log n}\left(\widehat{d}_1 - d_1 - \mu_n\right)\right] \tag{29}$$

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

$$s = \arg\min_{1\leqslant m<\lfloor n/2\rfloor} T_m. \tag{30}$$

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

$$Var_0\left(\sum_{j=1}^{m} W_j\right) = \sum_{j=1}^{m} Var_0\left(W_j\right) + 2\sum_{j=1}^{m}\sum_{\substack{k=1\\ j\neq k}}^{m} Cov_0\left(W_j, W_k\right) . \tag{31}$$

The variance of W_j in this case is

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

As $\{U_j\}$ and $\{V_k\}$ are normal random variables with a covariance structure as specified in Theorem 3, 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,

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

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

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

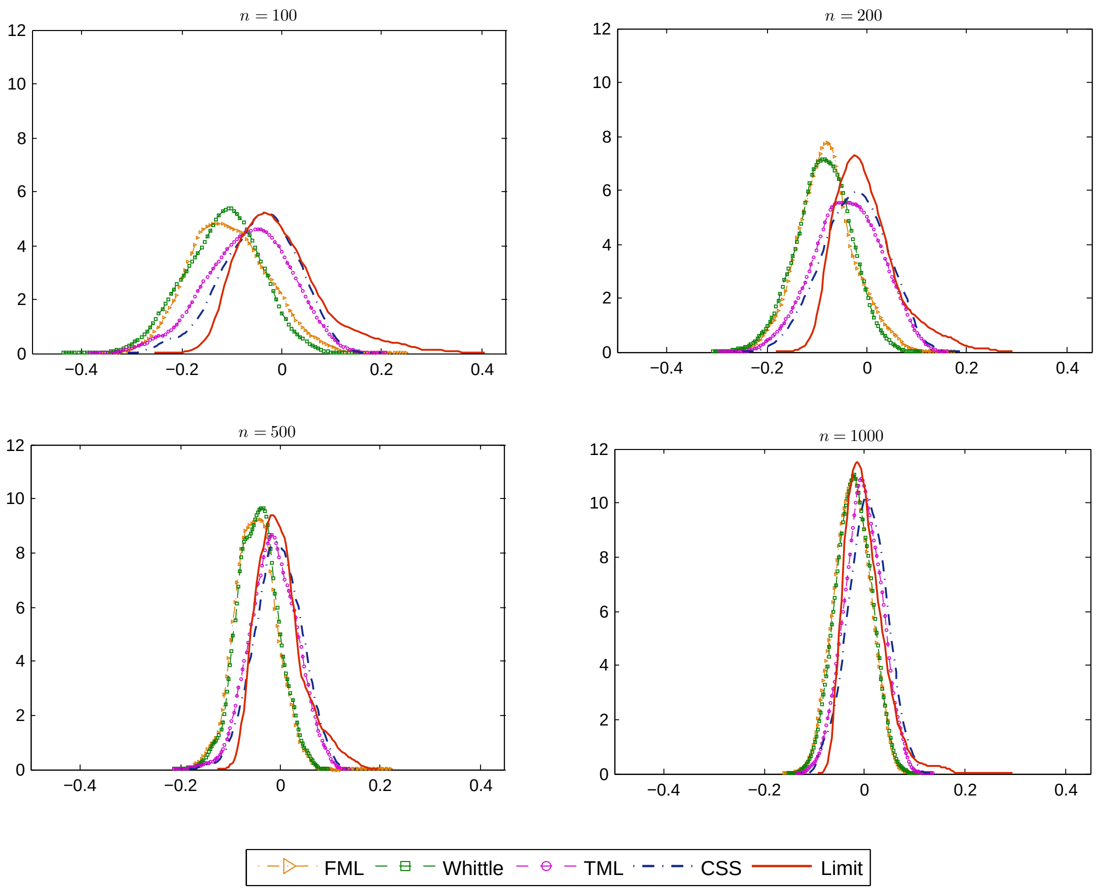
The expression in (31) can therefore be evaluated numerically using the formula for $Cov_0(U_j, V_k)$ to calculate the necessary moments required to determine s from (30).

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 (29) 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 (31) are also dependent on the form of both the TDGP and the MM 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 1.

Table 1 Truncation values s: ARFIMA $(0, d_0, 1)$ TDGP vis-à-vis ARFIMA $(0, d, 0)$ MM.

n	100	200	500	1000
s	36	75	162	230

Figure 2: 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$(0, d, 0)$ MM; $d^* > 0.25$.



Each panel in Figure 2 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 of s used are those given in Table 1. From Figure 2 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.

5.2.2 Case 2: $d^* = 0.25$

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

$$n^{1/2}[\overline{\Lambda}_{dd}]^{-1/2}\left(\widehat{d}_1 - d_1\right) \to^D N(0, b^{-2}), \tag{32}$$

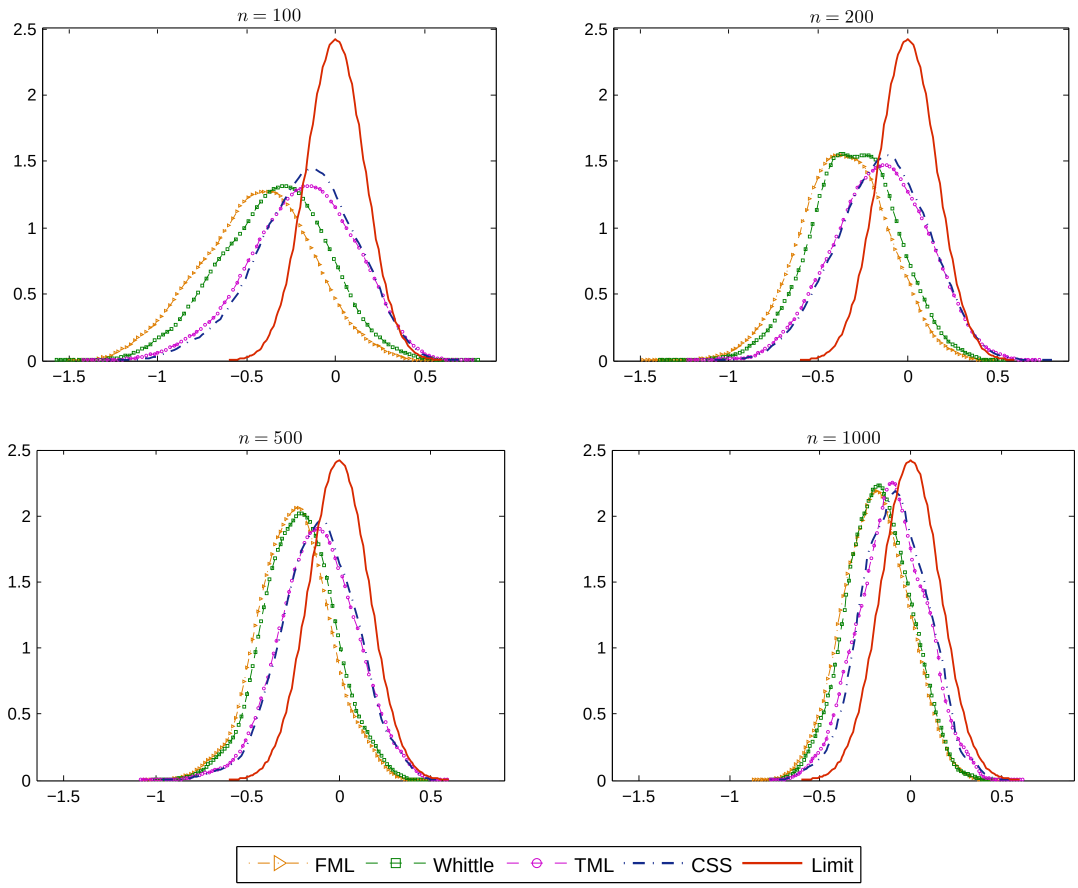
where

$$\overline{\Lambda}_{dd} = \frac{1}{n}\sum_{j=1}^{n/2}(1+\theta_0^2+2\theta_0\cos(\lambda_j))^2(2\sin(\lambda_j/2))^{-1}(2\log(2\sin(\lambda_j/2)))^2 \tag{33}$$

and b is as in (28). In both (33) and (28) $\theta_0 = -0.444978$, as $d^* = 0.25$ occurs at this particular value. Once again, $d_0 = 0.2$ in the TDGP.

Each panel of Figure 3 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 (32). Once again we observe a disparity between the time domain and frequency

Figure 3: 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$(0, d, 0)$ MM, d^*= 0.25.



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 $n = 1000$.

