EXPANSIONS FOR APPROXIMATE MAXIMUM LIKELIHOOD ESTIMATORS OF THE FRACTIONAL DIFFERENCE PARAMETER

BY

OFFER LIEBERMAN AND PETER C. B. PHILLIPS

COWLES FOUNDATION PAPER NO. 1157

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2006

http://cowles.econ.yale.edu/

Econometrics Journal (2005), volume **8**, pp. 367–379.

Expansions for approximate maximum likelihood estimators of the fractional difference parameter

OFFER LIEBERMAN* AND PETER C. B. PHILLIPS†

**Technion–Israel Institute of Technology*

†*Cowles Foundation for Research in Economics, Yale University, University of Auckland and University of York*

Received: November 2003

Summary This paper derives second-order expansions for the distributions of the Whittle and profile plug-in maximum likelihood estimators of the fractional difference parameter in the 'Autoregressive Fractionally Integrated Moving Average of order (0, d, 0)' with unknown mean and variance. Both estimators are shown to be second-order pivotal. This extends earlier findings of Lieberman and Phillips (2004a, *Econometric Theory, 20,* 464–84), who derived expansions for the Gaussian maximum likelihood estimator under the assumption that the mean and variance are known. One implication of the results is that the parametric bootstrap upper one-sided confidence interval provides an $o(n^{-1}\ln n)$ improvement over the delta method. For statistics that are not second-order pivotal, the improvement is generally only of the order $o(n^{-1/2}\ln n)$.

Key words: *ARFIMA*, *Bootstrap*, *Edgeworth expansion*, *Fractional differencing*, *Pivotal statistic*.

1. INTRODUCTION

We consider the model

$$(1-B)^d\,(X_t-\mu)=\varepsilon_t,\, t=1,\ldots,n, \tag{1}$$

where B is the backshift operator, $d\in(0,1/2)$, $\varepsilon_t \overset{\text{i.i.d.}}{\sim} N(0,\sigma^2)$ and μ, σ^2 are unknown. In the canonical case with known mean and variance, Lieberman and Phillips (2004a, henceforth, LP) showed that the distribution of normalized Gaussian maximum likelihood estimator (MLE) of d, $\hat{\delta}_n=\sqrt{n}(\hat{d}_n-d)$, admits the expansion

$$\tilde{H}^{(1),A}_{\hat{\delta}_n}(x/\sqrt{\kappa_{n,1,1}})=\Phi\,(x)+\frac{\zeta\,(3)}{\sqrt{n}\zeta^{3/2}\,(2)}\phi\,(x)\left\{x^2+2\right\}, \tag{2}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal cdf and pdf, respectively, $\zeta\ (\cdot)$ is the Riemann-zeta function and $\kappa_{n,1,1}$ is the variance of the score function. The expansion is uniform and valid in the sense that

$$\sup_{x\in R}\sup_{d\in D^*}\left|\Pr_{d_0}(\hat{\delta}_n\le x/\sqrt{\kappa_{n,1,1}})-\tilde{H}^{(1),A}_{\hat{\delta}_n}(x/\sqrt{\kappa_{n,1,1}})\right|=o(n^{-1/2}),$$

where d_0 is the true value of d and D^* is any compact subset of (0, 1/2).

To our knowledge, the formula (2) is the only explicit expansion known in a parametric long-memory context. It shows that $\hat{\delta}_n$ is second-order pivotal. This feature seems rare in time series contexts. In contrast, for example, even the asymptotic distribution of the first-order serial correlation coefficient depends on the autoregressive parameter in a stationary AR(1).

While the results for the canonical model may be interesting from a theoretical view point, they are of limited practical use, since the mean and variance are assumed known. While still specialized, the model with unknown μ and σ^2 is popular and has been applied in a number of disciplines, including economics and finance, so it is of interest to extend the higher-order analysis to this case. For applications of the model, see Geweke and Porter Hudak (1983) and Baillie (1996) and the references therein.

The main result of the present paper is that for the model (1) with unknown mean and variance, the second-order expansion for the distribution of either the Whittle MLE (WMLE) or the profile-plug-in MLE (PPMLE) of d is of the same form as (2). Thus, these estimators of d are second-order pivotal when the mean and variance are unknown.

A few remarks on context are in order. First, as far as we know, there are presently no explicit expansions for the WMLE in the long-memory literature, even for simple models like (1), although Taniguchi developed expansions under short memory (see Taniguchi and Kakizawa's review (2000, chapter 4)). Second, we show that the WMLE and PPMLE expansions have terms which generally differ by $O(n^{-\frac{1}{2}+\epsilon})$, $\forall \epsilon > 0$. The comparison between higher-order expansions of the two estimators is novel, refining earlier work confirming the asymptotic equivalence of the two estimators. Finally, the implication of the second-order pivotal result is that the improvement of the parametric bootstrap upper confidence interval for d over the delta method upper confidence interval is of the order $o(n^{-1} \ln(n))$, compared with an improvement of the order $o(n^{-1/2} \ln(n))$ for non-pivotal statistics. See Andrews and Lieberman (2002). This result shows that there is some practical import in the second-order expansion.

