Article:

Martin, G.M., Nadarajah, K. and Poskitt, D.S. (2020) Issues in the estimation of mis-specified models of fractionally integrated processes. Journal of Econometrics, 215 (2). pp. 559-573. ISSN 0304-4076

https://doi.org/10.1016/j.jeconom.2019.09.007

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

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

May 25, 2019

Abstract

This short paper provides a comprehensive set of new theoretical results on the impact of mis-specifying the short run dynamics in fractionally integrated processes. We show that four alternative parametric estimators – frequency domain maximum likelihood, Whittle, time domain maximum likelihood and conditional sum of squares – converge to the same pseudo-true value under common mis-specification, and that they possess a common asymptotic distribution. The results are derived assuming the true data generating mechanism is a fractional linear process driven by a martingale difference innovation. A completely general parametric specification for the short run dynamics of the estimated (mis-specified) fractional model is considered, and with long memory, short memory and antipersistence in both the model and the data generating mechanism accommodated. The paper can be seen as extending an existing line of research on mis-specification in fractional models, important contributions to which have appeared in the *Journal of Econometrics*. It also complements a range of existing asymptotic results on estimation in *correctly specified* fractional models. Open problems in the area are the subject of the final discussion.

Keywords: Long memory models, ARFIMA, pseudo-true parameter, frequency domain estimators, time domain estimators, mis-specified short memory dynamics.

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

1 Introduction

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

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

where $0 \leq |d_0| < 0.5$ and $L_0(\lambda)$ is a positive function that is slowly varying at 0. Prototypical examples of processes of this type are fractional noise, obtained as the increments of self-similar processes, and fractional autoregressive moving average processes. The process $\{y_t\}$ is said to exhibit long memory (or long-range dependence) when $0 < d_0 < 0.5$, short memory (or short-range dependence) when $d_0 = 0$, and antipersistence when $-0.5 < d_0 < 0$, and in this paper we undertake an extensive examination of the consequences for estimation of such processes of mis-specifying the short run dynamics. In so doing we provide a significant extension of earlier work on this particular form of mis-specification in Yajima (1992) and Chen and Deo (2006), as well as

[†]This research has been supported by Australian Research Council (ARC) Discovery Grants DP150101728 and DP170100729. The authors would like to thank the Editor, the Associate Editor and three anonymous referees for very helpful and constructive comments on an earlier draft of the paper.

[‡]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.

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

We begin by showing that four alternative parametric techniques – frequency domain maximum likelihood (FML), Whittle, time domain maximum likelihood (TML) and CSS – converge to a common pseudo-true parameter value when the short memory component is mis-specified.[1] Convergence is established for all three forms of dependence in the true data generating process (TDGP) - long memory, short memory and antipersistence. We establish convergence by demonstrating that when the mis-specified model is evaluated at points in the parameter space where the fractional index d exceeds $d_0 - 0.5$ the FML criterion function has a deterministic limit, but that the FML criterion function is *divergent* otherwise. The difference in the behaviour of the FML criterion function on subsets of the parameter space implies that the objective function does not behave uniformly. (See Robinson (1995), Hualde and Robinson (2011) and Cavaliere, Nielsen and Taylor (2017) for related discussion.) This lack of uniformity makes proofs of convergence across the whole parameter space more complex than usual, but solutions presented in the previously cited references can be tailored to the current situation. We then show that under common mis-specification the criterion functions that define all three alternative estimators behave in a manner similar to that of the FML criterion. All four estimators are, accordingly, shown to converge to the same pseudo-true parameter value – by definition the common value that optimizes all four limiting objective functions.

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

Thirdly, we extend the asymptotic theory established by Chen and Deo (2006) for the FML estimator in the long memory Gaussian process case to the other three estimators, under long memory, short memory and antipersistence for both the TDGP and the estimated model, and without the imposition of Gaussianity. As noted above, Chen and Deo (2006) derived their results by assuming that $\{y_t\}$ is a Gaussian process, thereby implying that finite sample Fourier sine and cosine transformations are normally distributed, and hence that various properties established in Moulines and Soulier (1999) could be employed. Here we use the same properties established in Moulines and Soulier (1999) by appealing to results of Lahiri (2003) showing that the Fourier transformations are asymptotically normal. We establish that all four estimation methods are asymptotically equivalent in that they converge in distribution under common mis-specification. The convergence rate and nature of the asymptotic distribution is determined by the deviation of the pseudo-true value of the fractional index, d_1 say, from the true value, d_0, with three critical ranges for $d^* = d_0 - d_1 < 0.5$ given by $d^* > 0.25$, $d^* = 0.25$ and $d^* < 0.25$. This nonstandard distributional behaviour for all four parametric estimators introduces a further degree of

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

complexity into the analysis, and contrasts sharply with earlier results established in correctly specified models, where separate modeling of the short-run and fractional dynamics results in the asymptotic distribution of the parameter estimates being normal and free of the fractional indices.

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 Whittle, TML and CSS estimators, as well as the FML estimator, are specified, and we demonstrate that all four estimators possess a common probability limit under mis-specification. The limiting form of the criterion function for a mis-specified ARFIMA model is presented in Section 3, 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 paper concludes in Section 5 with a brief summary and some discussion of several issues that arise from the work. In order to streamline the presentation, four lemmas used in the proof of Theorem 1 are placed in an appendix to the paper. The proofs of these lemmas and all other results presented in the paper are, in turn, assembled in a Supplementary Appendix, available on the journal website. The Supplementary Appendix also contains certain technical derivations referenced in the text. Results of an extensive set of simulation experiments, documenting the relative finite sample performance of all four mis-specified estimators, under varying degrees of mis-specification, are available from the authors on request.

2 Estimation Under Mis-specification of the Short Run Dynamics

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

$$\frac{\sigma_0^2}{2\pi} f_0(\lambda) = \frac{\sigma_0^2}{2\pi} g_0(\lambda) \left(2\sin(\lambda/2)\right)^{-2d_0}, \tag{1}$$

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

$$y_t = \mu_0 + \sum_{j=0}^{\infty} b_{0j}\varepsilon_{t-j}, \quad t \in \mathbb{Z} = 0, \pm 1, \ldots, \tag{2}$$

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

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

The conditions imposed on $c_0(z)$ imply that $g_0(\lambda)$ corresponds to the spectrum of an invertible short-memory process that is bounded and bounded away from zero for all $\lambda \in [-\pi, \pi]$ and the TDGP satisfies *Conditions A* of Hannan (1973, page 131). Assumption A.1 was introduced into time series analysis by Hannan (1973) and has been employed by several authors in investigations of both short memory and fractional linear processes since. The assumption that $\{\varepsilon_t\}$ is a conditionally homoscedastic martingale difference process circumvents the need to assume independence or identical distributions for the innovations, but rules out heteroscedasticity (see Cavaliere, Nielsen and Taylor, 2017, pages 5-6).[2]

