Submitted to the Annals of Statistics
arXiv: math.PR/0000000

SUPPLEMENT TO "GAUSSIAN PSEUDO-MAXIMUM LIKELIHOOD ESTIMATION OF FRACTIONAL TIME SERIES MODELS"

By Javier Hualde, Peter M. Robinson

Universidad Pública de Navarra and London School of Economics

This supplements [4] by providing a Monte Carlo study of finite sample performance, an application to two empirical time series, and proofs of the four lemmas in [4].

FINITE-SAMPLE PERFORMANCE

A Monte Carlo study was conducted in order to throw light on the performance of our estimates in small and moderate samples. We considered several versions of the FARIMA$(1, \delta_0, 0)$ model, which allows for simultaneous variation of both long- and short-range dependence. In (1.2) we have $\alpha(L) = (1 - \varphi_0 L)^{-1}$, $\beta(L) \equiv 1$, for $\varphi_0 = -0.5$, 0, 0.5, while the memory parameter values were $\delta_0 = -0.6$, -0.4, 0, 0.4, 0.6, 1, 1.5, 2, covering values either side of the stationarity and invertibility boundaries, as well as cases with one or two unit roots, and a value between these. We generated x_t, $t = 1, ..., n$, for $n = 64$, 128, 256 from (1.1), (1.2), using independent $N(0, 1)$ ε_t. We computed the estimate $\widehat{\boldsymbol{\tau}} = (\widehat{\delta}, \widehat{\varphi})'$ in (1.5) of $\boldsymbol{\tau}_0 = (\delta_0, \varphi_0)'$, using $\mathcal{T} =[-3, 3] \times [-0.999, 0.999]$ for each of the 8×3 $\boldsymbol{\tau}_0$ values. This was repeated over 5,000 independent replications, and Monte Carlo bias and standard deviation computed in each of the $8 \times 3 \times 3 = 72$ cases. From the same data sets we also computed these summary statistics for an estimate which correctly assumes the degree of integer differencing or aggregating needed to shift the process to the stationarity and invertibility region: we estimated the memory and autoregressive parameter of the appropriately integer differenced or aggregated sequence by the discrete-frequency Whittle pseudo likelihood estimate (i.e. the untapered version of the estimate in [8]) and then added to or subtracted from the former the appropriate integer, denoting the resulting estimate $\widehat{\boldsymbol{\tau}}_W = (\widehat{\delta}_W, \widehat{\varphi}_W)'$. Though $\widehat{\boldsymbol{\tau}}$ and $\widehat{\boldsymbol{\tau}}_W$ are equally asymptotically efficient, the additional information it employs leads one to expect $\widehat{\boldsymbol{\tau}}_W$ to be generally more accurate than our $\widehat{\boldsymbol{\tau}}$ in finite samples.

Monte Carlo biases of estimates of δ_0 are given in Table 1. Nearly all biases of both estimates are negative, and overall are worst when $\varphi_0 = 0$ and $n = 64$, though there is considerable improvement with increasing n.

The latter phenomenon is mostly repeated, albeit less dramatically, for the other values of φ_0, such that for $n = 256$ absolute bias tends to monotonically increase with φ_0. The relative performance of $\widehat{\delta}$ and $\widehat{\delta}_W$ also differs markedly between zero and non-zero φ_0. When $\varphi_0 = \pm 0.5$, $\widehat{\delta}$ is more (less) biased in 38 (6) cases out of 48, whereas when $\varphi_0 = 0$ the corresponding scores are 4 (19) out of 24 (though mention must be made of the relatively poor performance here of $\widehat{\delta}$ when $\delta_0 = 2$). Otherwise, biases of both estimates are fairly stable across δ_0. The overall superiority of $\widehat{\delta}_W$ here might be explained by the fact that it correctly uses the information on the unit length interval in which δ_0 lies.

TABLE 1. *Bias of estimates of* δ_0

	φ_0		.5			0			-.5	
δ_0	n	64	128	256	64	128	256	64	128	256
-.6	$\widehat{\delta}$	-.052	-.054	-.049	-.113	-.049	-.020	-.041	-.018	-.009
-.6	$\widehat{\delta}_W$	-.019	-.043	-.043	-.244	-.099	-.022	-.058	-.015	-.005
-.4	$\widehat{\delta}$	-.049	-.058	-.049	-.105	-.047	-.021	-.037	-.018	-.010
-.4	$\widehat{\delta}_W$	-.007	-.037	-.040	-.192	-.085	-.019	-.031	-.006	.000
0	$\widehat{\delta}$	-.056	-.059	-.050	-.106	-.052	-.019	-.037	-.020	-.008
0	$\widehat{\delta}_W$	-.034	-.053	-.050	-.210	-.098	-.026	-.052	-.019	-.008
.4	$\widehat{\delta}$	-.051	-.058	-.049	-.115	-.049	-.019	-.039	-.018	-.008
.4	$\widehat{\delta}_W$	-.021	-.054	-.044	-.245	-.104	-.021	-.048	-.015	-.004
.6	$\widehat{\delta}$	-.059	-.056	-.048	-.126	-.052	-.019	-.039	-.018	-.008
.6	$\widehat{\delta}_W$	-.027	-.045	-.033	-.240	-.124	-.021	-.037	-.005	.005
1	$\widehat{\delta}$	-.071	-.066	-.050	-.115	-.049	-.019	-.046	-.016	-.008
1	$\widehat{\delta}_W$	-.036	-.054	-.052	-.218	-.098	-.029	-.057	-.017	-.008
1.5	$\widehat{\delta}$	-.129	-.063	-.051	-.152	-.056	-.021	-.119	-.038	-.011
1.5	$\widehat{\delta}_W$	-.028	-.038	-.040	-.265	-.124	-.021	-.043	-.010	-.002
2	$\widehat{\delta}$	-.162	-.150	-.136	-.440	-.246	-.120	-.098	-.017	-.009
2	$\widehat{\delta}_W$	-.043	-.049	-.049	-.210	-.093	-.025	-.052	-.015	-.009