The work in this paper continues some recent literature on higher-order theory for fractional Gaussian processes. Validity of the Edgeworth expansion for the distribution of the Gaussian MLE and the WMLE under strong dependence was established in Lieberman *et al.* (2003) and Andrews and Lieberman (2005), respectively. Those papers prove validity but were not concerned with developing explicit expansions. Lieberman and Phillips (2004a) found explicit expansions in the canonical ARFIMA(0, d, 0) model and Andrews and Lieberman (2002) used results on Edgeworth expansions to prove higher-order improvements of the parametric bootstrap under strong dependence. In other work, Lieberman and Phillips (2004b) established the error rate of the integral limit of matrix product functionals of unbounded spectra and this result is used extensively in the development of the expansions in the present paper. In particular, the results are used in the investigation of the difference between the expansions for the WMLE and the PPMLE. The above results relate to parametric long-memory models. In a semi-parametric framework, Giraitis and Robinson (2003) gave an Edgeworth expansion for the local Whittle estimator of the memory parameter. The error rate in the semiparametric expansion is slower than the parametric rate and depends on the number of frequencies (m) used in estimation, where it is assumed that $m \to \infty$ so slowly that the bias effect is smaller than $m^{-1/2}$. Giraitis and Robinson found the Whittle estimator to be independent of d to order $m^{-1/2}$, which is analogous to the pivotal property found for the parametric estimator in the present paper on parametric estimation.

The plan for the rest of the paper is as follows. In Section 2 we identify 'exact' and 'approximate' expansions for the distribution of the WMLE. By 'exact', it is meant that the terms in the Edgeworth expansions depend on n and are $O(1)$ (see, e.g., Durbin (1980, equation (28))) and by 'approximate' it is meant that the limits of these terms have been taken. Section 3 develops similar expansions for the PPMLE. The approximate WMLE and PMLE expansions are identical to second order and agree with the expansion found by LP in the canonical case. However, the exact expansions differ. Final remarks are given in Section 4.

2. SECOND-ORDER EXPANSIONS FOR THE WHITTLE MLE

This section provides second-order Edgeworth expansion for the distribution of the WMLE and shows that this expansion is identical to the expansion obtained in LP for the exact Gaussian MLE in the canonical case, giving the second-order pivotal property of the WMLE of d. Using Szegö's identity, we express the Whittle likelihood as a summand of two terms, with dependence on d only through the second term in the summand, which is a scaled quadratic form in Gaussian long-memory variables. The decomposition reveals that the solution to the WMLE, as well as its distribution, are independent of the scale parameter and μ.

Let $\theta = (d, \sigma^2)$. The spectral density of the process is given by

$$f_\theta(\lambda) = \frac{\sigma^2}{2\pi} \left|1 - e^{i\lambda}\right|^{-2d}.$$

Denote the covariance matrix by $T_n(f_\theta)$. The Whittle log likelihood is given by

$$L_n^W(\theta) = -\frac{n}{4\pi} \int_\pi \left[\log f_\theta(\lambda) + \frac{I_n(\lambda)}{f_\theta(\lambda)}\right] d\lambda\,, \tag{3}$$

where

$$I_n(\lambda) = \frac{1}{2\pi n} \left|\sum_{j=1}^n e^{i\lambda j}(X_j - \bar{X}_n)\right|^2, \quad \bar{X}_n = n^{-1}\Sigma X_j.$$

We can write the second summand in (3) as

$$-\frac{1}{2}\sum_{j,k=1}^n (X_j - \bar{X}_n)(X_k - \bar{X}_n)\left\{\frac{1}{4\pi^2}\int_\Pi f_\theta^{-1}(\lambda) e^{i(j-k)\lambda} d\lambda\right\} = -\frac{1}{2} x_n' M_n T_{W,n} M_n x_n,$$

where $M_n = I_n - P_n$, $P_n = n^{-1} 1_n 1_n'$, 1_n is an n-vector of 1's, $x_n = (X_1, \ldots, X_n)'$ and

$$T_{W,n} = T_n\left(\frac{1}{4\pi^2 f_\theta}\right).$$

The matrix $T_{W,n}$ is the Whittle approximation to $T_n^{-1}(f_\theta)$. The Whittle log likelihood is thus given by

$$L_n^W(\theta) = -\frac{n}{4\pi}\int_\Pi \log f_\theta(\lambda)\, d\lambda - \frac{1}{2} x_n' M_n T_{W,n} M_n x_n.$$

Write $f_d(\lambda) = \frac{1}{2\pi}|1 - e^{-i\lambda}|^{-2d}$ and $T_{W,n}^d = T_n\left((4\pi^2 f_d)^{-1}\right)$. That is, $T_{W,n}^d$ is the Whittle matrix of the ARFIMA(0, d, 0) model with a unit variance. Szegö's identity

$$\log\sigma^2 = \frac{1}{2\pi}\int_\Pi \log\left(2\pi f_\theta(\lambda)\right) d\lambda,$$

implies that

$$\int_{\Pi} \log\left(2\pi f_d(\lambda)\right) d\lambda = 0.$$

Thus,

$$L_n^W = \frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}x_n' M_n T_{W,n}^d M_n x_n,$$

and the WMLE of d solves

$$\frac{\partial L_n^W}{\partial d} = -\frac{1}{2\sigma^2}x_n' M_n \dot{T}_{W,n}^d(\hat{d}_{W,n}) M_n x_n = 0,$$

where $\hat{d}_{W,n}$ is the WMLE of d. Evidently, the solution for $\hat{d}_{W,n}$ depends on d through the quadratic form only. More generally, Szegö's identity implies that for any ARFIMA(p, d, q) model, the Whittle likelihood depends on all the parameters apart from σ^2 through the quadratic form only. This is not the case with the exact likelihood, where there is dependence on the ARFIMA parameters through the logarithm of the determinant of the covariance matrix. Unlike the exact Gaussian MLE then, the solution for $\hat{d}_{W,n}$ does not depend on $\hat{\sigma}^2$ but does depend on $\bar{X}_n$ through M_n.