5.2.3 Case 3: $d^* < 0.25$

In this case we have

$$\sqrt{n}\left(\widehat{d}_1 - d_1\right) \to^D N(0, \upsilon^2)\,, \tag{34}$$

where

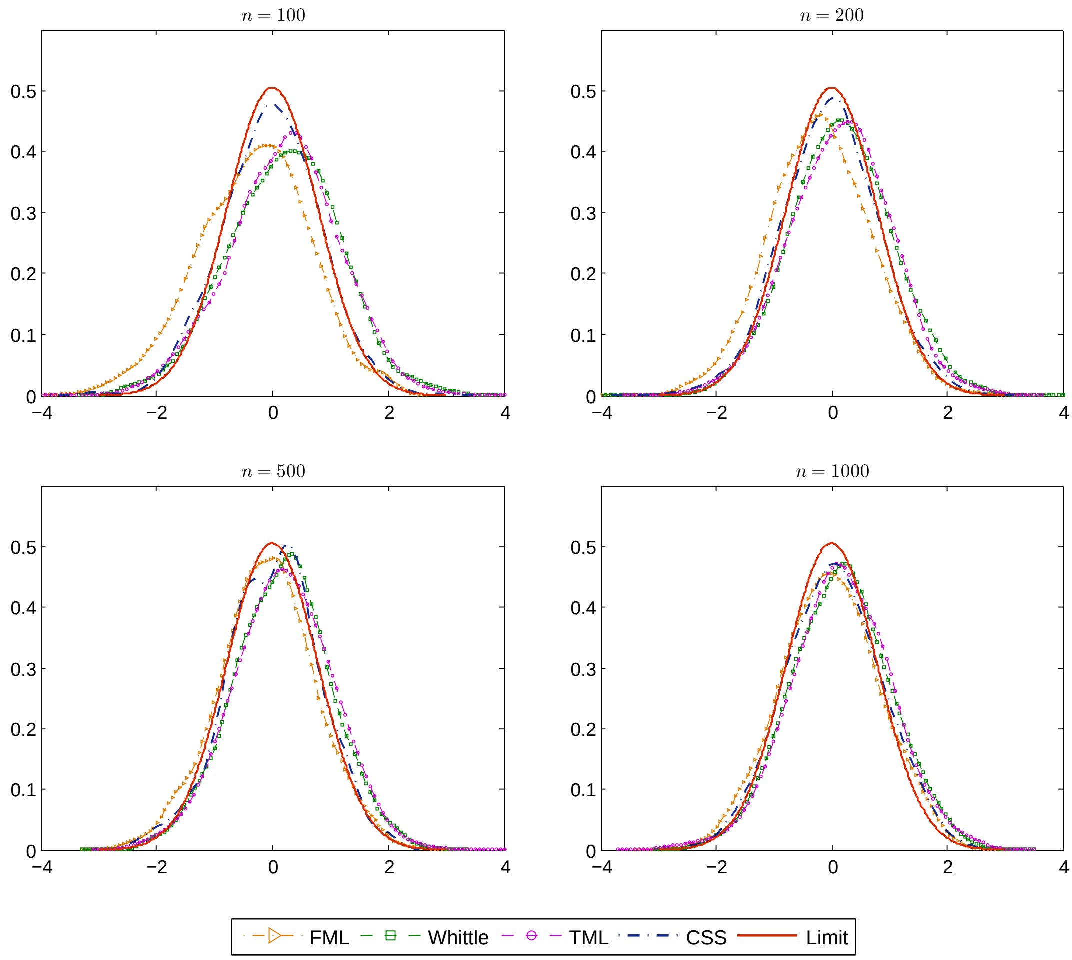
$$\upsilon^2 = \Lambda_{11}/b^{-2}\,, \tag{35}$$

with

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

and b as given in (28) evaluated at $\theta_0 = -0.3$ and $d^* = 0.1736$. Each panel in Figure 4 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 given in (34). In this case there is no clear visual differentiation between the time domain

Figure 4: 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$(0, d, 0)$ MM, $d^* < 0.25$.



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 $n = 1000$.

5.3 Finite sample bias and MSE of estimators of the pseudo-true parameter d_1

We supplement the graphical results in the previous section by documenting the finite sample bias, MSE and relative efficiency of the four alternative estimators, as estimators of the pseudo-true parameter d_1. The following standard formulae,

$$\widehat{\text{Bias}}_0\left(\widehat{d}_1^{(i)}\right) = \frac{1}{R}\sum_{r=1}^{R}\widehat{d}_r^{(i)} - d_1 \tag{36}$$

$$\widehat{Var}_0\left(\widehat{d}_1^{(i)}\right) = \frac{1}{R}\sum_{r=1}^{R}\left(\widehat{d}_{1,r}^{(i)}\right)^2 - \left(\frac{1}{R}\sum_{r=1}^{R}\widehat{d}_{1,r}^{(i)}\right)^2 \tag{37}$$

$$\widehat{\text{MSE}}_0\left(\widehat{d}_1^{(i)}\right) = \widehat{\text{Bias}}_0^2 + \widehat{Var}_0\left(\widehat{d}_1^{(i)}\right) \tag{38}$$

$$\widehat{r.eff}_0\left(\widehat{d}_1^{(i)}, \widehat{d}_1^{(j)}\right) = \frac{\widehat{\text{MSE}}_0\left(\widehat{d}_1^{(i)}\right)}{\widehat{\text{MSE}}_0\left(\widehat{d}_1^{(j)}\right)}, \tag{39}$$

are applied to all four estimators $i, j = 1, ..., 4$. Since all empirical expectations and variances are evaluated under the TGDP, we make this explicit with appropriate subscript notation. Results are produced for Example 1 in Tables 2 and 5 and for Example 2 in Tables 3 and 5, with additional results in Table 4. Values of $d^* = d_0 - d_1$ are documented across the key ranges, $d^* \lesseqqgtr 0.25$, along with associated values for the MA coefficient in the TGDP, 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.[5]

Consider first the bias and MSE results for Example 1 with $d_0 = 0.2$ displayed in the top panel of Table 2. As is consistent with the theoretical results (and the graphical illustration in the previous section) the bias and MSE of all four parametric estimators 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 MM 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 $d^* < 0.25$.

As indicated by the results in the bottom panel of Table 2 for $d_0 = 0.4$, the impact of an increase in d_0 (for any given value of d^* and n) is to (usually but not uniformly) increase the bias and MSE of all estimators, as estimators of d_1. That is, the ability of the four estimators to accurately estimate the pseudo-true parameter for which they are consistent tends to decline (overall) as the long memory in the TDGP increases. Nevertheless, these results show that the relativities between the estimators remain essentially the same as for the smaller value of d_0, with the CSS estimator now being uniformly preferable to all other estimators under mis-specification, and the FML estimator still performing the worst of all.

The results recorded in Table 3 for Example 2 illustrate that the presence of an AR term in the MM 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 tend to be smaller in (absolute) value than the corresponding values in Table 2. Results not presented here suggest, however, that when the value of θ_0 is near zero, estimation under

[5] 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 2 Estimates of the bias and MSE of $\widehat{d_1}$ for the FML, Whittle, TML and CSS estimators: Example 1.

			FML		Whittle		TML		CSS	
d^*	θ_0	n	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
			ARFIMA $(0, d_0, 1)$ TDGP d_0= 0.2 vis-à-vis ARFIMA $(0, d, 0)$ MM							
0.3723	-0.7	100	-0.1781	0.0915	-0.2466	0.0691	-0.1748	0.0481	**-0.1427**	**0.0315**
		200	-0.1620	0.0558	-0.1940	0.0431	-0.1287	0.0335	**-0.1110**	**0.0207**
		500	-0.1354	0.0211	-0.1308	0.0178	-0.0916	0.0138	**-0.0798**	**0.0097**
		1000	-0.1019	0.0141	-0.0996	0.0127	-0.0776	0.0103	**-0.0670**	**0.0065**
0.2500	-0.44	100	-0.1515	0.0393	-0.1184	0.0298	-0.0650	0.0170	**-0.0577**	**0.0119**
		200	-0.1010	0.0148	-0.0852	0.0117	-0.0434	0.0072	**-0.0400**	**0.0052**
		500	-0.0544	0.0048	-0.0487	0.0042	-0.0257	0.0027	**-0.0241**	**0.0021**
		1000	-0.0351	0.0023	-0.0323	0.0021	-0.0188	0.0015	**-0.0162**	**0.0012**
0.1736	-0.3	100	-0.1082	0.0217	-0.0712	0.0146	-0.0340	0.0095	**-0.0330**	**0.0087**
		200	-0.0663	0.0085	-0.0491	0.0064	**-0.0213**	0.0047	-0.0228	**0.0045**
		500	-0.0318	0.0026	-0.0251	0.0022	**-0.0106**	0.0017*	-0.0188	**0.0017**
		1000	-0.0184	0.0011	-0.0149	0.0010	**-0.0065**	0.0009*	-0.0180	**0.0009**
			ARFIMA $(0, d_0, 1)$ TDGP d_0= 0.4 vis-à-vis ARFIMA $(0, d, 0)$ MM							
0.3723	-0.7	100	-0.2786	0.0995	-0.2456	0.0724	-0.2210	0.0515	**-0.1957**	**0.0489**
		200	-0.2096	0.0601	-0.1942	0.0440	-0.1778	0.0357	**-0.1648**	**0.0340**
		500	-0.1598	0.0213	-0.1287	0.0181	-0.1347	0.0137	**-0.0871**	**0.0118**
		1000	-0.1123	0.0157	-0.0939	0.0143	-0.0812	0.0121	**-0.0648**	**0.0117**
0.2500	-0.44	100	-0.1903	0.0475	-0.1659	0.0383	-0.0911	0.0201	**-0.0550**	**0.0138**
		200	-0.1362	0.0237	-0.1227	0.0195	-0.0534	0.0103	**-0.0421**	**0.0089**
		500	-0.0796	0.0095	-0.0550	0.0082	-0.0249	0.0059	**-0.0224**	**0.0038**
		1000	-0.0360	0.0048	-0.0295	0.0042	-0.0180	0.0035	**-0.0175**	**0.0025**
0.1736	-0.3	100	-0.0990	0.0228	-0.0843	0.0152	-0.0422	0.0102	**-0.0321**	**0.0092**
		200	-0.0773	0.0092	-0.0505	0.0071	-0.0244	0.0057	**-0.0199**	**0.0048**
		500	-0.0407	0.0031	-0.0276	0.0025	-0.0129	0.0022	**-0.0087**	**0.0019**
		1000	-0.0172	0.0011	-0.0163	0.0010	-0.0077	0.0009	**-0.0052**	**0.0008**

the MM with an extraneous AR parameter causes an increase in (absolute) bias and MSE, relative to the case where the MM 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: 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 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 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,

Table 3 Estimates of the bias and MSE of $\widehat{d_1}$ for the FML, Whittle, TML and CSS estimators: Example 2.