More surprising are the Monte Carlo standard deviations of estimates of δ_0, displayed in Table 2. Again, for both estimates $\varphi_0 = 0$ (overspecification) with $n = 64$ is a bad case, there is improvement with increasing n, standard deviations tend to increase with φ_0 for large n, and there is reasonable stability across δ_0. However, with the notable exception of the 9 cases when $\boldsymbol{\tau}_0 = (1.5, -0.5)'$, $\boldsymbol{\tau}_0 = (2, 0)'$, $\boldsymbol{\tau}_0 = (2, 0.5)'$ for $n \geq 128$, and $\boldsymbol{\tau}_0 = (2, -0.5)'$ for $n = 64$, $\widehat{\delta}$ is consistently the more precise, in 63 out of 72 cases.

TABLE 2. *Standard deviation of estimates of* δ_0

	φ_0	.5			0			-.5		
δ_0	n	64	128	256	64	128	256	64	128	256
-.6	$\widehat{\delta}$	.252	.203	.165	.282	.176	.099	.150	.091	.062
-.6	$\widehat{\delta}_W$	.299	.236	.189	.417	.286	.137	.246	.107	.068
-.4	$\widehat{\delta}$	.256	.209	.165	.286	.176	.100	.148	.093	.062
-.4	$\widehat{\delta}_W$	.297	.239	.188	.408	.282	.135	.217	.110	.069
0	$\widehat{\delta}$	.255	.207	.165	.284	.179	.099	.151	.092	.062
0	$\widehat{\delta}_W$	.293	.235	.186	.406	.284	.135	.219	.106	.067
.4	$\widehat{\delta}$	.257	.204	.166	.295	.177	.097	.151	.091	.062
.4	$\widehat{\delta}_W$	.303	.238	.190	.424	.294	.136	.222	.105	.067
.6	$\widehat{\delta}$	.255	.207	.165	.305	.185	.102	.152	.091	.062
.6	$\widehat{\delta}_W$	.310	.251	.198	.441	.337	.169	.233	.113	.070
1	$\widehat{\delta}$	.255	.212	.166	.286	.179	.102	.172	.091	.061
1	$\widehat{\delta}_W$	.292	.233	.187	.410	.283	.145	.222	.106	.067
1.5	$\widehat{\delta}$	.259	.208	.166	.333	.192	.103	.328	.173	.076
1.5	$\widehat{\delta}_W$	.306	.243	.192	.441	.317	.141	.230	.108	.068
2	$\widehat{\delta}$	.271	.242	.220	.433	.376	.291	.298	.097	.061
2	$\widehat{\delta}_W$	.290	.232	.187	.404	.281	.128	.222	.104	.067

TABLE 3. *Bias of estimates of* φ_0

	φ_0	.5			0			-.5		
δ_0	n	64	128	256	64	128	256	64	128	256
-.6	$\widehat{\varphi}$	-.001	.022	.031	.095	.043	.017	.039	.018	.009
-.6	$\widehat{\varphi}_W$	-.034	.007	.021	.213	.091	.022	.064	.022	.010
-.4	$\widehat{\varphi}$	-.003	.025	.030	.087	.039	.017	.036	.017	.008
-.4	$\widehat{\varphi}_W$	-.039	.004	.019	.166	.077	.017	.049	.018	.007
0	$\widehat{\varphi}$	.004	.025	.032	.088	.043	.016	.036	.017	.008
0	$\widehat{\varphi}_W$	-.016	.015	.028	.177	.084	.022	.054	.020	.009
.4	$\widehat{\varphi}$	-.003	.026	.031	.095	.043	.016	.037	.017	.008
.4	$\widehat{\varphi}_W$	-.032	.016	.023	.215	.095	.020	.054	.020	.008
.6	$\widehat{\varphi}$	.006	.022	.030	.106	.044	.017	.035	.016	.008
.6	$\widehat{\varphi}_W$	-.040	-.004	.006	.221	.122	.027	.057	.021	.008
1	$\widehat{\varphi}$	.016	.035	.031	.094	.043	.016	.042	.016	.007
1	$\widehat{\varphi}_W$	-.017	.018	.029	.181	.087	.025	.057	.020	.009
1.5	$\widehat{\varphi}$	.076	.030	.032	.133	.049	.018	.128	.039	.011
1.5	$\widehat{\varphi}_W$	-.032	-.002	.016	.240	.118	.022	.056	.021	.009
2	$\widehat{\varphi}$	.107	.115	.112	.429	.245	.118	.101	.017	.008
2	$\widehat{\varphi}_W$	-.010	.015	.025	.176	.083	.020	.054	.019	.010