To proceed, we recall theorem 6 of LP, which gives a formal second-order expansion for the density of $\hat{\delta}_n$. The expansion is

$$\tilde{h}_{\hat{\delta}_n}^{(1)}(u; d) = \phi\left(u; \kappa_{n,1,1}^{-1}(d)\right)\left\{1 + \frac{1}{\sqrt{n}}\left[C_{n,1}^*(d)\,u + C_{n,3}^*(d)\,u^3\right]\right\}, \tag{4}$$

for terms $\kappa_{n,1,1}^{-1}(d)$, $C_{n,1}^*(d)$ and $C_{n,3}^*(d)$, which are functions of expected values of Gaussian log-likelihood derivatives under the known mean and variance assumption and are defined through equations (10)–(12) of LP. It is clear from the proof of theorem 6 of LP that the general form of (4) extends to the Whittle likelihood, but with terms $\kappa_{W,n,1,1}^{-1}(d)$, $C_{W,n,1}^*(d)$ and $C_{W,n,3}^*(d)$ replacing their analogues in (4). In other words, to identify the expansion for $\hat{\delta}_{W,n} = \sqrt{n}(\hat{d}_{W,n} - d_0)$, we need to find terms $\kappa_{W,n,1,1}^{-1}(d)$, $C_{W,n,1}^*(d)$ and $C_{W,n,3}^*(d)$, which are precisely the same functions of the Whittle likelihood derivatives as $\kappa_{n,1,1}^{-1}(d)$, $C_{n,1}^*(d)$ and $C_{n,3}^*(d)$ are of the Gaussian log-likelihood derivatives. To do so, we define

$$L_{W,n,j} = -\frac{1}{2\sigma^2}x' M_n\left(\frac{\partial^j}{\partial d^j}T_{W,n}^d\right)M_n x_n, \quad j = 1, 2, 3,$$

$$Z_{W,n,j} = \frac{1}{\sqrt{n}}(L_{W,n,j} - E_\theta(L_{W,n,j})), \quad j = 1, 2\,.$$

We find

$$\kappa_{W,n,1,1} = \mathrm{Var}(Z_{W,n,1}) = \frac{1}{2n}\mathrm{tr}\left[M_n \dot{T}_{W,n}^d M_n T_n^d\right]^2, \tag{5}$$

$$C_{W,n,1}^* = -\frac{2\mu_{W,n,3}' + 4\kappa_{W,n,1,2} + \kappa_{W,n,1,1,1}}{2\kappa_{W,n,1,1}}, \tag{6}$$

$$C_{W,n,3}^* = \frac{\kappa_{W,n,1,1,1}}{6} + \frac{\mu_{W,n,3}'}{2} + \kappa_{W,n,1,2}, \tag{7}$$

$$\mu'_{W,n,3} = \frac{1}{n}\, E_d(L_{W,n,3}) \\ = -\frac{1}{2n}\, \mathrm{tr}\big(M_n \ddot{T}^d_{W,n} M_n T^d_n\big), \tag{8}$$

$$\kappa_{W,n,1,2} = \mathrm{cov}(Z_{W,n,1}, Z_{W,n,2}) \\ = \frac{1}{2n}\, \mathrm{tr}\big(M_n \dot{T}^d_{W,n} M_n T^d_n M_n \ddot{T}^d_{W,n} M_n T^d_n\big) \tag{9}$$

$$\kappa_{W,n,1,1,1} = \frac{1}{n}\, \mathrm{tr}\left(M_n \dot{T}^d_{W,n} M_n T^d_n\right)^3. \tag{10}$$

With (5)–(10), the exact second-order expansion for the density of $\hat{\delta}_{W,n}$ is

$$\tilde{h}^{(1)}_{\hat{\delta}_{w,n}}(u;d) = \phi\big(u; \kappa^{-1}_{W,n,1,1}(d)\big)\left\{1 + \frac{1}{\sqrt{n}}\left[C^*_{W,n,1}(d)u + C^*_{W,n,3}(d)u^3\right]\right\}. \tag{11}$$

The expansion (11) is exact in the sense that the terms in it involve exact expectations of Whittle likelihood derivatives. The terms $\kappa^{-1}_{W,n,1,1}$, $C^*_{W,n,1}(d)$ and $C^*_{W,n,3}(d)$ depend on n and are $O(1)$. Edgeworth expansions with coefficients depending on n are standard in the literature. See, for example, Durbin (1980, equation (28)). The expansion is a nonlinear function of d through these terms. We proceed to obtain an approximate expansion by seeking limits of $\kappa^{-1}_{W,n,1,1}$, $C^*_{W,n,1}(d)$ and $C^*_{W,n,3}(d)$ as $n \to \infty$. It turns out that these limits do not depend on d. Hence, it turns out that $\hat{\delta}_{W,n}$ is second-order pivotal.