			FML		Whittle		TML		CSS	
d^*	θ_0	n	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
ARFIMA $(0, d_0, 1)$ TDGP d_0= 0.2 vis-à-vis ARFIMA $(1, d, 0)$ MM										
0.2915	-0.7	100	-0.1612	0.0541	-0.1169	0.0342	-0.0950	0.0295	**-0.0671**	**0.0236**
		200	-0.1143	0.0376	-0.0941	0.0262	-0.0760	0.0213	**-0.0482**	**0.0175**
		500	-0.0679	0.0165	-0.0604	0.0125	-0.0454	0.0110	**-0.0369**	**0.0089**
		1000	-0.0469	0.0089	-0.0432	0.0071	**-0.0250**	0.0067	-0.0303	**0.0059**
0.25	-0.64	100	-0.1339	0.0279	-0.0899	0.0175	-0.0655	0.0138	**-0.0457**	**0.0110**
		200	-0.0902	0.0125	-0.0700	0.0086	-0.0490	0.0067	**-0.0345**	**0.0062**
		500	-0.0490	0.0041	-0.0415	0.0030	-0.0323	0.0026	**-0.0230**	**0.0022**
		1000	-0.0316	0.0019	-0.0281	0.0015	-0.0181	0.0013	**-0.0176**	**0.0011**
0.0148	-0.3	100	-0.0508	0.0139	-0.0256	0.0086	0.0190	0.0067	**-0.0082**	**0.0054**
		200	-0.0266	0.0053	-0.0135	0.0036	0.0168	0.0028	**-0.0081**	**0.0025**
		500	-0.0093	0.0027	-0.0080	0.0019	0.0073	0.0016	**-0.0004**	**0.0014**
		1000	-0.0036	0.0010	-0.0023	0.0008	0.0067	0.0006*	**0.0003**	**0.0006**
ARFIMA $(0, d_0, 1)$ TDGP d_0= 0.4 vis-à-vis ARFIMA $(1, d, 0)$ MM										
0.2915	-0.7	100	-0.2299	0.0639	-0.1805	0.0419	-0.1279	0.0372	**-0.0699**	**0.0140**
		200	-0.1774	0.0395	-0.1599	0.0282	-0.1034	0.0245	**-0.0578**	**0.0190**
		500	-0.1294	0.0197	-0.1039	0.0150	-0.0816	0.0126	**-0.0294**	**0.0101**
		1000	-0.1089	0.0125	-0.0632	0.0099	-0.0462	0.0081	**-0.0109**	**0.0069**
0.25	-0.64	100	-0.1396	0.0257	-0.0979	0.0155	-0.0692	0.0145	**-0.0508**	**0.0103**
		200	-0.0868	0.0122	-0.0675	0.0077	-0.0401	0.0076	**-0.0357**	**0.0058**
		500	-0.0455	0.0065	-0.0342	0.0046	-0.0294	0.0041	**-0.0216**	**0.0033**
		1000	-0.0316	0.0027	-0.0192	0.0021	-0.0177	0.0018	**-0.0122**	**0.0014**
0.0148	-0.3	100	-0.0650	0.0162	-0.0422	0.0115	0.0246	0.0082	**-0.0132**	**0.0067**
		200	-0.0312	0.0095	-0.0164	0.0075	0.0107	0.0053	**-0.0094**	**0.0047**
		500	-0.0205	0.0042	-0.0133	0.0034	0.0079	0.0026	**-0.0035**	**0.0023**
		1000	-0.0136	0.0021	-0.0088	0.0018	0.0053	0.0014	**-0.0017**	**0.0013**

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. This latter result is in accord with the findings in Nielsen and Frederiksen (2005), where the TML estimator is compared with the CSS and Whittle estimators for a fractional noise model and a deterioration in relative performance of the TML estimator as a result of estimating the unknown mean is observed, an effect previously documented in Cheung and Diebold (1994). □

The results in Tables 2, 3 and 4 highlight that the CSS estimator has the smallest MSE of all four 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. The absolute value of its bias is also the smallest in the vast majority of such cases. This superiority presumably reflects a certain in-built robustness of least squares methods to mis-specification and incorrect model formulation. This is further emphasized in Table 5 which

Table 4 Estimates of the bias and MSE of $\widehat{d_1}$ for the FML, Whittle, TML and CSS estimators: ARFIMA $(0, d_0, 0)$ TDGP $d_0 = 0.2$, $d^* = 0.0$.

	FML		Whittle		TML		CSS	
n	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
				Correct ARFIMA $(0, d, 0)$ model				
100	-0.0502	0.0113	-0.0173	0.0102	**0.0066**	**0.0087**	0.0094	0.0096
200	-0.0279	0.0044	-0.0110	0.0041	**0.0043**	**0.0037**	0.0063	0.0039
500	-0.0089	0.0015	-0.0062	0.0014	**0.0026**	**0.0013**	0.0031	0.0014
1000	-0.0045	0.0006*	-0.0037	0.0006*	**0.0016**	**0.0006**	0.0025	0.0006*
				Over-Parameterized ARFIMA $(1, d, 0)$ model				
100	-0.0455	0.0177	0.0371	0.0121	0.0255	0.0107	**0.0158**	**0.0087**
200	-0.0216	0.0081	0.0196	0.0058	0.0107	0.0052	**0.0092**	**0.0042**
500	-0.0120	0.0065	0.0091	0.0049	0.0078	0.0043	**0.0055**	**0.0037**
1000	-0.0074	0.0027	0.0055	0.0021	0.0034	0.0019	**0.0028**	**0.0016**

records the relative efficiencies of the estimators. The relative efficiencies are calculated by taking the ratio of the MSE of $\widehat{d_1}$ for all estimation methods to the MSE of the FML estimator, as per (39), and for each combination of d_0, θ_0 and n the minimum MSE ratio is bolded. The relative efficiency results recorded in Table 5 confirm that the CSS estimator is between (approximately) 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. Also, the MSE of the FML estimator exceeds the corresponding values for all three other estimators, with all relative efficiency values recorded in Table 5 being less than one. Accordingly, across all parameter settings we have documented where mis-specification or incorrect model formulation obtains, the CSS estimator is almost universally dominant.

6 Summary and Conclusions

This paper presents theoretical and simulation-based results relating to the estimation of mis-specified models for long range dependent processes. We show that under mis-specification four classical parametric estimation methods, frequency domain maximum likelihood (FML), Whittle, time domain maximum likelihood (TML) and conditional sum of squares (CSS) converge to the same pseudo-true parameter value. A general closed-form solution for the limiting criterion function for the four alternative parametric estimation methods 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, 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. 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

Table 5 Estimates of the efficiency of the Whittle, TML and CSS estimators of the long memory parameter relative to the FML estimator: Examples 1 and 2

			Whittle	TML	CSS	Whittle	TML	CSS
d^*	θ_0	n		d_0= 0.2			d_0= 0.4	
			ARFIMA $(0, d_0, 1)$ TDGP vis-à-vis ARFIMA $(0, d, 0)$ MM					
0.3723	-0.7	100	0.7552	0.5257	**0.3443**	0.7276	0.5176	**0.4915**
		200	0.7724	0.6004	**0.3710**	0.7321	0.5940	**0.5657**
		500	0.8436	0.6540	**0.4597**	0.8498	0.6432	**0.5540**
		1000	0.9007	0.7305	**0.4610**	0.9108	0.7707	**0.7452**
0.2500	-0.44	100	0.7583	0.4326	**0.3028**	0.8063	0.4232	**0.2905**
		200	0.7905	0.4865	**0.3514**	0.8228	0.4346	**0.3755**
		500	0.8750	0.5625	**0.4375**	0.8632	0.6211	**0.4000**
		1000	0.9130	0.6522	**0.5217**	0.8750	0.7292	**0.5208**
0.1736	-0.3	100	0.6728	0.4378	**0.4009**	0.6667	0.4474	**0.4035**
		200	0.7529	0.5529	**0.5294**	0.7717	0.6196	**0.5217**
		500	0.8462	0.6538	**0.6362**	0.8065	0.7097	**0.6129**
		1000	0.9091	0.8182	**0.7730**	0.9091	0.8182	**0.7636**
			ARFIMA $(0, d_0, 1)$ TDGP vis-à-vis ARFIMA $(1, d, 0)$ MM					
0.2915	-0.7	100	0.6322	0.5453	**0.4362**	0.6557	0.5822	**0.4224**
		200	0.6968	0.5665	**0.4654**	0.7139	0.6203	**0.4810**
		500	0.7576	0.6667	**0.5394**	0.7614	0.6396	**0.5127**
		1000	0.7978	0.7528	**0.6629**	0.7920	0.6480	**0.5520**
0.25	-0.64	100	0.6272	0.4946	**0.3943**	0.6031	0.5642	**0.4008**
		200	0.6880	0.5360	**0.4960**	0.6311	0.6230	**0.4754**
		500	0.7317	0.6341	**0.5366**	0.7077	0.6308	**0.5077**
		1000	0.7895	0.6842	**0.5789**	0.7778	0.6667	**0.5185**
0.0148	-0.3	100	0.6187	0.4820	**0.3885**	0.7099	0.5062	**0.4136**
		200	0.6792	0.5283	**0.4717**	0.7895	0.5579	**0.4947**
		500	0.7148	0.5926	**0.5185**	0.8095	0.6190	**0.5476**
		1000	0.7632	0.6400	**0.5600**	0.8571	0.6667	**0.6190**

pseudo-true parameter for which they are both consistent.