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

$$\frac{\sigma^2}{2\pi} f_1(\boldsymbol{\eta}, \lambda) = \frac{\sigma^2}{2\pi} g_1(\boldsymbol{\beta}, \lambda)(2\sin(\lambda/2))^{-2d}, \tag{3}$$

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

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

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

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

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

$$\phi(L)(1-L)^d\{y_t - \mu\} = \theta(L)\varepsilon_t, \tag{4}$$

where $\mu = E(y_t)$, L is the lag operator such that $L^k y_t = y_{t-k}$, and $\phi(z) = 1 + \phi_1 z + ... + \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + ... + \theta_q z^q$ are the autoregressive and moving average operators respectively, where it is assumed that $\phi(z)$ and $\theta(z)$ have no common roots and that the roots lie outside the unit circle. The errors $\{\varepsilon_t\}$ are assumed to be a white noise sequence with finite variance $\sigma_\varepsilon^2 > 0$. 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 (4) is correct, $\{y_t\}$ is defined as the limit in mean square of a covariance-stationary process. When $0 < d < 0.5$ neither the moving average coefficients nor the autocovariances of the process are absolutely summable, declining at a hyperbolic rate rather than the exponential rate typical of an ARMA process, with the term 'long memory' invoked accordingly. Thus, for an ARFIMA model we have $g_1(\boldsymbol{\beta}, \lambda) = |\theta(e^{i\lambda})|^2/|\phi(e^{i\lambda})|^2$ where $\boldsymbol{\beta} = (\phi_1, \phi_2, ..., \phi_p, \theta_1, \theta_2, ..., \theta_q)^T$ and an $ARFIMA(p, d, q)$

[2]Note, however, that the assumption of long memory has become a standard one to adopt in the *modelling* of the variance, or volatility of financial returns. In particular, the ARFIMA models that underpin our results have been used in the direct modelling of long memory in observable measures of volatility like, for instance, the logarithm of realized variance (e.g. Andersen, Bollerslev, Diebold and Labys, 2003; Pong, Shackleton, Taylor and Wu, 2004; Koopman, Jungbacker and Hol, 2005; Martin, Reidy and Wright, 2009). The theoretical results in the paper would apply to such cases. The applicability of the results to settings in which volatility is modelled as a latent long memory process (e.g. Baillie, Bollerslev and Mikkelsen, 1996; Breidt, Crato and de Lima, 1998; Comte and Renault, 1998; Deo and Hurvich, 2001; Hurvich and Ray, 2003; Hurvich, Moulines and Soulier, 2005) would need further work to confirm.

model will be mis-specified if the realizations are generated from a true $ARFIMA(p_0, d_0, q_0)$ process and any of $\{p \neq p_0 \cup q \neq q_0\} \setminus \{p_0 \leq p \cap q_0 \leq q\}$ obtain.

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

$$\mathbb{E}_\delta = \mathbb{D}_\delta \times \mathbb{B} \quad \text{where} \quad \mathbb{D}_\delta = \{d : |d| \ \leq 0.5 - \delta\}\,, \text{ for some } 0 < \delta \ll 0.5\,. \tag{5}$$

The bound on $|d|$ must be set by the practitioner via some criterion that reflects numerical precision. Under mis-specification the generic estimator, denoted by $\widehat{\boldsymbol{\eta}}_1$ for the time being, is obtained by minimizing $Q_n(\boldsymbol{\eta})$ assuming that $\{y_t\}$ follows the MisM.[3] In Section 2.1 we specify the form of $Q_n(\boldsymbol{\eta})$ associated with the FML estimator considered in Chen and Deo (2006) and outline its relationship with the criterion functions underlying two alternative versions of the frequency domain estimator introduced by Whittle, making it clear which form of Whittle estimator is the focus of our theoretical investigations. In Section 2.2 we define the two time domain estimators that we consider here, TML and CSS, and their associated criterion functions.

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

$$Q(\boldsymbol{\eta}) = \lim_{n\to\infty} E_0\left[Q_n(\boldsymbol{\eta})\right] = \frac{\sigma_0^2}{2\pi}\int_0^\pi \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta}, \lambda)} d\lambda\,. \tag{6}$$

If, on the other hand, $Q_n(\boldsymbol{\eta})$ is evaluated on a subset of $\mathbb{E}_\delta$ where $\mathbb{D}_\delta$ intersects with $\{d : (d_0 - d) \geq 0.5\}$, then the criterion function is divergent. The latter corresponds to the integral on the right hand side in (6) being assigned the value ∞ if $(d_0 - d) \geq 0.5$ (see the comment by Hannan on his Lemma 2 in Hannan, 1973, page 134). This difference in behaviour of the criterion function about the point $d_0 - d = 0.5$ implies that $Q_n(\boldsymbol{\eta})$ does not converge uniformly on subsets of the parameter space that include this point. Nevertheless, as will be demonstrated below, provided that $\boldsymbol{\eta}_1 \in \mathbb{E}_\delta$, where $\boldsymbol{\eta}_1$ is the minimizer of $Q(\boldsymbol{\eta})$, $Q_n(\widehat{\boldsymbol{\eta}}_1)$ will converge to $Q(\boldsymbol{\eta}_1)$ and $\widehat{\boldsymbol{\eta}}_1$ will converge to $\boldsymbol{\eta}_1$ as a consequence.

In Section 2.3 we derive our asymptotic results pertaining to the convergence of $Q_n(\boldsymbol{\eta})$ and demonstrate the relationships between the limiting criterion functions of the Whittle, TML and CSS estimators and the limiting criterion function of the FML estimator. The value that minimizes the limiting criterion function of all four estimators is shown to be identical, and the asymptotic convergence of all four estimators to the common pseudo-true parameter, $\boldsymbol{\eta}_1$, is thereby established.

We highlight the fact that the FML and Whittle estimators are mean invariant by virtue of being defined on the non-zero fundamental Fourier frequencies. The same is not true, however, for either of the two time domain based methods. Hence, in Section 2.1 the parameter μ, which characterizes the assumed model in (4), does not feature in either criterion function, whilst in Section 2.2 it does. In all theoretical derivations we adopt the assumption of a known mean for both the true and estimated model, with a zero value specified without loss of generality. As a consequence of their invariance to the mean, all theoretical results as they pertain to the FML and Whittle estimators also hold for a process that has an arbitrary (non-zero) mean, which may be unknown. Such is not the case for the TML and CSS estimators, and we revisit this point in the Discussion.