Define $C(\lambda) = \log\left[2(1-\cos\lambda)\right]$ and note that for any $j \in N$,

$$\frac{\partial^j f_d^{-1}(\lambda)}{\partial d^j} = \frac{C^j(\lambda)}{f_d(\lambda)} = O\left(|\lambda|^{2d-\epsilon}\right) \quad \text{as } \lambda \to 0, \quad \forall \epsilon > 0. \tag{12}$$

We know from theorem 3 of Andrews and Lieberman (2005), which holds under (12), that

$$\frac{1}{n}\,\mathrm{tr}\left(M_n \overset{*}{T}{}^d_{W,n} M_n T^d_n\right)^h = \frac{1}{n}\,\mathrm{tr}\left(\overset{*}{T}{}^d_{W,n} T^d_n\right)^h + O\left(n^{-1+\epsilon}\right), \forall \epsilon > 0,$$

where h is any finite positive integer and the $*$ on $T^d_{W,n}$ denotes any number of derivatives. Applying theorem 5 of Lieberman and Phillips (2004b), which also holds under (12), we get

$$\left|\frac{1}{n}\,\mathrm{tr}\left(\overset{*}{T}{}^d_{W,n} T^d_n\right)^h - (2\pi)^{-1}\int_\Pi \left[\frac{\partial^j f_d^{-1}(\lambda)}{\partial d^j} f(\lambda)\right]^h d\lambda\right| = O\left(n^{-1+2\epsilon}\right), \quad \forall\, \epsilon > 0,$$

where j is the number of dots (derivatives) signified in $*$. Using the results (e.g., p. 13 of LP)

$$\int_\Pi C^2(\lambda) = \frac{2\pi^3}{3}, \qquad \int_\Pi C^3(\lambda) = -24\pi\xi(3), \qquad \int_\Pi C^4(\lambda) = 228\pi\xi(4).$$

We obtain

$$\begin{aligned}\kappa_{W,n,1,1} &= \frac{1}{2n}\operatorname{tr}\left(M_n\dot{T}^d_{W,n}M_nT^d_n\right)^2\\ &= \frac{1}{2n}\operatorname{tr}\left(\dot{T}^d_{W,n}T^d_n\right)^2 + O\left(n^{-1+\epsilon}\right)\\ &= \frac{1}{4\pi}\int_\Pi\left(\frac{\dot{f}_d(\lambda)}{f_d(\lambda)}\right)^2 d\lambda + O\left(n^{-1+2\epsilon}\right)\\ &= \frac{1}{4\pi}\int_\Pi C^2(\lambda)\,d\lambda + O\left(n^{-1+2\epsilon}\right)\\ &= \frac{\pi^2}{6} + O\left(n^{-1+2\epsilon}\right). \qquad (13)\end{aligned}$$

Set

$$\kappa_{W,1,1} = \frac{\pi^2}{6}. \qquad (14)$$

Further,

$$\begin{aligned}\kappa_{W,n,1,1,1} &= \frac{1}{n}\operatorname{tr}\left(M_n\dot{T}^d_{W,n}M_nT^d_n\right)^3\\ &= \frac{1}{n}\operatorname{tr}\left(\dot{T}^d_{W,n}T^d_n\right)^3 + O\left(n^{-1+\epsilon}\right)\\ &= \frac{1}{2\pi}\int_\Pi\left(\frac{\dot{f}_d(\lambda)}{f_d(\lambda)}\right)^3 d\lambda + O\left(n^{-1+2\epsilon}\right)\\ &= -\frac{1}{2\pi}\int_\Pi C^3(\lambda)\,d\lambda + O\left(n^{-1+2\epsilon}\right)\\ &= 12\xi(3) + O\left(n^{-1+2\epsilon}\right).\end{aligned}$$

Set

$$\kappa_{W,1,1,1} = 12\xi(3). \qquad (15)$$

Applying again theorem 5 of Lieberman and Phillips (2004b), we have

$$\begin{aligned}\kappa_{W,n,1,2} &= \frac{1}{2n}\operatorname{tr}\left(\dot{T}^d_{W,n}T^d_n\ddot{T}^d_{W,n}T^d_n\right) + O\left(n^{-1+\epsilon}\right)\\ &= \frac{1}{4\pi}\int_\Pi\left[\frac{\partial f_d^{-1}(\lambda)}{\partial d}\,\frac{\partial^2 f_d^{-1}(\lambda)}{\partial d^2}f^2(\lambda)\right]d\lambda + O\left(n^{-1+2\epsilon}\right)\\ &= \frac{1}{4\pi}\int_\Pi C^3(\lambda)\,d\lambda + O\left(n^{-1+2\epsilon}\right)\\ &= -6\zeta\,(3) + O\left(n^{-1+2\epsilon}\right).\end{aligned}$$

Set

$$\kappa_{W,1,2} = -6\xi(3). \qquad (16)$$

Also,

$$\mu'_{W,n,3} = -\frac{1}{2n}\,\mathrm{tr}\left(\dddot{T}^{d}_{W,n}T^{d}_{n}\right) + O\left(n^{-1+\epsilon}\right).$$

The third-order derivative of $f_d^{-1}(\lambda)$ is

$$\frac{\partial^3 f_d^{-1}}{\partial d^3} = -6f_d^{-4}\dot{f}_d^3 + 6f_d^{-3}\dot{f}_d\ddot{f}_d - f_d^{-2}\,\dddot{f}_d\,.$$

So,

$$\begin{aligned}\mu'_{W,n,3} &= -\frac{1}{4\pi}\int_{\Pi}\left(-6\,\frac{\dot{f}_d^3(\lambda)}{f_d^3(\lambda)} + 6\,\frac{\dot{f}_d(\lambda)\ddot{f}_d(\lambda)}{f_d^2(\lambda)} - \frac{\dddot{f}_d\,(\lambda)}{f_d(\lambda)}\right)d\lambda + O\left(n^{-1+2\epsilon}\right)\\ &= -\frac{1}{4\pi}\int_{\Pi}C^3(\lambda)d\lambda + O\left(n^{-1+2\epsilon}\right)\\ &= 6\xi(3) + O\left(n^{-1+2\epsilon}\right).\end{aligned}$$