There are several interesting issues that arise from the results that we have established, including the following: First, the necessity to suppose that $\{y_t\}$ is a Gaussian process in order to appeal to existing results in the literature where this assumption is made is unfortunate. It seems reasonable to suppose that our results can be extended to long range dependent linear processes, given that under current assumptions the series will have such a representation, but extension to more general processes is not likely to be straightforward. Second, a relaxation of the restriction that only values of $d \in (0, 0.5)$ be considered seems desirable, particularly as the relationship between the true value d_0 and the pseudo-true value d_1 depends upon the interaction between the TDGP and the MM and $d_0 \in (0, 0.5)$ does not imply the same is true of d_1. The extension of our results to short memory, $d = 0$, anti-persistent, $d < 0$, and non-stationary, $d \geq 0.5$, cases will facilitate the consideration of a broader range of circumstances. To some extent other values of d might be covered by means of appropriate pre-filtering, for example, the use of first-differencing when $d \in (1, 3/2)$, 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 non-stationary case with $d \in (0, 3/2)$, say, 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, 1999) 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 $\mathbf{\Lambda}$. 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) 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.

References

Beran, J. (1994). *Statistics for Long-Memory Processes*, vol. 61 of *Monographs on Statistics and Applied Probability.* Chapman and Hall, New York.

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* **B 57** 654–672.

Box, G. and Jenkins, G. (1970). *Time Series Analysis: Forecasting and Control.* Holden Day, San Francisco.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods.* Springer Series in Statistics. Springer-Verlag, New York, 2nd ed.

Chen, W. W. and Deo, R. S. (2006). Estimation of mis-specified long memory models. *Journal of Econometrics* **53** 257 – 281.

Cheung, Y. W. and Diebold, F. X. (1994). On maximum likelihood estmation of the differencing parameter of fractionally integrated noise with unknown mean. *Journal of Econometrics* **62** 301–316.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Annals of Statistics* **17** 1749–1766.

Doornik, J. A. and Ooms, M. (2001). Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. *Computational Statistics & Data Analysis* **42** 333–348. Also a 2001 Nuffield discussion paper.

Fox, R. and Taqqu, M. S. (1986). Large sample properties of parameter estimates for strongly dependent stationary gaussian time series. *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.

Gradshteyn, I. S. and Ryzhik, I. M. (2007). *Tables of Integrals, Series and Products.* Academic Press, Sydney.

Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Times Series.* J. Wiley, New York.

Grenander, U. and Szego, G. (1958). *Toeplitz Forms and Their Application.* University of California Press, Berkeley.

Hannan, E. J. (1973). The asymptotic theory of linear time series models. *Advances in Applied Probability* **10** 130–145.

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.

Moulines, E. and Soulier, P. (1999). Broad band log-periodogram regression of time series with long range dependence. *Annals of Statistics* **27** 1415–1439.

Nielsen, M. . and Frederiksen, P. H. (2005). Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. *Econometric Reviews* **24** 405–443.

Robinson, P. M. (1995a). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* **23** 1630–1661.

Robinson, P. M. (1995b). Log periodogram regression of time series with long memory. *Annals of Statistics* **23** 1048–1072.

Robinson, P. M. (2006). *Time Series and Related Topics: In Memory of Ching-Zong Wei*, vol. 52 of *IMS Lecture Notes–Monograph Series*, chap. Conditional-Sum-of-Squares Estimation of Models for Stationary Time Series with Long Memory. Institue of Mathematical Statistics, Beachwood, 130–137.

Sowell, F. (1992). Maximum likelihood estmation 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.

Velasco, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis* **20** 87–127.

Walker, A. M. (1964). Asymptotic properties of least squares estimates of the parameters of the spectrum of a stationary non–deterministic time series. *Journal of the Australian Mathematical Society* **4** 363–384.

Whittle, P. (1962). Gaussian estimation in stationary time series. *Bulletin International Statistical Institute* **39** 105–129.

Whittle, P. (1984). *Prediction and Regulation By Linear Least Squares Methods.* Basil Blackwell.

Yajima, Y. (1992). 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, eds., *New Directions in Time Series Analysis, Part II*, vol. 46 of *IMA Volumes in Mathematics and Its Applications.* Springer-Verlag, New York, 375–382.

Appendix A: Proofs

In the proofs we will need to consider stochastic Rieman-Stieltges integrals of the periodogram. These are dealt with in the following lemma.

Lemma A.1 *Assume that $I(\lambda)$ is calculated from a realization of a stationary Gaussian process with a spectral density as given in* (1)*, and that $h(\lambda)$ is an even valued periodic function with period 2π that is continuously differentiable on $(0,\pi]$. Set*

$$\nabla_I(h) = \int_0^\pi I(\lambda)h(\lambda)d\lambda - \frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor} I(\lambda_j)h(\lambda_j)\,.$$

Then $\nabla_I(h) = O_p(n^{-1})$ and $\lim_{n\to\infty}|\nabla_I(h)| = 0$ almost surely.

Proof. Using the partition of $(0,\pi]$ induced by $\lambda_j = 2\pi j/n$, $j = 1,\ldots,\lfloor n/2\rfloor$, gives the decomposition

$$\begin{aligned}\nabla_I(h) = \int_0^{2\pi/n} I(\lambda)h(\lambda)d\lambda + &\sum_{j=1}^{\lfloor n/2\rfloor-1}\int_{2\pi j/n}^{2\pi(j+1)/n}\{I(\lambda)h(\lambda) - I(\lambda_j)h(\lambda_j)\}d\lambda \\ &+ \int_{2\pi\lfloor n/2\rfloor/n}^{\pi} I(\lambda)h(\lambda)d\lambda - I(\lambda_{\lfloor n/2\rfloor})h(\lambda_{\lfloor n/2\rfloor})\frac{2\pi}{n}\,,\end{aligned}$$

which can be rearranged to give $\nabla_I(h) = T_1 + T_2 + T_3$ where $T_1 = \int_0^{2\pi/n} I(\lambda)h(\lambda)d\lambda$,

$$T_2 = \sum_{j=1}^{\lfloor n/2\rfloor-1}\int_{2\pi j/n}^{2\pi(j+1)/n} I(\lambda)\{h(\lambda) - h(\lambda_j)\}d\lambda + \int_{2\pi\lfloor n/2\rfloor/n}^{\pi} I(\lambda)\{h(\lambda) - h(\lambda_{\lfloor n/2\rfloor})\}d\lambda$$

and

$$T_3 = \sum_{j=1}^{\lfloor n/2\rfloor-1}\int_{2\pi j/n}^{2\pi(j+1)/n}\{I(\lambda) - I(\lambda_j)\}h(\lambda_j)d\lambda + \int_{2\pi\lfloor n/2\rfloor/n}^{\pi} I(\lambda)h(\lambda_{\lfloor n/2\rfloor})d\lambda - I(\lambda_{\lfloor n/2\rfloor})h(\lambda_{\lfloor n/2\rfloor})\frac{2\pi}{n}\,.$$

By the Mean Value Theorem (first for integrals and then derivatives), the first term $T_1 = I(\lambda')h(\lambda')\frac{2\pi}{n}$, $\lambda' \in (0, \lambda_1)$, and $I(\lambda') - I(\lambda_1) = I'(\lambda")(\lambda_1 - \lambda')$, $\lambda" \in (\lambda', \lambda_1)$, so $|T_1| \leq (I(\lambda_1)\frac{2\pi}{n} + |I'(\lambda")|(\frac{2\pi}{n})^2)|h(\lambda')|$. For the second term we have

$$|T_2| \leq \sum_{j=1}^{\lfloor n/2 \rfloor} \int_{2\pi j/n}^{2\pi(j+1)/n} I(\lambda)|h(\lambda) - h(\lambda_j)|d\lambda\,.$$

But for all $\lambda \in (2\pi j/n, 2\pi(j+1)/n)$, $|h(\lambda) - h(\lambda_j)| \leq \mathcal{M}\frac{2\pi}{n}$ where $\mathcal{M} = \sup_{\lambda \in [\frac{2\pi}{n}, \pi]} |h'(\lambda)|$. Hence we can conclude that $|T_2| \leq 2\int_0^\pi I(\lambda)d\lambda\mathcal{M}2\pi/n = 2\pi\mathcal{M}\sum_{t=1}^n y_t^2/n^2$. Similarly,

$$|T_3| \leq \sum_{j=1}^{\lfloor n/2 \rfloor} I(\lambda_j) \int_{2\pi j/n}^{2\pi(j+1)/n} \left|\frac{I(\lambda)}{I(\lambda_j)} - 1\right| |h(\lambda_j)|d\lambda$$

and it follows that $|T_3| \leq \frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2 \rfloor} I(\lambda_j)\mathcal{M}'2\pi/n = 2\pi\mathcal{M}'\sum_{t=1}^n y_t^2/n^2$ where

$$\mathcal{M}' = \frac{\sup_{\lambda \in [\frac{2\pi}{n}, \pi]} |I'(\lambda)| \sup_{\lambda \in [\frac{2\pi}{n}, \pi]} |h(\lambda)|}{\inf_{\lambda \in [\frac{2\pi}{n}, \pi]} I(\lambda)}\,.$$