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

2.1 Frequency domain estimators

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

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

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. We have labeled this the FML estimator. The objective function in (7) is an approximation to the frequency domain Gaussian (negative) log-likelihood introduced initially by Whittle (1953) for short-range dependent processes, namely

$$W_n(\sigma^2, \boldsymbol{\eta}) = \int_{-\pi}^{\pi} \left\{ \log \frac{\sigma^2}{2\pi} f_1(\boldsymbol{\eta}, \lambda) + \frac{2\pi I(\lambda)}{\sigma^2 f_1(\boldsymbol{\eta}, \lambda)} \right\} d\lambda\,, \tag{8}$$

and it coincides with the frequency domain objective function considered in Hannan (1973). Concentrating out σ^2 in (8) and minimizing the associated profile function with respect to $\boldsymbol{\eta}$ produces what we refer to as the exact Whittle estimator.

An alternative approximation to the Whittle criterion function in (8), considered for example in Beran (1994), is

$$Q_n^{(2)}(\sigma^2, \boldsymbol{\eta}) = \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \log \left[\frac{\sigma^2}{2\pi} f_1(\eta, \lambda_j) \right] + \frac{(2\pi)^2}{\sigma^2 n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{f_1(\boldsymbol{\eta}, \lambda_j)}. \tag{9}$$

Taking $\boldsymbol{\eta}$ as the parameter of interest and concentrating $Q_n^{(2)}(\sigma^2, \boldsymbol{\eta})$ with respect to σ^2 indicates that the value of σ^2 that minimises (9) is given by $\widehat{\sigma}^2(\boldsymbol{\eta}) = 2Q_n^{(1)}(\boldsymbol{\eta})$. Substituting back in to (9) yields the (negative) profile likelihood,

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

Minimization of $Q_n^{(2)}(\boldsymbol{\eta})$ with respect to $\boldsymbol{\eta}$ yields what we call (simply) the Whittle estimator, and which is the form of Whittle procedure that features in our theoretical derivations. Since $\lim_{n\to\infty} \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \log f_1(\boldsymbol{\eta}, \lambda_j) = 0$ (see Supplementary Appendix A) it follows that this estimator is equivalent to the FML estimator for large n. In common with the FML approach, this form of Whittle estimator is invariant to the mean of the process, as noted above.

2.2 Time domain estimators

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

- 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 $\sigma^2 \boldsymbol{\Sigma}_\eta = [\gamma_1 (i - j)]$, $i, j = 1, 2, ..., n$, where

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

  The Gaussian log-likelihood function for the TML estimator is

  $$-\frac{1}{2} \left(n \log(2\pi\sigma^2) + \log |\boldsymbol{\Sigma}_\eta| + \frac{1}{\sigma^2} (\mathbf{Y} - \mu \mathbf{l})^T \boldsymbol{\Sigma}_\eta^{-1} (\mathbf{Y} - \mu \mathbf{l}) \right)\,, \tag{10}$$

  where $\mathbf{l}^T = (1, 1, ..., 1)$, and maximizing (10) is equivalent to minimizing the criterion function

  $$Q_n^{(3)}(\sigma^2, \boldsymbol{\eta}) = \log \sigma^2 + \frac{1}{n} \log |\boldsymbol{\Sigma}_\eta| + \frac{1}{n\sigma^2} (\mathbf{Y} - \mu \mathbf{l})^T \boldsymbol{\Sigma}_\eta^{-1} (\mathbf{Y} - \mu \mathbf{l})\,. \tag{11}$$

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

$$(1-z)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} z^j \,, \tag{12}$$

where $\Gamma(\cdot)$ is the gamma function. Furthermore, since $g_1(\boldsymbol{\beta},\lambda)$ is bounded, by Assumption $(A.3)$, we can employ the method of Whittle (Whittle, 1984, §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

$$Q_n^{(4)}(\boldsymbol{\eta}) = \frac{1}{n}\sum_{t=1}^{n} e_t(\boldsymbol{\eta})^2 \,, \tag{13}$$

where

$$e_t(\boldsymbol{\eta}) = \sum_{i=0}^{t-1} \tau_i(\boldsymbol{\eta})\,(y_{t-i}-\mu)\;, \quad t=1,\ldots,n\,, \tag{14}$$

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

$$\tau_j(\boldsymbol{\eta}) = \sum_{s=0}^{j} \frac{\alpha_{j-s}(\beta)\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}\;, \quad j=1,2,\ldots\,. \tag{15}$$

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

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

2.3 Convergence Properties

In Chen and Deo (2006) it is shown that if $\{y_t\}$ is a long-range dependent Gaussian process, then on subsets of the parameter space of the form $(\delta, 0.5-\delta)\times\Phi$, where $0<\delta<0.25$ and Φ is a compact convex set, we have (for $Q_n^{(1)}(\boldsymbol{\eta})$ defined in (7)) $\text{plim}_{n\to\infty}|Q_n^{(1)}(\boldsymbol{\eta}) - Q(\boldsymbol{\eta})| = 0$ (Chen and Deo, 2006, Lemma 2). The minimum of the limiting objective function $Q(\eta)$ then defines a pseudo-true parameter value to which the FML estimator will converge, since with the addition of the assumption that there exists a unique vector $\boldsymbol{\eta}_1 = (d_1, \boldsymbol{\beta}_1^T)^T \in (\delta, 0.5-\delta)\times\Phi$ that minimizes $Q(\boldsymbol{\eta})$, it follows that the FML estimator will converge to $\boldsymbol{\eta}_1$.

Because Chen and Deo assumed that the TDGP was a long memory process and that in the MisM the fractional index was similarly confined to the long memory region, they did not explicitly consider the case where $(d_0-d)\geq 0.5$. In contrast, as noted with reference to the TDGP in (1), our work allows for $0\leq|d_0|<0.5$, and involves the specification of the appropriate user-assigned compact subset for $\boldsymbol{\eta}=(d,\boldsymbol{\beta}^T)^T$ in (5). This implies a wider range of values for (d_0-d) and, hence, the need for our analysis to deal with the differing behaviour of $Q_n^{(1)}(\boldsymbol{\eta})$ about the point $d_0-d=0.5$ alluded to above. To achieve this, we divide the parameter space $\mathbb{E}_\delta$ into three disjoint sub-sets:

1. $\mathbb{E}_\delta^0 = \mathbb{D}_\delta^0\times\mathbb{B}$ where $\mathbb{D}_\delta^0 = \mathbb{D}_\delta\cap\{d: -(1-2\delta)\leq(d_0-d)\leq 0.5-\delta\}$,
2. $\overline{\mathbb{E}}_{\delta1}^0 = \overline{\mathbb{D}}_{\delta1}^0\times\mathbb{B}$ where $\overline{\mathbb{D}}_{\delta1}^0 = \mathbb{D}_\delta\cap\{d: 0.5-\delta<(d_0-d)<0.5\}$ and
3. $\overline{\mathbb{E}}_{\delta2}^0 = \overline{\mathbb{D}}_{\delta2}^0\times\mathbb{B}$ where $\overline{\mathbb{D}}_{\delta2}^0 = \mathbb{D}_\delta\cap\{d: 0.5\leq(d_0-d)\leq 1-2\delta\}$.