Set

$$\mu'_{W,3} = 6\xi(3)\,. \tag{17}$$

Using (5)–(10) and (13)–(17),

$$\begin{aligned}C^*_{W,n,1} &= -\left[\frac{12\xi(3) - 24\xi(3) + 12\xi(3)}{2\frac{\pi^2}{6}}\right] + O\left(n^{-1+2\epsilon}\right)\\ &= O\left(n^{-1+2\epsilon}\right),\end{aligned} \tag{18}$$

and

$$\begin{aligned}C^*_{W,n,3} &= \frac{12\xi(3)}{6} + \frac{6\xi(3)}{2} - 6\xi(3) + O\left(n^{-1+2\epsilon}\right)\\ &= -\xi(3) + O\left(n^{-1+\epsilon}\right).\end{aligned} \tag{19}$$

Using (13)–(19) in (11), the approximate second-order expansion to the density of $\delta^*_{W,n} = \sqrt{\kappa_{W,n,1,1}}\hat{\delta}_{W,n}$ is given by

$$\tilde{h}^{(1),A}_{\delta^*_{W,n}} = \phi\,(u)\left\{1 - \frac{\xi(3)}{\sqrt{n}\zeta^{3/2}\,(2)}u^3\right\}, \tag{20}$$

which is *identical* to the approximate expansion for the density of $\delta^*_n = \sqrt{\kappa_{n,1,1}}\hat{\delta}_n$, given in corollary 7 of LP. Thus, the same pivotal result, with the same coefficients, holds for the Gaussian MLE in the ARFIMA(0, d, 0) model with known mean and variance and the WMLE in the same model but with σ^2 and μ unknown. Furthermore, it follows from the proof of theorem 8 of LP that

$$\sup_{x\in R}\,\sup_{d\in D^*}\left|\Pr_{d_0}\left(\hat{\delta}_{W,n} \le x/\sqrt{\kappa_{W,n,1,1}}\right) - \tilde{H}^{(1),A}_{\hat{\delta}_{W,n}}(x/\sqrt{\kappa_{W,n,1,1}})\right| = o\left(n^{-1/2}\right),$$

where

$$\tilde{H}^{(1),A}_{\hat{\delta}_{W,n}}(x) = \int_{-\infty}^{x}\tilde{h}^{(1),A}_{\hat{\delta}_{W,n}}(u)\,du.$$

In other words, the distribution expansion based on the integral of the density expansion (20) is a valid asymptotic expansion.

3. EXPANSIONS FOR THE PPMLE

The Gaussian log likelihood is given by

$$L_n(\theta, \mu) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2}\log\det T_n^d - \frac{1}{2\sigma^2}(x_n - \mu 1_n)' T_n^{-1}(f_d)(x_n - \mu 1_n).$$

We reduce the dimensionality of the problem by projecting μ out and profiling the resulting *plug-in log-likelihood* with respect to σ^2. Replacing μ by $\bar{x}_n$ then, we obtain

$$L_n(d, \sigma^2; \bar{x}_n) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2}\log\det T_n^d - \frac{1}{2\sigma^2}x_n' M_n T_n^{-1}(f_d) M_n x_n.$$

We could replace μ by any $n^{\frac{1}{2}-d}$-consistent estimator but the choice of $\bar{x}_n$ is the most popular in applied work and leads to tractable results. See Dahlhaus (1989). Set $Q_n = x_n' \, M_n T_n^{-1} \, (f_d) \, M_n x_n$. The plug-in score is

$$\frac{\partial L_n(d, \sigma^2; \bar{x}_n)}{\partial d} = -\frac{1}{2}\mathrm{tr}\big(T_n^{-1}\,(f_d)\,\dot{T}_n^d\big) + \frac{1}{\sigma^2}x_n' M_n A_{n,1}^d M_n x_n,$$

with

$$A_{n,1}^d = -\frac{1}{2}\frac{\partial T_n^{-1}\,(f_d)}{\partial d},$$

and the plug-in MLE (PMLE) of σ^2 is

$$\hat{\sigma}^2 = \frac{Q_n}{n}.$$

So, the profiled-plug-in score is

$$\frac{\partial L_n(d; \hat{\sigma}^2, \bar{x}_n)}{\partial d} = -\frac{1}{2}\mathrm{tr}\big(T_n^{-1}\,(f_d)\,\dot{T}_n^d\big) + \frac{x_n' M_n A_{n,1}^d M_n x_n}{Q_n/n}.$$