By Lemma 4 of Moulines and Soulier (1999) $I(\lambda_1)/f_0(\lambda_1)$ converges in distribution to a $\chi^2(2)$ variate, and by Theorem 4 of Hosking (1996) $\sum_{t=1}^n y_t^2/n$ converges to $E_0(y_t^2)$. We can therefore conclude that $\nabla_I(h)$ is bounded above by three terms each of order $O_p(n^{-1})$. Moreover, since each term has a variance of order $O(n^{-2})$ or smaller it follows from Markov's inequality and the Borel-Cantelli lemma that $\nabla_I(h)$ converges to zero almost surely. ■

A.1: Proof of Proposition 1:

A.1.1: Whittle estimation

Following the development in Beran (1994, p. 116) we have

$$\lim_{n\to\infty} \frac{4}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \log f_1(\boldsymbol{\eta}, \lambda_j) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f_1(\boldsymbol{\eta}, \lambda)d\lambda,$$

where

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

From standard results for trigonometric integrals in Gradshteyn and Ryzhik (2007, p. 583) we have

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

and by Assumption A.6 $\int_{-\pi}^{\pi} \log g_1(\boldsymbol{\beta},\lambda)d\lambda = 0$. The limit of the first component of $Q_n^{(2)}(\sigma_\varepsilon^2, \boldsymbol{\eta})$ is therefore $2\log\left(\sigma_\varepsilon^2/2\pi\right)$. Applying the result in (6) to the second component it follows that

$$Q_n^{(2)}(\sigma_\varepsilon^2, \boldsymbol{\eta}) \to^p \mathcal{Q}^{(2)}(\sigma_\varepsilon^2, Q(\boldsymbol{\eta})) = 2\log\left(\frac{\sigma_\varepsilon^2}{2\pi}\right) + \frac{4}{\sigma_\varepsilon^2}Q(\boldsymbol{\eta})$$

uniformly in σ_ε^2 and $\boldsymbol{\eta}$. Concentrating $\mathcal{Q}^{(2)}(\sigma_\varepsilon^2, Q(\boldsymbol{\eta}))$ with respect to σ_ε^2, as the parameter of interest here is $\boldsymbol{\eta}$, we find that the minimum of $\mathcal{Q}^{(2)}(\sigma_\varepsilon^2, Q(\boldsymbol{\eta}))$ is given by $2\log(Q(\boldsymbol{\eta}_1))/\pi)+2$. Thus we conclude that plim $\widehat{\boldsymbol{\eta}}_1^{(2)} = \boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_1$ is the pseudo-true parameter for the Whittle estimator. ■

A.1.2: Time domain maximum likelihood estimation

From Grenander and Szego (1958) we have

$$\lim_{n\to\infty} \frac{1}{n} \log |\boldsymbol{\Sigma}_\eta| = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f_1(\boldsymbol{\eta},\lambda) d\lambda = 0$$

for the second term in (9). To determine the limit of the third component set $\mathbf{A}_\eta = [a_{s-r}(\eta)]$ where

$$a_{s-r}(\eta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{f_1(\boldsymbol{\eta},\lambda)} \exp(i(s-l)\lambda) d\lambda\,, \quad r,s = 1,\ldots,n\,. \tag{A.1}$$

Then

$$\begin{aligned}
\frac{1}{n}\mathbf{Y}^T\mathbf{A}_\eta\mathbf{Y} &= \frac{1}{n}\sum_{k=0}^{n-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{f_1(\boldsymbol{\eta},\lambda)} \exp(ik\lambda) d\lambda \sum_{t=k}^{n} y_t y_{t-k} \\
&= \int_{-\pi}^{\pi} \frac{1}{f_1(\boldsymbol{\eta},\lambda)} \left(\frac{1}{2\pi n} \sum_{k=0}^{n-1} \exp(ik\lambda) \sum_{t=k}^{n} y_t y_{t-k} \right) d\lambda \\
&= \int_{-\pi}^{\pi} \frac{I(\lambda)}{f_1(\boldsymbol{\eta},\lambda)} d\lambda\,.
\end{aligned}$$

From the triangular inequality we have

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

and it follows by Lemma A.1 and application of (6) that as $n\to\infty$

$$\left| \frac{1}{n}\mathbf{Y}^T\mathbf{A}_\eta\mathbf{Y} - \int_{-\pi}^{\pi} \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta},\lambda)} d\lambda \right| \to^p 0\,. \tag{A.2}$$

Now, $E_0[\mathbf{Y}^T(\boldsymbol{\Sigma}_\eta^{-1} - \mathbf{A}_\eta)\mathbf{Y}] = \mathrm{tr}(\boldsymbol{\Sigma}_\eta^{-1} - \mathbf{A}_\eta)\boldsymbol{\Sigma}_0$, where $\boldsymbol{\Sigma}_0 = E_0[\mathbf{Y}\mathbf{Y}^T]$, and $|\mathrm{tr}(\boldsymbol{\Sigma}_\eta^{-1} - \mathbf{A}_\eta)\boldsymbol{\Sigma}_0| \le \|\mathbf{I} - \boldsymbol{\Sigma}_\eta^{1/2}\mathbf{A}_\eta\boldsymbol{\Sigma}_\eta^{1/2}\|\|\boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Sigma}_0\|$. By Lemma 5.2 of Dahlhaus (1989) $\|\mathbf{I} - \boldsymbol{\Sigma}_\eta^{1/2}\mathbf{A}_\eta\boldsymbol{\Sigma}_\eta^{1/2}\| = O(n^\delta)$ for all $\delta \in (0, d/2)$ and Theorem 5.1 of Dahlhaus (1989) implies that $\|\boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Sigma}_0\| = O(n^{1/2})$. It follows that $|\mathrm{tr}(\boldsymbol{\Sigma}_\eta^{-1} - \mathbf{A}_\eta)\boldsymbol{\Sigma}_0| = O(n^{1/2+\delta})$. Similarly $Var_0[\mathbf{Y}^T(\boldsymbol{\Sigma}_\eta^{-1} - \mathbf{A}_\eta)\mathbf{Y}] = \mathrm{tr}((\boldsymbol{\Sigma}_\eta^{-1} - \mathbf{A}_\eta)\boldsymbol{\Sigma}_0)^2 = \|(\boldsymbol{\Sigma}_\eta^{-1} - \mathbf{A}_\eta)\boldsymbol{\Sigma}_0\|^2 = O(n^{1+2\delta})$. Markov's inequality therefore implies that

$$Pr\left(n^{-1}\left|\mathbf{Y}^T(\boldsymbol{\Sigma}_\eta^{-1} - \mathbf{A}_\eta)\mathbf{Y}\right| > \epsilon\right) = O(n^{-(1-2\delta)})$$

for all $\epsilon > 0$, and hence $\mathrm{plim}_{n\to\infty} n^{-1}\left|\mathbf{Y}^T\boldsymbol{\Sigma}_\eta^{-1}\mathbf{Y} - \mathbf{Y}^T\mathbf{A}_\eta\mathbf{Y}\right| = 0$. We thus have

$$\frac{1}{n}\mathbf{Y}^T\boldsymbol{\Sigma}_\eta^{-1}\mathbf{Y} \to^p \int_{-\pi}^{\pi} \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta},\lambda)} d\lambda\,.$$

The limiting value of the criterion function $Q_n^{(3)}(\sigma_\varepsilon^2, \boldsymbol{\eta})$ is therefore

$$\begin{aligned}\mathcal{Q}^{(3)}(\sigma_\varepsilon^2, Q(\boldsymbol{\eta})) &= \log \sigma_\varepsilon^2 + \frac{1}{\sigma_\varepsilon^2}\int_{-\pi}^{\pi} \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta},\lambda)} d\lambda \\ &= \log \sigma_\varepsilon^2 + \frac{2Q(\boldsymbol{\eta})}{\sigma_\varepsilon^2}\end{aligned}$$

uniformly in σ_ε^2 and $\boldsymbol{\eta}$. Concentrating $\mathcal{Q}^{(3)}(\sigma_\varepsilon^2, Q(\boldsymbol{\eta}))$ with respect to σ_ε^2 we find that the minimum of $\mathcal{Q}^{(3)}(\sigma_\varepsilon^2, Q(\boldsymbol{\eta}))$ is given by $\log(2Q(\boldsymbol{\eta}_1)) + 1$ and we conclude that plim $\widehat{\boldsymbol{\eta}}_1^{(3)} = \boldsymbol{\eta}_1$. Once again $\boldsymbol{\eta}_1 = \arg\min_{\boldsymbol{\eta}} Q(\boldsymbol{\eta})$ is the pseudo-true parameter for the estimator under mis-specification. ■

A.1.3: Conditional sum of squares estimation

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. Then from (A.1) we can deduce that $\mathbf{A}_\eta = \mathbf{T}_\eta \mathbf{T}_\eta^T + \mathbf{H}_\eta \mathbf{H}_\eta^T$. From (11) and (12) it follows that $Q_n^{(4)}(\boldsymbol{\eta}) = \frac{1}{n}\mathbf{Y}^T\mathbf{T}_\eta\mathbf{T}_\eta^T\mathbf{Y}$, and it is shown below that $\frac{1}{n}\mathbf{Y}^T\mathbf{H}_\eta\mathbf{H}_\eta^T\mathbf{Y} = o_p(1)$. We can therefore conclude that $\left|Q_n^{(4)}(\boldsymbol{\eta}) - \frac{1}{n}\mathbf{Y}^T\mathbf{A}_\eta\mathbf{Y}\right|$ converges to zero in probability, and hence, using (A.2), the limiting value of the criterion function $Q_n^{(4)}(\boldsymbol{\eta})$ is $2Q(\boldsymbol{\eta})$. That the pseudo-true parameter for the CSS estimator under mis-specification is $\boldsymbol{\eta}_1 = \arg\min_{\boldsymbol{\eta}} Q(\boldsymbol{\eta})$ and plim $\widehat{\boldsymbol{\eta}}_1^{(4)} = \boldsymbol{\eta}_1$ is now obvious.