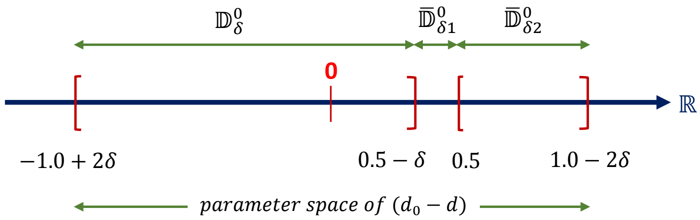


Figure 1: Graphical illustration of the division of the parameter space of $(d_0 - d)$

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

We establish that on the subset $\mathbb{E}^0_\delta$ we have $\lim_{n\to\infty} Q^{(1)}_n(\boldsymbol{\eta}) = Q(\boldsymbol{\eta})$ almost surely and uniformly in $\boldsymbol{\eta}$, where $Q(\boldsymbol{\eta})$ is defined as in (6), whereas $Q^{(1)}_n(\eta)$ is of order $O(\delta^{-1})$ on $\overline{\mathbb{E}}^0_{\delta 1}$ and is divergent as $n \to \infty$ on $\overline{\mathbb{E}}^0_{\delta 2}$. This is the content of Lemmas 1, 2, 3 and 4, which are given in the appendix to the paper. The following proposition, which establishes the convergence of the FML estimator to $\boldsymbol{\eta}_1$ under the same generality for both the TDGP and MisM as highlighted above (*cf.* Chen and Deo, 2006, Corollary 1) now follows as an almost immediate corollary if we suppose that the following additional assumption holds:

$(A.4)$ There exists a unique pseudo-true parameter vector $\eta_1 = (d_1, \boldsymbol{\beta}_1^T)^T$ belonging to the subset $\mathbb{E}^0_\delta$ that satisfies $\boldsymbol{\eta}_1 = \arg\min_{\boldsymbol{\eta}} Q(\boldsymbol{\eta})$.

Proposition 1 *Suppose that the TDGP of* $\{y_t\}$ *is as prescribed in equations* (1) *and* (2) *and that the MisM is specified as in* (3)*. Assume also that Assumptions* $A.1 - A.4$ *are satisfied. Let* $\widehat{\boldsymbol{\eta}}^{(1)}_1$ *denote the FML estimator obtained by minimising the criterion function* $Q^{(1)}_n(\boldsymbol{\eta})$ *over* $\mathbb{E}_\delta$*. Then* $\lim_{n\to\infty} Q^{(1)}_n(\widehat{\boldsymbol{\eta}}^{(1)}_1) = Q(\boldsymbol{\eta}_1)$ *and* $\widehat{\boldsymbol{\eta}}^{(1)}_1 \to \boldsymbol{\eta}_1$ *almost surely.*

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

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

Before proceeding, we note that Cavaliere, Nielsen and Taylor (2017) have shown that if $g_0(\lambda)$ has a parametric form that is known, but the parameter values that characterise it are not, then the CSS parameter estimates will converge to a pseudo-true value if incorrect linear parameter constraints are imposed (Cavaliere, Nielsen and Taylor, 2017, Theorem 5(ii)). Theorem 1 provides a generalisation of this result: Firstly, by extending it to the FML, Whittle and TML estimators; Secondly, and perhaps more importantly, by allowing for the possibility that the parametric form of the model may itself be mis-specified and the characterisation of the short run dynamics by the function $g_1(\lambda)$ incorrect.

Having established that the four parametric estimators converge towards a common $\boldsymbol{\eta}_1$, we can as a consequence now broaden the applicability of the asymptotic distributional results derived by Chen and Deo (2006) for the FML estimator. This we do in Section 4 by establishing that all four alternative parametric estimators converge in distribution for all three forms of memory

- long memory, short memory and antipersistence. Prior to doing this, however, we indicate the precise form of the limiting objective function $Q(\boldsymbol{\eta})$, and the associated first-order conditions that define the pseudo-true value $\boldsymbol{\eta}_1$ of the four estimation procedures, in the ARFIMA case, once again under complete generality.

3 Pseudo-True Parameters Under ARFIMA Mis-Specification

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

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

$$\frac{\sigma_0^2}{2\pi} f_0(\lambda) = \frac{\sigma_0^2}{2\pi} \frac{\left|1 + \theta_{10}e^{i\lambda} + ... + \theta_{q_00}e^{iq_0\lambda}\right|^2}{\left|1 + \phi_{10}e^{i\lambda} + ... + \phi_{p_00}e^{ip_0\lambda}\right|^2} |2\sin(\lambda/2)|^{-2d_0},$$

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

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

Substituting $f_0(\lambda)$ and $f_1(\boldsymbol{\eta}, \lambda)$ into the limiting objective function in (6), we obtain the representation

$$Q(\boldsymbol{\eta}) = \frac{\sigma_0^2}{2\pi} \int_0^{\pi} \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta}, \lambda)} d\lambda = \frac{\sigma_0^2}{2\pi} \int_0^{\pi} \frac{|A_\beta(e^{i\lambda})|^2}{|B_\beta(e^{i\lambda})|^2} |2\sin(\lambda/2)|^{-2(d_0 - d)} d\lambda\,, \tag{16}$$

where

$$A_\beta(z) = \sum_{j=0}^{\underline{q}} a_j z^j = \theta_0(z)\phi(z) = (1 + \theta_{10}z + ... + \theta_{q_00}z^{q_0})(1 + \phi_1 z + ... + \phi_p z^p), \tag{17}$$

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

$$B_\beta(z) = \sum_{j=0}^{\underline{p}} b_j z^j = \phi_0(z)\theta(z) = (1 + \phi_{10}z + ... + \phi_{p_00}z^{p_0})(1 + \theta_1 z + ... + \theta_q z^q)\,, \tag{18}$$

with $\underline{p} = p_0 + q$. The expression for $Q(\boldsymbol{\eta})$ in (16) takes the form of the variance of an ARFIMA process with MA operator $A_\beta(z)$, AR operator $B_\beta(z)$ and fractional index $d_0 - d$. It follows that $Q(\boldsymbol{\eta})$ could be evaluated using the procedures presented in Sowell (1992). Sowell's algorithms are based upon series expansions in gamma and hypergeometric functions however, and although they are suitable for numerical calculations, they do not readily lend themselves to the analytical investigation of $Q(\eta)$. 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 (17) and (18) respectively. Then (16) can be expanded to give