The PPMLE of d is a solution to the estimating equation

$$L_{n,1}(d) = -\frac{1}{2}C_n Q_n + x_n' M_n A_{n,1}^d M_n x_n = 0, \tag{21}$$

with

$$C_n = \frac{1}{n}\mathrm{tr}\left(T_n^{-1}\,(f_d)\,\dot{T}_n^d\right).$$

Equation (21) differs from the score in the canonical case in that the latter does not involve Q_n and M_n. Moreover, (21) is not a proper score but rather it is an estimating equation. Nevertheless, it is easy to see that the general form of (4) is also valid for the estimator based on (21). For simplicity, we do not distinguish the notation between the MLE and the PPMLE. We will write

$$L_{n,j+1} = \frac{\partial^j L_{n,1}}{\partial d^j}, \quad j = 1, 2$$

$$\hat{\delta}_n = \sqrt{n}(\hat{d}_n - d_0),$$

and set

$$\begin{aligned}
Z_{n,1} &= \frac{1}{\sqrt{n}}[L_{n,1} - E(L_{n,1})], \\
Z_{n,2} &= \frac{1}{\sqrt{n}}[L_{n,2} - E(L_{n,2})], \\
Z_n &= (Z_{n,1}, Z_{n,2})', \\
\kappa_{n,1,1} &= E_d\left(Z_{n,1}^2\right), \\
\mu'_{n,3} &= \frac{1}{n}\ E_d(L_{n,3}(d)), \\
D_n &= A_{n,1}^d - \frac{1}{2}\ C_n T_n^{-1}\left(f_d\right).
\end{aligned}$$

For the exact expansion, the analogue of (4), we need the following terms:

$$\begin{aligned}
\kappa_{n,1,1} &= E_d\left(Z_{n,1}^2\right) \\
&= \frac{2}{n}\ \mathrm{tr}\left(M_n D_n M_n T_n^d\right)^2, \qquad (22)
\end{aligned}$$

$$\kappa_{n,1,2} = \frac{2}{n}\mathrm{tr}\left(M_n D_n M_n T_n^d M_n D_n' M_n T_n^d\right),$$

$$\kappa_{n,1,1,1} = \frac{8}{n}\mathrm{tr}\left(M_n D_n M_n T_n^d\right)^3,$$

$$\begin{aligned}
\mu'_{n,3} &= \frac{1}{n} E_d[L_{n,3}(d)] \\
&= \frac{1}{2n}\mathrm{tr}\left(4\left[T_n^{-1}\left(f_d\right)\dot{T}_n^d\right]^3 - 3T_n^{-1}\left(f_d\right)\dot{T}_n^d T_n^{-1}\left(f_d\right)\ddot{T}_n^d\right).
\end{aligned}$$

Substituting the last expressions into $C_{n,1}^*$ and $C_{n,3}^*$, which are of the same form as (6) and (7), we obtain the expansion (4). Simplification of the last terms is required to achieve the approximate expansions, with coefficients not depending on n. In (22),

$$\frac{2}{n}\ \mathrm{tr}(M_n D_n M_n T_n)^2 = \frac{2}{n}\ \Sigma\ \mathrm{tr}\prod_{j=1}^{2}\left\{(-P_n)^{\chi_j} D_n(-P_n)^{\xi_j} T_n\right\}, \qquad (23)$$

where χ_j, ξ_j are either zero or one and satisfy $0 \leq \sum_{j=1}^{2}(\chi_j + \xi_j) \leq 4$. The summation in (23) is over all possible $2^4 = 16$ configurations $(\chi_1, \xi_1, \chi_2, \xi_2)$ and $P_n^0 = I$. From theorem 7 of Lieberman and Phillips (2004b) and the fact that

$$\int_\Pi \frac{\dot{f}_d(\lambda)}{f_d(\lambda)} d\lambda = 0,$$

we see that

$$C_n = O\left(n^{-\frac{1}{2}+\epsilon}\right), \quad \forall\epsilon > 0.$$

The leading term in (23), corresponding to $\Sigma\,(\chi_j + \xi_j) = 0$, is therefore

$$\begin{aligned}\frac{2}{n}\mathrm{tr}(D_n T_n)^2 &= \frac{2}{n}\mathrm{tr}\left[\left(A^d_{n,1} - \frac{1}{2}C_n T_n^{-1}(f_d)\right)T_n^d\right]^2 \\ &= \frac{2}{n}\mathrm{tr}\left\{\left(\frac{1}{2}T_n^{-1}(f_d)\,\dot T_n^d\right)^2 + \frac{1}{4}\,C_n^2 I - \frac{1}{2}\left(T_n^{-1}(f_d)\,\dot T_n^d\right)C_n\right\} \\ &= O(1) + O\left(n^{-1+2\epsilon}\right) + O\left(n^{-1+2\epsilon}\right) \\ &= O(1).\end{aligned}$$

So,

$$\frac{2}{n}\,\mathrm{tr}(D_n T_n^d)^2 = \frac{1}{2n}\mathrm{tr}\big(T_n^{-1}(f_d)\,\dot T_n^d\big)^2 + O\left(n^{-1+2\epsilon}\right).$$

Now, consider the configuration (1, 0, 0, 0). This gives $\frac{2}{n}\,\mathrm{tr}\,\{P_n\big(D_n T_n^d\big)^2\}$. We have

$$D_n T_n = \frac{1}{2}T_n^{-1}(f_d)\,\dot T_n^d - \frac{1}{2}C_n I,$$

and

$$\frac{2}{n}\,\mathrm{tr}\left[P_n\big(D_n T_n^d\big)^2\right] = \frac{2}{n^2}\,1'\left[\frac{1}{2}\big(T_n^{-1}(f_d)\,\dot T_n^d\big)^2 + \frac{1}{4}C_n^2 I - \frac{1}{2}\big(T_n^{-1}(f_d)\,\dot T_n^d\big)C_n\right]1. \qquad (24)$$