It remains for us to establish that $\frac{1}{n}\mathbf{Y}^T\mathbf{H}_\eta\mathbf{H}_\eta^T\mathbf{Y} = o_p(1)$. Suppressing the dependence on the parameter $\boldsymbol{\eta}$ for notational simplicity, set $\mathbf{M} = \mathbf{H}\mathbf{H}^T$. Then $\mathbf{M} = [m_{ij}]_{i,j=1,\ldots,n}$ where $m_{ij} = \sum_{u=0}^{\infty} \tau_{u+n-i}\tau_{u+n-j}$. Since $|\tau_k| \sim k^{-(1+d)}\mathcal{C}_\tau$, $\mathcal{C}_\tau < \infty$, the series $\sum_{k=0}^{\infty}\tau_k$ is absolutely convergent and square summable; moreover, $\sum_{k=0}^{\infty} k^\alpha|\tau_k|^2 < \infty$ for all $\alpha \in (0, 1+2d)$, from which we can deduce that $|m_{ij}|^2 \sim \{(n-i+1)(n-j+1)\}^{-(1+d)}\mathcal{C}_m$, $\mathcal{C}_m < \infty$, and hence (with $r = n-i+1$ and $s = n-j+1$) that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} |m_{ij}|^2 \sim \sum_{r=1}^{n}\sum_{s=1}^{n}(rs)^{-(1+d)}\mathcal{C}_m < \infty\,. \tag{A.3}$$

By (A.3) we have $\|\mathbf{M}\|^2 = O(n^{-(1+d)})$ and Theorem 5.1 of Dahlhaus (1989) implies that $\|\boldsymbol{\Sigma}_0\| = O(n^{1/2})$. It follows that $|E_0[\mathbf{Y}^T\mathbf{M}\mathbf{Y}]| = |\mathrm{tr}\mathbf{M}\boldsymbol{\Sigma}_0| = O(n^{-d//2})$ and $Var_0[\mathbf{Y}^T\mathbf{M}\mathbf{Y}] = \mathrm{tr}(\mathbf{M}\boldsymbol{\Sigma}_0)^2 = O(n^{-d})$. We therefore have that

$$Pr\left(n^{-1}\left|\mathbf{Y}^T\mathbf{M}\mathbf{Y}\right| > \epsilon\right) = O(n^{-(2+d)})$$

for all $\epsilon > 0$ by Markov's inequality. Since ϵ is arbitrary it follows that $n^{-1}\left|\mathbf{Y}^T\mathbf{M}\mathbf{Y}\right| \to 0$ almost surely, by the Borell-Cantelli lemma, giving the desired result. ■

A.2: Proof of Theorem 2:

First note that

$$Q_N(\boldsymbol{\psi}) = \left\{\pi\frac{\sigma_{\varepsilon 0}^2}{\sigma_\varepsilon^2}\frac{\Gamma(1-2(d_0-d))}{\Gamma^2(1-(d_0-d))}\right\}K_N(\boldsymbol{\eta}) \tag{A.4}$$

by the same argument that gives (17). Now let $\Delta C_N(z) = \sum_{j=N+1}^{\infty} c_j z^j = C(z) - C_N(z)$. Then

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

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

$$R_{1N} = \left(\frac{\sigma_{\varepsilon 0}^2}{\sigma_\varepsilon^2}\right)\int_0^\pi |\Delta C_N(e^{i\lambda})|^2|2\sin(\lambda/2)|^{-2(d_0-d)}d\lambda \tag{A.5}$$

and

$$R_{2N} = \left(\frac{\sigma_{\varepsilon 0}^2}{\sigma_\varepsilon^2}\right)\int_{-\pi}^\pi \Delta C_N(e^{i\lambda})C_N(e^{-i\lambda})|2\sin(\lambda/2)|^{-2(d_0-d)}d\lambda\,. \tag{A.6}$$

The first integral in (A.5) equals

$$\left\{\pi\frac{\sigma_{\varepsilon 0}^2}{\sigma_\varepsilon^2}\frac{\Gamma(1-2(d_0-d))}{\Gamma^2(1-(d_0-d))}\right\}\left(\sum_{j=N+1}^\infty c_j^2 + 2\sum_{k=N+1}^\infty\sum_{j=k+1}^\infty c_jc_k\rho(j-k)\right)\,.$$

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 $0 < d, d_0 < 0.5$ it follows that $|d_0 - d| < 0.5$ and Sterling's approximation can therefore be used to show that $|\rho(h)| < \mathcal{C}'h^{2(d_0-d)-1}$, $h = 1, 2, \ldots$, for some $\mathcal{C}' < \infty$. This implies that

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

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 (A.6) enables us to bound $|R_{2N}|$ by $2(\sigma_{\varepsilon 0}/\sigma_\varepsilon)\sqrt{I_N \cdot R_{1N}}$. 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 $Q(\boldsymbol{\psi})$ with respect to $\boldsymbol{\eta}$ is

$$\frac{\partial Q(\boldsymbol{\psi})}{\partial\boldsymbol{\eta}} = \left(\frac{\sigma_{\varepsilon 0}^2}{\sigma_\varepsilon^2}\right)\int_{-\pi}^\pi \frac{C(e^{i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}}\frac{\partial}{\partial\boldsymbol{\eta}}\{\frac{C(e^{-i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}}\}d\lambda$$

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

$$R_{3N} = \left(\frac{\sigma_{\varepsilon 0}^2}{\sigma_\varepsilon^2}\right)\int_{-\pi}^\pi \frac{C_N(e^{i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}}\frac{\partial}{\partial\eta_j}\{\frac{\Delta C_N(e^{-i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}}\}d\lambda$$

and

$$R_{4N} = \left(\frac{\sigma_{\varepsilon 0}^2}{\sigma_\varepsilon^2}\right)\int_{-\pi}^\pi \frac{\Delta C_N(e^{i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}}\frac{\partial}{\partial\eta_j}\{\frac{C(e^{-i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}}\}d\lambda\,.$$

The Cauchy-Schwarz inequality now yields the inequalities

$$|R_{3N}|^2 \leq R_{1N}\left(\frac{\sigma_{\varepsilon 0}^2}{\sigma_\varepsilon^2}\right)\int_{-\pi}^\pi \frac{|C_N(e^{i\lambda})|^2}{|2\sin(\lambda/2)|^{2(d_0-d)}}\left|\frac{\partial}{\partial\eta_j}\{\log\frac{\Delta C_N(e^{-i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}}\}\right|^2 d\lambda$$

and

$$|R_{4N}|^2 \leq R_{1N}\left(\frac{\sigma_{\varepsilon 0}^2}{\sigma_\varepsilon^2}\right)\int_{-\pi}^\pi \left|\frac{\partial}{\partial\eta_j}\{\frac{C(e^{-i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}}\}\right|^2 d\lambda\,,$$

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

A.3: Proof of Theorem 3:

To establish (26) we will first show that for the Whittle estimator we have $\frac{\sigma_\varepsilon^2}{4}\partial Q_n^{(2)}(\boldsymbol{\eta})/\partial\boldsymbol{\eta} = \partial Q_n^{(1)}(\boldsymbol{\eta})/\partial\boldsymbol{\eta} + o_p(n^{-1/2})$. For the TML and CSS estimators we will then show that $2R_n\partial Q_n^{(3)}(\boldsymbol{\eta}_1)/\partial\boldsymbol{\eta}$ and $R_n\partial Q_n^{(2)}(\boldsymbol{\eta}_1)/\partial\boldsymbol{\eta}$ converge in distribution, and that $n^{1/2}\partial\{Q_n^{(4)}(\boldsymbol{\eta}) - Q_n^{(2)}(\boldsymbol{\eta})\}/\partial\boldsymbol{\eta} = o_p(1)$, respectively.