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

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

$$Q(\boldsymbol{\eta}) = \{ \frac{\sigma_0^2 \Gamma(1-2(d_0-d))}{2\Gamma^2(1-(d_0-d))} \} K(\boldsymbol{\eta}) , \tag{19}$$

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

Using (19) we now derive the form of the first-order conditions that define $\boldsymbol{\eta}_1$, namely $\partial Q(\boldsymbol{\eta})/\partial \boldsymbol{\eta} = 0$. Differentiating $Q(\boldsymbol{\eta})$ first with respect to β_r, $r = 1, \ldots, l$, and then d gives:

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

where

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

and

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

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

$$\frac{\partial K(\boldsymbol{\eta})}{\partial \beta_r} = 0 , \quad r = 1, 2, ..., l, \tag{20}$$

and

$$2(\Psi[1-2(d_0-d)] - \Psi[1-(d_0-d)]) K(\boldsymbol{\eta}) + \frac{\partial K(\boldsymbol{\eta})}{\partial d} = 0 \tag{21}$$

for β_{r1}, $r = 1, \ldots, l$, and d_1 using appropriate algebraic and numerical procedures. A corollary of the following theorem is that $\boldsymbol{\eta}_1$ can be calculated to any desired degree of numerical accuracy by truncating the series expansions in the expressions for $K(\boldsymbol{\eta})$, $\partial K(\boldsymbol{\eta})/\partial \boldsymbol{\beta}$ and $\partial K(\boldsymbol{\eta})/\partial d$ after a suitable number of N terms before substituting into (20) and (21) and solving (numerically) for ϕ_{i1}, $i = 1, 2, ..., p$, θ_{j1}, $j = 1, 2, ..., q$, and d_1.

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

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

where

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

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

By way of illustration, consider the case of mis-specifying a true $ARFIMA(0, d_0, q_0)$ process by an $ARFIMA(p, d, 0)$ model. When $p_0 = q = 0$ we have $B_\beta(z) \equiv 1$ and $C(z)$ is polynomial, $C(z) = 1 + \sum_{j=1}^{\underline{q}} c_j z^j$ where $c_j = \sum_{r=\max\{0,j-p\}}^{\min\{j,p\}} \theta_{(j-r)0}\phi_r$. Abbreviating the latter to $\sum_r \theta_{(j-r)0}\phi_r$, this then gives us:

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

and setting $\theta_{s0} \equiv 0$ for when s does not belong to the set $\{0, 1, \ldots, q_0\}$,

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

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

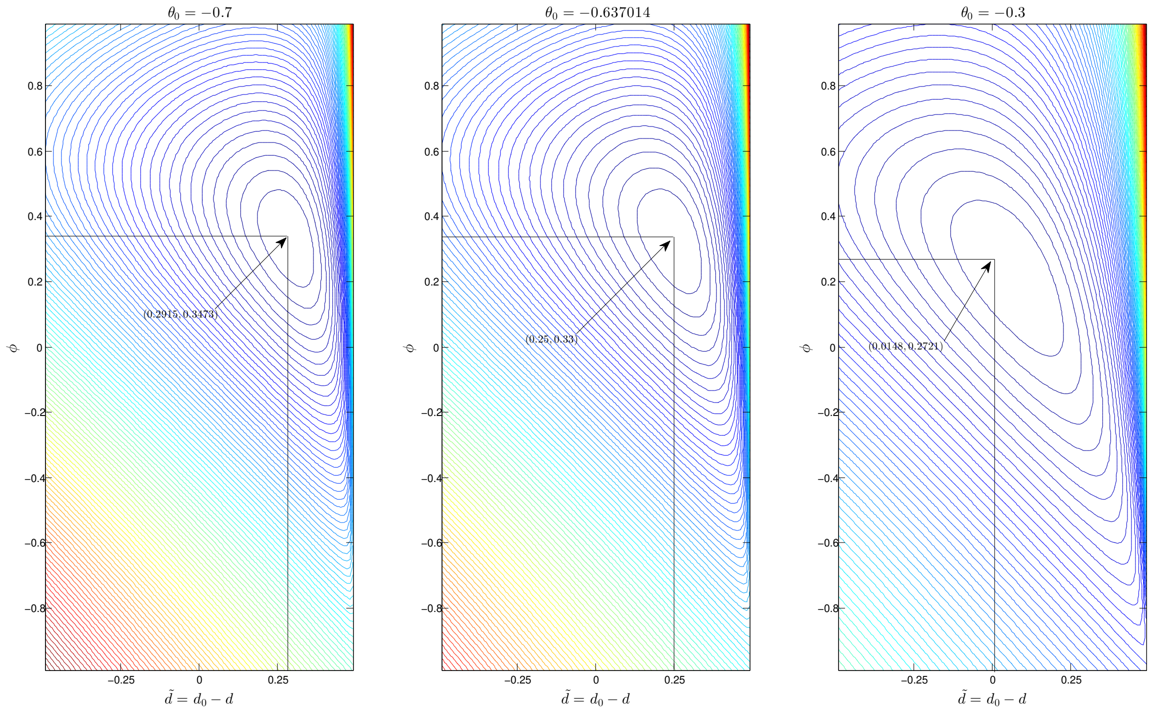
$$\begin{aligned} \frac{\partial K(d,\phi_1,\ldots,\phi_p)}{\partial d} = \; &2\sum_{k=0}^{\underline{q}-1}\sum_{j=k+1}^{\underline{q}} (\sum_r \theta_{(j-r)0}\phi_r)(\sum_r \theta_{(k-r)0}\phi_r)\rho(j-k) \times \\ &(2\Psi[1-(d_0-d)] - \Psi[1-(d_0-d)+(j-k)] - \Psi[1-(d_0-d)-(j-k)]) \end{aligned}$$

for the required derivatives. The pseudo-true values ϕ_{r1}, $r = 1, \ldots, p$, and d_1 can now be obtained by solving (20) and (21) having inserted these exact expressions for $K(d,\phi_1,\ldots,\phi_p)$, $\partial K(d,\phi_1,\ldots,\phi_p)/\partial\phi_r$, $r = 1,\ldots,p$, and $\partial K(d,\phi_1,\ldots,\phi_p)/\partial d$ into the equations.