Now,

$$\begin{aligned}\left|1'\left(T_n^{-1}(f_d)\,\dot T_n^d\right)^2 1\right| &= \left|1' T_n^{-1/2}(f_d)\,T_n^{-1/2}(f_d)\,\dot T_n^d T_n^{-1}(f_d)\,\dot T_n^{d1/2}\dot T_n^{d1/2}1\right| \\ &\le \sqrt{1' T_n^{-1}(f_d)\,1}\,\sqrt{1'\dot T_n^d 1}\;\left\|T_n^{-\frac{1}{2}}(f_d)\,\dot T_n^d T_n^{-1}(f_d)\,\dot T_n^{d\frac{1}{2}}\right\|.\end{aligned}$$

Using theorem 5.2 of Adenstedt (1974),

$$\sqrt{1' T_n^{-1}(f_d)\,1} \le K n^{(1-2d+\epsilon)/2}.$$

For $(1'\dot T_n^d 1)$, we use the fact that

$$\frac{\partial f_d(\lambda)}{\partial d} = C(\lambda) f_d(\lambda) = O\left(|\lambda|^{-2d-\epsilon}\right),$$

as $\lambda \to 0$. Hence,

$$\sqrt{1'\dot T_n^d 1} \le K n^{(1+2d+\epsilon)/2}.$$

Finally,

$$\begin{aligned}\left\|T_n^{-\frac{1}{2}}(f_d)\dot T_n^d T_n^{-1}(f_d)\,\dot T_n^{\frac{1}{2}}(f_d)\right\| &\le \left\|T_n^{-\frac{1}{2}}(f_d)\,\dot T_n^{\frac{1}{2}}(f_d)\right\|\left\|\dot T_n^{\frac{1}{2}}(f_d)\,T_n^{-\frac{1}{2}}(f_d)\right\|\left\|T_n^{-\frac{1}{2}}(f_d)\,\dot T_n^{\frac{1}{2}}(f_d)\right\| \\ &\le K n^{3\epsilon},\end{aligned}$$

by lemma 5.3 of Dahlhaus (1989). So, the first term in (24) is

$$\frac{2}{n^2}\left|1'\left(\frac{1}{2}\left(T_n^{-1}(f_d)\,\dot T_n^d\right)^2\right)1\right| \le K n^{-1+2\epsilon}.$$

The second term in (24) is

$$\frac{2}{n^2} 1' \left(\frac{1}{4} C_n^2 I \right) 1 = \frac{1}{2n} C_n^2 = O\left(n^{-2+2\epsilon}\right)$$

and the last term is

$$\frac{2}{n^2}\, C_n \frac{1}{2} \left| 1' T_n^{-1}\left(f_d\right) \dot{T}_n^d 1 \right| \leq K n^{-2+\epsilon}.$$

Similarly, we can show that terms involving more than one P are of smaller order of magnitude. Returning to (22), we have

$$\begin{aligned} \kappa_{n,1,1} &= \frac{2}{n} \operatorname{tr} \left[M_n \left(A_{n,1} - \frac{1}{2}\, C_n T_n^{-1}\left(f_d\right) \right) M_n T_n^d \right]^2 \\ &= \frac{1}{2n} \operatorname{tr} \left(T_n^{-1}\left(f_d\right) \dot{T}_n^d \right)^2 + O(n^{-1+2\epsilon}) \\ &= \frac{\pi^2}{6} + O(n^{-1/2+\epsilon}), \end{aligned}$$

so that the leading term is as in the canonical case. Continuing

$$\begin{aligned} \kappa_{n,1,2} &= \frac{2}{n} \operatorname{tr}(M_n D_n M_n T_n^d M_n D_n' M_n T_n^d) \\ &= \frac{2}{n} \operatorname{tr} \left(A_{n,1}^d T_n^d A_{n,1}^{d\prime} T_n^d \right) + O\left(n^{-1+2\epsilon}\right) \\ &= -6\zeta\left(3\right) + O\left(n^{-1/2+\epsilon}\right). \end{aligned}$$

The further terms needed in $C_{n,1}^*$ and $C_{n,3}^*$ are

$$\begin{aligned} \kappa_{n,1,1,1} &= \frac{8}{n} \operatorname{tr}(A_n D_n M_n T_n^d)^3 \\ &= \frac{8}{n} \operatorname{tr}(A_{n,1}^d T_n^d)^3 + O\left(n^{-1+2\epsilon}\right) \\ &= 12\zeta\left(3\right) + O\left(n^{-1/2+\epsilon}\right) \end{aligned}$$

and

$$\begin{aligned} \mu_{n,3}' &= \frac{1}{n}\, E_d[L_{n,3}(d)] \\ &= \frac{1}{2n} \operatorname{tr}(4T_n^{-1}\left(f_d\right) \dot{T}_n^d - 3T_n^{-1}\left(f_d\right) \dot{T}_n^d T_n^{-1}\left(f_d\right) \ddot{T}_n^d) + O\left(n^{-1+2\epsilon}\right) \\ &= 6\zeta\left(3\right) + O\left(n^{-1/2+\epsilon}\right). \end{aligned}$$

So, the approximate Edgeworth expansion for the PPMLE is identical to (2).