For the Whittle estimator we have

$$\frac{\partial\{\frac{\sigma_\varepsilon^2}{4}Q_n^{(2)}(\boldsymbol{\eta}) - Q_n^{(1)}(\boldsymbol{\eta})\}}{\partial\boldsymbol{\eta}} = \frac{\sigma_\varepsilon^2}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{\partial\log\left[f_1(\boldsymbol{\eta},\lambda_j)\right]}{\partial\boldsymbol{\eta}},$$

a deterministic function of $\boldsymbol{\eta}$. Following the development in Chen and Deo's proof of their Lemma 4 (see Chen and Deo, 2006, p. 270) gives

$$\frac{\sigma_\varepsilon^2}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{\partial\log\left[f_1(\boldsymbol{\eta},\lambda_j)\right]}{\partial\boldsymbol{\eta}} = O(n^{-1}\log^2 n) = o(n^{-1/2})$$

and $n^{1/2}\partial\{\frac{\sigma_\varepsilon^2}{4}Q_n^{(2)}(\boldsymbol{\eta}) - Q_n^{(1)}(\boldsymbol{\eta})\}/\partial\boldsymbol{\eta} = o(1)$ almost surely. The asymptotic equivalence of the two gradients now follows; in Case 1 because $n^{1-2d^*}/n^{1/2}\log n \to 0$ as $n\to\infty$ when $d^* > 0.25$, in Case 2 because $n^{1/2}[\overline{\Lambda}_{dd}]^{-1/2} \propto \left(n/\log^3 n\right)^{1/2}$ when $d^* = 0.25$ by Lemma 10 of Chen and Deo (2006) and, trivially, $1/\log^{3/2} n \to 0$ as $n\to\infty$, and directly in Case 3 when $d^* < 0.25$.

For $R_n\partial\{2Q_n^{(3)}(\boldsymbol{\eta}_1) - Q_n^{(2)}(\boldsymbol{\eta}_1)\}/\partial\boldsymbol{\eta}$ we begin by noting that by Theorem 5.1 of Dahlhaus (1989), and definition of the Riemann-Stieltjes integral,

$$\frac{1}{n}\frac{\partial\log|\boldsymbol{\Sigma}_\eta|}{\partial\boldsymbol{\eta}} = \frac{1}{n}\mathrm{tr}\boldsymbol{\Sigma}_\eta^{-1}\frac{\partial\boldsymbol{\Sigma}_\eta}{\partial\boldsymbol{\eta}} \sim \frac{2}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{\partial\log\left[f_1(\boldsymbol{\eta},\lambda_j)\right]}{\partial\boldsymbol{\eta}}.$$

Our task therefore reduces to a consideration of the properties of

$$\frac{1}{n}\frac{\partial\mathbf{Y}^T\boldsymbol{\Sigma}_\eta^{-1}\mathbf{Y}}{\partial\boldsymbol{\eta}} - \frac{2}{n}\sum_{j=1}^{\lfloor n/2\rfloor}I(\lambda_j)\frac{\partial f_1(\boldsymbol{\eta},\lambda_j)^{-1}}{\partial\boldsymbol{\eta}},$$

which we rewrite as $\mathbf{a}-\mathbf{b}$ where

$$\mathbf{a} = \frac{1}{n}\frac{\partial\mathbf{Y}^T\boldsymbol{\Sigma}_\eta^{-1}\mathbf{Y}}{\partial\boldsymbol{\eta}} - \frac{1}{n}\mathrm{tr}\frac{\partial\boldsymbol{\Sigma}_\eta^{-1}}{\partial\boldsymbol{\eta}}\boldsymbol{\Sigma}_0$$

and

$$\mathbf{b} = \frac{2}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\left(\frac{I(\lambda_j)}{f_0(\lambda_j)} - 1\right)f_0(\lambda_j)\frac{\partial f_1(\boldsymbol{\eta},\lambda_j)^{-1}}{\partial\boldsymbol{\eta}}$$

recognizing, via Theorem 5.1 of Dahlhaus (1989) once again, that

$$\begin{aligned}E_0\left(\frac{1}{n}\frac{\partial\mathbf{Y}^T\boldsymbol{\Sigma}_\eta^{-1}\mathbf{Y}}{\partial\boldsymbol{\eta}}\right) =& \frac{1}{n}\mathrm{tr}\frac{\partial\boldsymbol{\Sigma}_\eta^{-1}}{\partial\boldsymbol{\eta}}\boldsymbol{\Sigma}_0\\ =& -\frac{1}{n}\mathrm{tr}\boldsymbol{\Sigma}_\eta^{-1}\frac{\partial\boldsymbol{\Sigma}_\eta}{\partial\boldsymbol{\eta}}\boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Sigma}_0\\ \sim& \frac{2}{n}\sum_{j=1}^{\lfloor n/2\rfloor}f_0(\lambda_j)\frac{\partial f_1(\boldsymbol{\eta},\lambda_j)^{-1}}{\partial\boldsymbol{\eta}}.\end{aligned}$$

Using expression (A.7) below we can therefore deduce that

$$E_0(\mathbf{a}-\mathbf{b}) = \begin{cases} O(n^{2d^*-1}\log n), & 0 < d^* < 0.5\,; \\ O(n^{-1}\log^3 n), & 0.5 < d^* \leq 0\,. \end{cases}$$

From the binomial expansion of $(a-b)^r$, $r \geq 2$, it follows that the higher order cumulants will converge to zero if the corresponding cumulants of $a = \boldsymbol{\lambda}^T\mathbf{a}$ and $b = \boldsymbol{\lambda}^T\mathbf{b}$ are asymptotically equal (modulo a constant multiple) for every fixed vector $\boldsymbol{\lambda} \neq \mathbf{0}$. The desired result then follows, implicitly invoking the Cramér-Wold device, since the cumulants are convergence determining for the limiting distributions in Theorem 3.

We will show that a and b asymptotically share the same cumulants in the special case where $\boldsymbol{\lambda}^T = (1, 0, \ldots, 0)$. This corresponds to considering the asymptotic distribution of the estimate of d and demonstrates the detailed particulars required to deal with the two critical cases involving convergence rates less than $n^{1/2}$. Denoting the rth cumulant of a by $\kappa_0^r(a)$, we obtain for $r \geq 2$

$$\begin{aligned}\kappa_0^r(a) =& n^{-r}(r-1)!2^{r-1}\mathrm{tr}\left\{\frac{\partial\boldsymbol{\Sigma}_\eta^{-1}}{\partial d}\boldsymbol{\Sigma}_0\right\}^r \\ \sim& n^{-(r-1)}(r-1)!2^{r-1}\frac{2}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\left(\frac{f_0(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)}\right)^r\left(\frac{\partial\{\log f_1(\boldsymbol{\eta},\lambda_j)\}}{\partial d}\right)^r,\end{aligned}$$

using Theorem 5.1 of Dahlhaus (1989) once more. For b, let

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

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

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

Then $\mathbf{X}^T$ is Gaussian with zero mean and covariance $\mathbf{I}+\boldsymbol{\Delta}$ where $\triangle_{jk} = O(j^{-d_0}k^{d_0-1}\log k)$ for $1 \leq j \leq k \leq \lfloor n/2\rfloor$, (see Moulines and Soulier ,1999, Lemma 4). Moreover,

$$\frac{2}{n}\sum_{j=1}^{\lfloor n/2\rfloor} I(\lambda_j)\frac{\partial f_1(\boldsymbol{\eta},\lambda_j)^{-1}}{\partial d} = \frac{2}{n}\mathbf{X}^T\mathbf{F}_0\mathbf{D}_1\mathbf{X}$$

where $\mathbf{D}_1 = \partial\mathbf{F}_1^{-1}/\partial d$,

$$\mathbf{F}_1 = \mathrm{diag}(f_1(\boldsymbol{\eta},\lambda_1), f_1(\boldsymbol{\eta},\lambda_1)\ldots, f_1(\boldsymbol{\eta},\lambda_{\lfloor n/2\rfloor}), f_1(\boldsymbol{\eta},\lambda_{\lfloor n/2\rfloor}))\,,$$

from which it follows that

$$\begin{aligned}\frac{2}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\left(\frac{E_0(I(\lambda_j))}{f_0(\lambda_j)}-1\right)f_0(\lambda_j)\frac{\partial f_1(\boldsymbol{\eta},\lambda_j)^{-1}}{\partial d} &= \frac{2}{n}\mathrm{tr}\mathbf{F}_0\mathbf{D}_1\boldsymbol{\Delta} \qquad \text{(A.7)} \\ &= O\left(n^{2d^*-1}\sum_{j=1}^{\lfloor n/2\rfloor} j^{-(1+2d^*)}\log^2 j\right) \\ &= \begin{cases} O(n^{2d^*-1}\log n), & 0 < d^* < 0.5\,; \\ O(n^{-1}\log^3 n), & -0.5 < d^* \leq 0\,, \end{cases}\end{aligned}$$

since

$$\frac{\partial f_1(\boldsymbol{\eta},\lambda)^{-1}}{\partial d} = -2\frac{\log 2|\sin\lambda/2|}{f_1(\boldsymbol{\eta},\lambda)} = O(\lambda^{2d_1}\log\lambda)\,,$$

cf. Lemma 4 of Chen and Deo (2006). For $r \geq 2$ we have

$$\kappa_0^r(b) = 2^r n^{-r}(r-1)!2^{r-1}\mathrm{tr}\left\{\mathbf{F}_0\mathbf{D}_1(\mathbf{I}+\boldsymbol{\Delta})\right\}^r$$

and the expansion $\mathrm{tr}\left\{\mathbf{F}_0\mathbf{D}_1(\mathbf{I}+\boldsymbol{\Delta})\right\}^r = \sum_{j=0}^r \binom{r}{j}\mathrm{tr}\{\mathbf{F}_0\mathbf{D}_1\}^{r-j}\{\mathbf{F}_0\mathbf{D}_1\boldsymbol{\Delta}\}^j$ yields the result that

$$\begin{aligned}\mathrm{tr}\left\{\mathbf{F}_0\mathbf{D}_1(\mathbf{I}+\boldsymbol{\Delta})\right\}^r = &\mathrm{tr}\left\{\mathbf{F}_0\mathbf{D}_1\right\}^r + \mathrm{tr}\left\{\mathbf{F}_0\mathbf{D}_1\right\}^r \boldsymbol{\Delta} \\ &+ O\left(\sum_{j=2}^r \binom{r}{j}\mathrm{tr}\{\mathbf{F}_0\mathbf{D}_1\}^{r-j}\{\mathbf{F}_0\mathbf{D}_1\boldsymbol{\Delta}\}^j\right). \end{aligned} \tag{A.8}$$