Let us further highlight some features of this special case by focussing on the example where the TDGP is an $ARFIMA(0, d_0, 1)$ and the MisM an $ARFIMA(1, d, 0)$. In this example $\underline{q} = 2$ and $C(z) = 1 + c_1 z + c_2 z^2$ where, neglecting the first order MA and AR coefficient subscripts, $c_1 = (\theta_0 + \phi)$ and $c_2 = \theta_0\phi$. The second factor of the criterion function in (19) is now

$$\begin{aligned} K(d,\phi) = &1 + (\theta_0+\phi)^2 + (\theta_0\phi)^2 \\ &+ \frac{2\left[\theta_0\phi(d_0-d+1) - (1+\theta_0\phi)(\theta_0+\phi)(d_0-d-2)\right](d_0-d)}{(d_0-d-1)(d_0-d-2)}\,. \end{aligned} \tag{22}$$

The derivatives $\partial K(d,\phi)/\partial\phi$ and $\partial K(d,\phi)/\partial d$ can be readily determined from (22) and hence the pseudo-true values d_1 and ϕ_1 evaluated.



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

Figure 2: Contour plot of $Q(d, \phi)$ against $\widetilde{d} = d_0 - d$ and ϕ for the mis-specification of an ARFIMA$(0, d_0, 1)$ TDGP by an ARFIMA$(1, d, 0)$ 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) $(0.0148, 0.2721)$.

It is clear from (22) that for given values of $|\theta_0| < 1$ we can treat $K(d, \phi)$ as a function of $\widetilde{d} = (d_0 - d)$ and ϕ, and hence treat $Q(d, \phi) = Q(\boldsymbol{\eta})$ similarly. Figure 2 depicts the contours of $Q(d, \phi)$ graphed as a function of $\widetilde{d}$ and ϕ for the values of $\theta_0 = \{-0.7, -0.637014, -0.3\}$. 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 2 clearly demonstrate the divergence in the asymptotic criterion function that occurs as $\widetilde{d} = (d_0 - d)$ approaches 0.5 and they illustrate that although the location of (d_1, ϕ_1) may be unambiguous, the sensitivity of $Q(d, \phi)$ to perturbations in (d, ϕ) can be very different depending on the value of $d^* = d_0 - d_1$.[4] In Figure 2a 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 2b when $d^* = 0.25$. The contours of $Q(d, \phi)$ in Figure 2c, corresponding to $d^* < 0.25$, are more elliptical and suggest that in this case the limiting criterion function is far closer to being globally quadratic around (d_1, ϕ_1). It turns out that these three different forms of $Q(d, \phi)$, reflecting the most, intermediate, and the least mis-specified cases, correspond to the three different forms of asymptotic distribution presented in the following section.

4 Asymptotic Distributions

In this section we show that the asymptotic distribution of the FML estimator derived in Chen and Deo (2006) in the context of long-range dependence is also applicable to the Whittle, TML and CSS estimators, and that all four estimators are, hence, asymptotically equivalent under mis-specification. As was highlighted by Chen and Deo, the rate of convergence and the nature of the asymptotic distribution of the FML estimator is determined by the deviation of the pseudo-true

[4] These graphs have been produced using MATLAB *2011b*, version *7.13.0.564 (R2011b)*.

value d_1 from the true value d_0.[5]. Theorem 3 shows that in the event that any one of the FML, Whittle, TML or CSS estimators possesses one of the asymptotic distributions as described in the theorem, then all four estimators will share the same asymptotic distribution, and this will hold for all three forms of memory in the TDGP and the mis-specified model. We comment further on this matter below.

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

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

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

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

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

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

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

Assumption $(A.1')$ implies that $\{\varepsilon_t\}$ is completely regular, and that $\{\varepsilon_t^p\}$ is a uniformly integrable sequence for any $p \leq 4$. This assumption is closely related to Assumption $(A.1)$ of Lahiri (2003), which specifies a set of weak dependence and moment conditions on $\{\varepsilon_t\}$ based on α-mixing. When deriving asymptotic distributions it is typical to assume finite bounds on the first four moments of the innovation process (see Cavaliere, Nielsen and Taylor, 2017, for a detailed explanation of the importance of such bounds).

Theorem 3 *Suppose that the TDGP of* $\{y_t\}$ *is as prescribed in equations* (1) *and* (2)*, and that the MisM is specified as in* (3)*, and assume that Assumptions* $A.1'$ *and* $A.2 - A.5$ *hold. Let*

$$\mathbf{B} = -\frac{\sigma_0^2}{\pi}\int_{-\pi}^{\pi}\frac{f_0(\lambda)}{f_1^3(\boldsymbol{\eta}_1,\lambda)}\frac{\partial f_1(\boldsymbol{\eta}_1,\lambda)}{\partial\boldsymbol{\eta}}\frac{\partial f_1(\boldsymbol{\eta}_1,\lambda)}{\partial\boldsymbol{\eta}^T}d\lambda + \frac{\sigma_0^2}{2\pi}\int_{-\pi}^{\pi}\frac{f_0(\lambda)}{f_1^2(\boldsymbol{\eta}_1,\lambda)}\frac{\partial^2 f_1(\boldsymbol{\eta}_1,\lambda)}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T}d\lambda\,, \tag{23}$$

and set $\boldsymbol{\mu}_n = \mathbf{B}^{-1}E_0\left(\frac{\partial Q_n(\boldsymbol{\eta}_1)}{\partial\boldsymbol{\eta}}\right)$ *where* $Q_n(\cdot)$ *denotes the objective function that defines* $\widehat{\boldsymbol{\eta}}_1$.[6] *Let* $\widehat{\boldsymbol{\eta}}_1$ *denote the estimator obtained by minimising* $Q_n(\boldsymbol{\eta})$ *over the compact set* $\mathbb{E}_\delta$ *where* $\boldsymbol{\eta}_1 \in \mathbb{E}_\delta$,

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

[6]Heuristically, $\boldsymbol{\mu}_n$ measures the bias associated with the estimator $\widehat{\boldsymbol{\eta}}_1$. That is, $\boldsymbol{\mu}_n \approx E_0(\widehat{\boldsymbol{\eta}}_1) - \boldsymbol{\eta}_1$. Note that the expression for $\boldsymbol{\mu}_n$ given in Chen and Deo (2006, p 263) contains a typographical error; the proofs in that paper use the correct expression.

and assume that $\boldsymbol{\eta}_1$ does not belong to $\partial\mathbb{E}_\delta$, the boundary of the set $\mathbb{E}_\delta$. Then the FML, Whittle, TML or CSS estimators are asymptotically equivalent with a common limiting distribution as delineated in Cases 1, 2 and 3:

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

$$\frac{n^{1-2d^*}}{\log n}\left(\widehat{\boldsymbol{\eta}}_1 - \boldsymbol{\eta}_1 - \boldsymbol{\mu}_n\right) \xrightarrow{D} \mathbf{B}^{-1}\left(\sum_{j=1}^{\infty} W_j, 0, ...0\right)^T, \tag{24}$$

where $\sum\limits_{j=1}^{\infty} W_j$ 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

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

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

$$n^{1/2}\left[\overline{\Lambda}_{dd}\right]^{-1/2}\left(\widehat{\boldsymbol{\eta}}_1 - \boldsymbol{\eta}_1\right) \xrightarrow{D} \mathbf{B}^{-1}\left(Z, 0, ..., 0\right)^T, \tag{25}$$

where

$$\overline{\Lambda}_{dd} = \frac{1}{n}\sum_{j=1}^{n/2}\left(\frac{\sigma_0^2 f_0(\lambda_j)}{2\pi f_1(\boldsymbol{\eta}_1, \lambda_j)}\frac{\partial \log f_1(\boldsymbol{\eta}_1,\lambda_j)}{\partial d}\right)^2,$$

and Z is a standard normal random variable.