4. COMMENTS

The results here show agreement between the second-order distributions of the WMLE and the PPMLE. To highlight the higher-order difference between the two, it is sufficient to compare the terms

$$\kappa_{W,n,1,1} = \frac{1}{2n} \text{tr} \left[M_n \dot{T}^d_{W,n} M_n T^d_n \right]^2$$

and

$$\kappa_{n,1,1} = \frac{2}{n} \text{ tr} \left[M_n \left(A_{n,1} - \frac{1}{2} \, C_n T_n^{-1} \right) M_n T_n \right]^2 .$$

Note that $\kappa_{n,1,1}$ and $\kappa_{W,n,1,1}$ are the inverses of the variances of the normal leading terms of the distributions of the WMLE and PPMLE, respectively. Their common limit is $\pi^2/6$, which is the inverse of the variance of the asymptotic distribution. Their difference is given by

$$\begin{aligned} \left| \kappa_{W,n,1,1} - \kappa_{n,1,1} \right| &= \left| \frac{1}{2n} \text{tr} \left(T^d_{W,n} \dot{T}^d_n \right)^2 - \frac{1}{2n} \text{ tr} \left(T_n^{d-1} \dot{T}^d_n \right)^2 + O(n^{-1+2\epsilon}) \right| \\ &= O\left(n^{-1/2+\epsilon} \right). \end{aligned}$$

That is, the difference is dominated by the error of the Whittle approximation to T_n^{-1}, and not by the addition of the M_n matrix or the fact that the PPMLE is not a solution to a standard score. Similar analysis extends to the other terms in the exact expansion. Hence, the exact PPMLE and WMLE second-order Edgeworth expansions differ by $o(n^{-1/2})$.

The t statistic for the hypothesis $H_0 : d = d_0$ is

$$t_n \left(d_0 \right) = \pi \sqrt{\frac{n}{6}} \left(\hat{d}_n - d_0 \right).$$

The upper one-sided $100(1-\alpha)\%$ confidence interval (CI) for d_0 is defined by

$$\Delta C I_{up} \left(\hat{d}_n \right) = \left[\hat{d}_n - z_\alpha \sqrt{6} / \left(\pi \sqrt{n} \right), \infty \right], \tag{25}$$

where z_α is the $(1-\alpha)$ quantile of the normal distribution. By theorem 1(c) of Andrews and Lieberman (2002),

$$\sup_{d_0 \in D^*} \left| \Pr_{d_0} \left(d_0 \in \Delta C I_{up} \left(\hat{d}_n \right) \right) - (1-\alpha) \right| = O\left(n^{-1/2} \right),$$

giving a uniform error rate. The upper one-sided bootstrap $100(1-\alpha)\%$ CI for d_0 is defined as

$$\Delta C I^*_{up} \left(\hat{d}_n \right) = [\hat{d}_n - z^*_{t,\alpha} \sqrt{6} / \left(\pi \sqrt{n} \right), \infty),$$

where $z^*_{t,\alpha}$ is the $(1-\alpha)$ quantile of the parametric bootstrap t statistic, $\tilde{t}^*_n \left(\tilde{d}_n \right)$, and $\tilde{d}_n$ is a bootstrap-generating estimator, such as the WMLE or the PPMLE. Our second-order pivotal result implies that

$$\sup_{d_0 \in D^*} \left| \Pr_{d_0} \left(d_0 \in \Delta C I^*_{up} \left(\hat{d}_n \right) \right) - (1-\alpha) \right| = o\left(n^{-3/2} \ln(n) \right), \tag{26}$$

see theorem 2(b) and comment 5 of Andrews and Lieberman (2002). In contrast, for the non-pivotal statistics, the right hand side of (26) is only $o(n^{-1} \ln(n))$.

Finally, the results of the present paper are restricted to the case in which the mean estimate is $\bar{X}_n$. Extensions to other mean estimates, such as the GLS, are left for future research.

ACKNOWLEDGEMENT

Phillips thanks the NSF for support under Grant No. SES 0092509.

REFERENCES

Adenstedt, R. K. (1974). On large-sample estimation for the mean of a stationary random sequence. *Annals of Statistics 2*, 1095–107.

Andrews, D. W. K., O. Lieberman and V. Marmer (2002). Higher-order improvement of the parametric bootstrap for long-memory Gaussian processes. Cowles Foundation Discussion Paper No. 1378, Yale University. Accepted for publication in *Journal of Econometrics*.

Andrews, D. W. K. and O. Lieberman (2005). Valid Edgeworth expansions for the Whittle maximum likelihood estimator for stationary long-memory Gaussian time series. *Econometric Theory 21*, 710–34.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics 73*, 5–59.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Annals of Statistics 17*, 1749–66.

Durbin, J. (1980). Approximation for densities of sufficient estimators. *Biometrika 67*, 311–33.

Geweke, J. and S. Porter-Hudak (1983). The estimation and application of long-memory time series models. *Journal of Time Series Analysis 4*, 221–38.

Giraitis, L. and P. M. Robinson (2003). Edgeworth expansions for the semiparametric Whittle estimation of long memory. *Annals of Statistics 31*, 1325.

Lieberman, O. and P. C. B. Phillips (2004a). Expansions for the distribution of the maximum likelihood estimator of the fractional difference parameter. *Econometric Theory 20*, 464–84.

Lieberman, O. and P. C. B. Phillips (2004b). Error bounds and asymptotic expansions for Toeplitz product functionals of unbounded spectra. *Journal of Time Series Analysis 25*, 733–53.

Lieberman, O., J. Rousseau and D. M. Zucker (2003). Valid Edgeworth expansions for the maximum likelihood estimator of the parameter of a stationary, Gaussian, strongly dependent series. *Annals of Statistics 31*, 586–612.

Taniguchi, M. and Y. Kakizawa (2000). *Asymptotic Theory of Statistical Inference for Time Series*. New York: Springer-Verlag.