Evaluating the terms on the right hand side of (A.8) gives

$$\mathrm{tr}\left\{\mathbf{F}_0\mathbf{D}_1\right\}^r = 2\sum_{j=1}^{\lfloor n/2\rfloor}\left(\frac{f_0(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)}\right)^r\left(\frac{\partial\{\log f_1(\boldsymbol{\eta},\lambda_j)\}}{\partial d}\right)^r,$$

$$\begin{aligned}\mathrm{tr}\left\{\mathbf{F}_0\mathbf{D}_1\right\}^r \boldsymbol{\Delta} =&O\left(n^{2rd^*}\sum_{j=1}^{\lfloor n/2\rfloor} j^{-(1+2rd^*)}\log^{(r+1)} j\right)\\ =&\begin{cases} O(n^{2rd^*}\log n), & 0 < d^* < 0.5\,;\\ O(\log^{(r+2)} n), & -0.5 < d^* \leq 0\,,\end{cases}\end{aligned}$$

and, similarly

$$\mathrm{tr}\{\mathbf{F}_0\mathbf{D}_1\}^{r-j}\{\mathbf{F}_0\mathbf{D}_1\boldsymbol{\Delta}\}^j = \begin{cases} O(n^{2rd^*}\log^{(j-1+2d_0)} n), & 0 < d^* < 0.5\,;\\ O(\log^{(r+2(1+d_0))} n), & -0.5 < d^* \leq 0\,,\end{cases}$$

for $j = 2, \ldots, r$. It follows that

$$\frac{\kappa_0^r(2a)-\kappa_0^r(b)}{(r-1)!2^{r-1}} = \begin{cases} O(n^{r(2d^*-1)}\log n) + \sum_{j=2}^r\binom{r}{j}O(n^{r(2d^*-1)}\log^{(j-1+2d_0)} n), & 0 < d^* < 0.5\,;\\ O(n^{-r}\log^{(r+2)} n) + O(n^{-r}\log^{(r+2(1+d_0))} n), & -0.5 < d^* \leq 0\,,\end{cases}$$

leading to the desired result, namely that $R_n\partial\{2Q_n^{(3)}(\boldsymbol{\eta}_1) - Q_n^{(2)}(\boldsymbol{\eta}_1)\}/\partial d \to^D 0$ where $R_n = n^{1-2d^*}/\log n$ when $d^* > 0.25$, Case 1, $R_n = (n/\log^3 n)^{1/2}$ when $d^* = 0.25$, Case 2, and $R_n = n^{1/2}$ when $d^* < 0.25$, Case 3.

The corresponding results for arbitrary $\boldsymbol{\lambda} \neq \mathbf{0}$ can be obtained by reexpressing the linear combinations as $a = \boldsymbol{\lambda}^T\mathbf{a} = \partial Q_n^{(a)} - E_0(\partial Q_n^{(a)})$ and $b = \boldsymbol{\lambda}^T\mathbf{b} = \partial Q_n^{(b)} - E_0(\partial Q_n^{(b)})$ where the quadratic forms are given by

$$\partial Q_n^{(a)} = \frac{1}{n}(\mathbf{Y}^T\otimes(1,\ldots,1))\left[\langle\frac{\partial}{\partial\boldsymbol{\eta}}\{\boldsymbol{\Sigma}_\eta^{-1}\}\rangle\otimes\langle\boldsymbol{\lambda}\rangle\right](\mathbf{Y}\otimes(1,\ldots,1)^T)\,,$$

where

$$\left[\langle\frac{\partial}{\partial\boldsymbol{\eta}}\{\boldsymbol{\Sigma}_\eta^{-1}\}\rangle\otimes\langle\boldsymbol{\lambda}\rangle\right] = \mathrm{diag}(\frac{\partial}{\partial\eta_1}\{\boldsymbol{\Sigma}_\eta^{-1}\},\ldots,\frac{\partial}{\partial\eta_{l+1}}\{\boldsymbol{\Sigma}_\eta^{-1}\})\otimes\mathrm{diag}(\lambda_1,\ldots,\lambda_{l+1})\,,$$

and

$$\partial Q_n^{(b)} = \frac{1}{n}(\mathbf{X}^T\otimes(1,\ldots,1))\left[\langle\frac{\partial}{\partial\boldsymbol{\eta}}\{\mathbf{F}_0\mathbf{F}_1^{-1}\}\rangle\otimes\langle\boldsymbol{\lambda}\rangle\right](\mathbf{X}\otimes(1,\ldots,1)^T)\,.$$

The cumulants of a and b of order $r \geq 2$ can then be evaluated in the same manner as described above for the special case $\boldsymbol{\lambda}^T = (1,0,\ldots,0)$, the remaining details involve only more complex notational and bookkeeping conventions.

For the difference between $\partial Q_n^{(4)}(\boldsymbol{\eta})/\partial\boldsymbol{\eta}$ and $\partial Q_n^{(2)}(\boldsymbol{\eta})/\partial\boldsymbol{\eta}$ we have

$$\frac{\partial\{Q_n^{(4)}(\boldsymbol{\eta}_1)\}}{\partial\boldsymbol{\eta}} = \frac{\partial}{\partial\boldsymbol{\eta}}\{\frac{\mathbf{Y}^T\mathbf{A}_\eta\mathbf{Y}}{n}\} - \frac{\partial}{\partial\boldsymbol{\eta}}\{\frac{\mathbf{Y}^T\mathbf{M}_\eta\mathbf{Y}}{n}\},$$

and by Lemma A.1 it follows that

$$\frac{\partial}{\partial\boldsymbol{\eta}}\{\frac{\mathbf{Y}^T\mathbf{A}_\eta\mathbf{Y}}{n}\} - \frac{2}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)}\frac{\partial\{\log f_1(\boldsymbol{\eta},\lambda_j)\}}{\partial\boldsymbol{\eta}} = o_p(n^{-1/2}).$$

Now let $\dot{\eta} = (\eta_1,\ldots,\dot{\eta}_j,\ldots,\eta_{l+1})^T$ and set

$$\nabla\mathbf{M}(\dot{\eta}_j) = \begin{cases} \frac{\mathbf{M}_{\dot{\eta}}-\mathbf{M}_\eta}{\dot{\eta}_j-\eta_j}, & \dot{\eta}_j \neq \eta_j\,; \\ \frac{\partial\{\mathbf{M}_\eta\}}{\partial\eta_j}, & \dot{\eta}_j = \eta_j\,. \end{cases}$$

Then $\lim_{\dot{\eta}_j\to\eta_j}\nabla\mathbf{M}(\dot{\eta}_j) = \nabla\mathbf{M}(\eta_j)$ and for all $\dot{\eta}_j \neq \eta_j$ we can employ an argument that parallels that following (A.3) to deduce that

$$Pr\left(n^{-1/2}\left|\mathbf{Y}^T\nabla\mathbf{M}(\dot{\eta}_j)\mathbf{Y}\right| > \epsilon\right) = O(n^{-(3+2d)/2})$$

for all $\epsilon > 0$, and hence that

$$n^{-1/2}\frac{\partial\{\mathbf{Y}^T\mathbf{M}_\eta\mathbf{Y}\}}{\partial\eta_j} = \lim_{\dot{\eta}_j\to\eta_j}\frac{\mathbf{Y}^T\nabla\mathbf{M}(\dot{\eta}_j)\mathbf{Y}}{n^{1/2}} = o_p(1).$$

This establishes that $n^{1/2}\partial\{Q_n^{(4)}(\boldsymbol{\eta}) - Q_n^{(2)}(\boldsymbol{\eta})\}/\partial\boldsymbol{\eta} = o_p(1)$, and the asymptotic equivalence stated in (26) now follows, because $n^{1-2d^*}/n^{1/2}\log n \to 0$ as $n\to\infty$ in Case 1, in Case 2 because $1/\log^{3/2} n \to 0$ as $n\to\infty$, and directly in Case 3.

The preceding derivations, in conjunction with (25), imply that for the Whittle estimator $R_n(\widehat{\boldsymbol{\eta}}_1^{(2)} - \widehat{\boldsymbol{\eta}}_1^{(1)}) \to^D 0$, and that for the TML and CSS estimators $R_n(\widehat{\boldsymbol{\eta}}_1^{(i)} - \widehat{\boldsymbol{\eta}}_1^{(2)}) \to^D 0$ for $i = 3$ and 4, for all three values of R_n. The asymptotic equivalence of all four estimators now follows since an immediate corollary is that $R_n(\widehat{\boldsymbol{\eta}}_1^{(i)} - \widehat{\boldsymbol{\eta}}_1^{(j)}) \to^D 0$, $i,j = 1,2,3$ and 4, for all three values of R_n. ■

Appendix B: 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

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

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

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

which has expectation

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

where $\boldsymbol{\Sigma}_0 = [\gamma_0(|i-j|)]$ and $\sigma_\varepsilon^2\boldsymbol{\Sigma}_\eta = [\gamma_1(|i-j|)]$, $i, j = 1, 2, ..., n$. The criterion function for the CSS estimator in (11) 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 (13). 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

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