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

$$\sqrt{n}\left(\widehat{\boldsymbol{\eta}}_1 - \boldsymbol{\eta}_1\right) \xrightarrow{D} N(0, \boldsymbol{\Xi}), \tag{26}$$

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

$$\boldsymbol{\Lambda} = \frac{\sigma_0^4}{2\pi}\int_0^{\pi}\left(\frac{f_0(\lambda)}{f_1(\boldsymbol{\eta}_1,\lambda)}\right)^2\left(\frac{\partial \log f_1(\boldsymbol{\eta}_1,\lambda)}{\partial \boldsymbol{\eta}}\right)\left(\frac{\partial \log f_1(\boldsymbol{\eta}_1,\lambda)}{\partial \boldsymbol{\eta}}\right)^T d\lambda\,.$$

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 $d^* = 0.25$).[7]

For the form of limiting distribution that obtains in Cases 1, 2 and 3 we refer to Chen and Deo (2006, Theorems 1, 3 and 2), wherein these distributions were produced specifically for the FML

[7] As part of a companion set of simulation experiments documenting the relative finite sample performance of the four estimators under varying degrees of mis-specification (available on request) we develop a method for obtaining the non-standard limiting distribution applicable when $d^* > 0.25$. For all four estimation methods, the derivation of the bias-adjustment term $\boldsymbol{\mu}_n$ which is relevant in this case is provided in Supplementary Appendix B.

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

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

$$\left\| \frac{\partial^2 \{ \mathcal{C}_n(\boldsymbol{\eta}_1) \cdot Q_n^{(i)}(\boldsymbol{\eta}_1) - Q_n^{(j)}(\boldsymbol{\eta}_1) \}}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} \right\| = o_p(1) \tag{27}$$

and

$$\operatorname{plim}_{n\to\infty} \quad R_n \left\| \mathcal{C}_n(\boldsymbol{\eta}_1) \cdot \frac{\partial Q_n^{(i)}(\boldsymbol{\eta}_1)}{\partial \boldsymbol{\eta}} - \frac{\partial Q_n^{(j)}(\boldsymbol{\eta}_1)}{\partial \boldsymbol{\eta}} \right\| = 0 \, . \tag{28}$$

The condition in (27) is established by showing that $\partial^2 \{ Q_n^{(1)}(\boldsymbol{\eta}_1) \} / \partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'$ converges in probability to $\mathbf{B}$, as defined in (23), and that for each $i = 2, 3$ and 4 the corresponding Hessian is proportional to $\partial^2 \{ Q_n^{(1)}(\boldsymbol{\eta}_1) \} / \partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'$ with probability approaching one. The proof of (27) is fairly conventional, whereas the proof of (28) – which implicitly invokes the Cramér-Wold device since the moments (cumulants) of the asymptotically normal gradient vector are convergence determining for the limiting distributions in Theorem 3 – is more involved because of the presence of the scaling factor R_n. In Supplementary Appendix A we present the steps necessary to prove (27) and (28) for each estimator, and for TDGPs with fractional indices in the range $-0.5 < d_0 < 0.5$.

5 Discussion

This paper presents theoretical results relating to the estimation of mis-specified models for fractionally integrated processes. We show that under mis-specification four classical parametric estimation methods, frequency domain maximum likelihood (FML), Whittle, time domain maximum likelihood (TML) and conditional sum of squares (CSS) converge to the same pseudo-true parameter value. Consistency of the four estimators for the pseudo-true value is proved for fractional exponents of both the true and estimated models in the long memory, short memory and antipersistent ranges. A general closed-form solution for the limiting criterion function for the four alternative estimators is derived in the case of ARFIMA models. This enables us to link the form of mis-specification of the short memory dynamics to 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.

There are several interesting issues that arise from the results that we have established, including the following:

First, as already noted, although the known (zero) mean assumption is inconsequential for the FML and Whittle estimators, this is not the case for the time domain estimators. Estimation of μ will impact on the limiting distribution of the time domain estimators; if the sample mean is used, for example, then the limiting behaviour of the estimators may be influenced by the slower than usual $n^{(0.5-d_0)}$ convergence rate of the sample mean, given that the rate of convergence of the estimators (for all ranges of values of $d^* = d_0 - d_1$) differs from this rate when the true mean is known. This is a matter that we have not pursued for the current paper, but is the subject of other ongoing research.

Second, the extension of our results to non-stationary cases will facilitate the consideration of a broader range of circumstances. To some extent non-stationary values of d might be covered by means of appropriate pre-filtering, for example, the use of first-differencing when $d_0 \in [1/2, 3/2)$, 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 interval $d \in [0, 3/2)$, say, allowing for both stationary and non-stationary cases perhaps offers a better approach as prior knowledge of the characteristics of the process would then be unnecessary. The latter also seems particularly relevant given that estimates near the boundaries $d = 0.5$ and $d = 1$ are not uncommon in practice. Previous developments in the analysis of non-stationary fractional processes (see, inter alios, Beran, 1995; Tanaka, 1999; Velasco, 1999; Velasco and Robinson, 2000) might offer a sensible starting point for such an investigation.

Third, in the spirit of Chen and Deo (2006), on which this work builds, the focus of the asymptotic results derived here (and the simulation results documented on the first author's website) is on estimation of the pseudo-true value, d_1 (see also Yajima, 1992). This focus has derived from the very nature of the exercise, namely to characterize the impact of mis-specification of the short memory dynamics on estimation of the long memory parameter by documenting the link between the *degree* of mis-specification and the extent of the deviation of d_1 from d_0. The extent of this deviation determines, in turn, the convergence, or otherwise, of the four estimators to d_1, the rate of that convergence and the nature of the asymptotic distribution centred on d_1, including the (finite sample) bias adjustment term, $\boldsymbol{\mu}_n$, that obtains in the case of the most extreme mis-specification (i.e. $d^* > 0.25$). In other words, by the very nature of the analysis, d_1, and the estimation thereof, is the focal point, with the results providing guidance about the extent, and type, of sampling variation one can observe when estimating the long memory parameter in a mis-specified model.

As a first step, this is a critical piece of knowledge with which an empirical researcher should be armed. The second step however, is to understand the sampling properties of the different estimators - across the mis-specification spectrum - *as estimators of* d_0. Some insight into the potential difference between this behaviour and the accuracy of the methods as estimators of d_1 is gained as follows. The relationships between the bias and MSE of the parametric estimators ($\widehat{d}_1$, generically) 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:

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

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 any 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: 1) Does the ranking of the mis-specified estimators change if the true value of d is the reference value and does the finite sample bias adjustment $\boldsymbol{\mu}_n$, relevant for the case where $d^* > 0.25$, play a role in that switch? 2) 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? 3) Can knowledge of the asymptotic distribution of the mis-specified parametric estimators, as estimators of d_1, be used to undertake inference (e.g. produce confidence intervals) for the true parameter, d_0, that is, in some sense robust to certain forms of mis-specification? One possible approach to 3) might be to make a worst-case Bias_d_0 correction to $\widehat{d_1}$ in order to form a confidence region for d_0, perhaps using the analysis in Section 3 to gauge the extent of $d_1 - d_0$ under any particular form of misspecification of the short-run dynamics that the practitioner has in mind. This idea, plus that of using the distributional results in Section 4 to develop appropriate inferential tools and diagnostic devices are beyond the scope of this short paper, but are the focus of current and ongoing research.

References

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica* **71** 579–625.

Baillie, R. T., Bollerslev, T. and Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of econometrics* **74** 3–30.

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.

Breidt, F. J., Crato, N. and De Lima, P. (1998). The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics* **83** 325–348.

Boes, D. C., Davis, R. A. and Gupta, S. N. (2006). Parameter estimation in low order fractionally differenced ARMA processes. *Stochastic Hydrology and Hydraulics* **3** 97–110.

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

Cavaliere, G., Nielsen, M. O. and Taylor, A. M. R. (2017). Quasi-Maximum Likelihood Estimation and Bootstrap Inference in Fractional Time Series Models with Heteroskedasticity of Unknown Form. *Journal of Econometrics* **198** 165 – 188.

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.

Comte, F. and Renault, E. (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance* **8** 291–323.

Crato, N. and Taylor, H. M. (1996). Stationary persistent time series misspecified as non-stationary ARIMA. *Statistical Papers* **37** 215–266.

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

Deo, R. S. and Hurvich, C. M. (2001). On the log periodogram regression estimator of the memory parameter in long memory stochastic volatility models. *Econometric Theory* **17** 686–710.

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.

Giraitis, L. and Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probability Theory and Related Fields* **86** 87–104.

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

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.

Hassler, U. (1994). (Mis)Specification of long memory in seasonal time series. *Journal of Time Series Analysis* **14** 19–30.

Hosking, J. R. M. (1996). Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long memory time series. *Journal of Econometrics* **73** 261–284.

Hualde, J. and Robinson, P.M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models. Annals of Statistics. **39**, 3152–3181.

Hurvich, C. M. and Beltrao, K. I. (1993). Asymptotics for the low-frequency ordinates of the periodogram of a long memory time series. *Journal of Time Series Analysis* **14** 455–472.

Hurvich, C. M., Moulines, E. and Soulier, P. (2005). Estimating long memory in volatility. *Econometrica* **73** 1283–1328.

Hurvich, C. M. and Ray, B. K. (2003). The local Whittle estimator of long-memory stochastic volatility. *Journal of Financial Econometrics* **1** 445–470.

Koopman, S. J., Jungbacker, B. and Hol, E. (2005). Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance* **12** 445–475.

Lahiri, S. N. (2003). A necessary and sufficient condition for asymptotic independence of discrete fourier transforms under short- and long-range dependence. *Annals of Statistics* **31** 613-641

Lieberman, O. and Phillips, P. C. B. (2004). Error bounds and asymptotic expansions for toeplitz product functionals of unbounded spectra. *Journal of Time Series Analysis* **25** 732–753.

Martin, G., Reidy, A. and Wright, J. (2009). Does the option market produce superior forecasts of noise-corrected volatility measures?. *Journal of Applied Econometrics* **24** 77–104.

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

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

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

Pong, S., Shackleton, M. B., Taylor, S. J. and Xu, X. (2004). Forecasting currency volatility: A comparison of implied volatilities and AR(FI)MA models. *Journal of Banking & Finance* **28** 2541–2563.

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

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.

Robinson, P. M. (2014). The estimation of misspecified long memory models *Journal of Econometrics*, **178** 225–230.

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.

Velasco, C. and Robinson, P. M. (2000). Whittle pseudo-maximum likelihood estimation for nonstationary time series. *Journal of the American Statistical Association* **95:452** 1229–1243.

Whittle, P. (1953). Estimation and information in stationary time series. *Ark. Mat.* **2** 423–434.

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: Lemmas 1-4

Lemma 1 *Suppose that the TDGP of* $\{y_t\}$ *is as prescribed in equations* (1) *and* (2) *and that the MisM is specified as in* (3). *Assume also that Assumptions* $A.1 - A.3$ *are satisfied. Then for any constant* $\nu_f > 0$

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

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

Since, obviously, $f_1(\boldsymbol{\eta}, \lambda) < f_1(\boldsymbol{\eta}, \lambda) + \nu_f$ it follows from Lemma 1 that,

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

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

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

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

Lemma 2 *Suppose that the conditions of Lemma 1 hold. Set*

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

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

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

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

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

Lemma 3 *Suppose that the conditions of Lemmas 1 and 2 hold. Then*

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

Lemma 4 then indicates that for points in $\mathbb{E}_\delta$ where $(d_0 - d) > 0.5 - \delta$, $0 < \delta < 0.5$, uniform convergence of the criterion function $Q_n^{(1)}(\boldsymbol{\eta})$ fails.

Lemma 4 *Suppose that the TDGP of* $\{y_t\}$ *is as prescribed in equations* (1) *and* (2) *and that the MisM is specified as in* (3). *Assume also that Assumptions* $A.1 - A.3$ *are satisfied. Then for all* $\boldsymbol{\eta} \in \overline{\mathbb{E}}_{\delta 1}^0$ *we have* $\liminf_{n\to\infty} Q_n^{(1)}(\boldsymbol{\eta}) = O(\delta^{-1})$ *and for* $\boldsymbol{\eta} \in \overline{\mathbb{E}}_{\delta 2}^0$

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

almost surely for all C, *no matter how large.*

Note that Lemma 4 implies that as n increases, and for all δ sufficiently small, $\widehat{\boldsymbol{\eta}}_1^{(1)} = \arg\min_{\boldsymbol{\eta}} Q_n^{(1)}(\boldsymbol{\eta})$ cannot lie in $\overline{\mathbb{E}}_{\delta 1}^0 \cup \overline{\mathbb{E}}_{\delta 2}^0$.