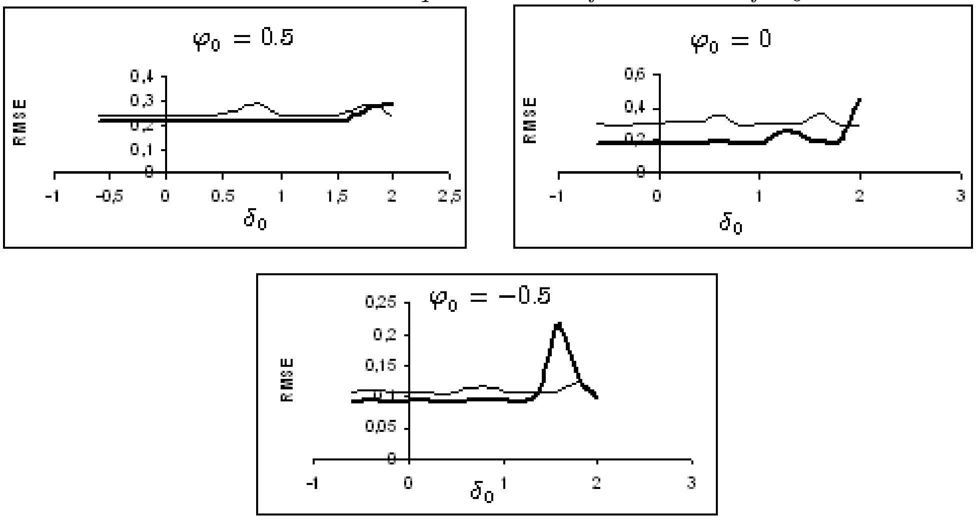
In Table 3, we compare the estimates of φ_0 in terms of bias. There are similar overall patterns to those in Table 1, but now, while $\widehat{\varphi}$ is the more biased when $\varphi_0 = 0.5$ (in 18 against 6 cases, the latter being ones when $n = 64$ and $\delta_0 \leq 1$), for both $\varphi_0 = 0$ and -0.5 $\widehat{\varphi}$ is superior (in 37 against 18 cases, the latter mostly being ones when $\boldsymbol{\tau}_0 = (1.5, -0.5)'$, $(2, 0)'$. The Monte Carlo standard deviations of estimates of φ_0, in Table 4, show a broadly similar picture to Table 2, with $\widehat{\varphi}$ clearly dominating, though for $\varphi_0 = 0$ and $n = 64$ $\widehat{\varphi}$ is even more imprecise.

TABLE 4. *Standard deviation of estimates of* φ_0

	φ_0		.5			0			-.5	
δ_0	n	64	128	256	64	128	256	64	128	256
-.6	$\widehat{\varphi}$	.255	.207	.167	.300	.197	.117	.159	.100	.068
-.6	$\widehat{\varphi}_W$	.273	.226	.185	.406	.291	.147	.230	.105	.069
-.4	$\widehat{\varphi}$	.258	.211	.166	.303	.195	.118	.156	.098	.067
-.4	$\widehat{\varphi}_W$	.275	.228	.184	.389	.283	.145	.198	.103	.067
0	$\widehat{\varphi}$	.257	.208	.166	.300	.200	.117	.154	.097	.066
0	$\widehat{\varphi}_W$	.268	.224	.181	.387	.285	.146	.202	.102	.068
.4	$\widehat{\varphi}$	.260	.208	.167	.311	.199	.115	.156	.098	.065
.4	$\widehat{\varphi}_W$	.280	.227	.185	.413	.297	.146	.205	.102	.066
.6	$\widehat{\varphi}$	.257	.209	.167	.319	.206	.121	.156	.097	.068
.6	$\widehat{\varphi}_W$	.288	.243	.195	.416	.333	.176	.219	.106	.068
1	$\widehat{\varphi}$	.255	.210	.166	.304	.199	.120	.177	.096	.065
1	$\widehat{\varphi}_W$	.268	.223	.181	.389	.285	.154	.207	.101	.066
1.5	$\widehat{\varphi}$	.247	.209	.166	.348	.212	.120	.382	.201	.086
1.5	$\widehat{\varphi}_W$	.280	.234	.187	.424	.326	.152	.212	.104	.067
2	$\widehat{\varphi}$	.238	.214	.191	.468	.411	.311	.346	.104	.066
2	$\widehat{\varphi}_W$	.267	.222	.183	.387	.282	.140	.204	.098	.068

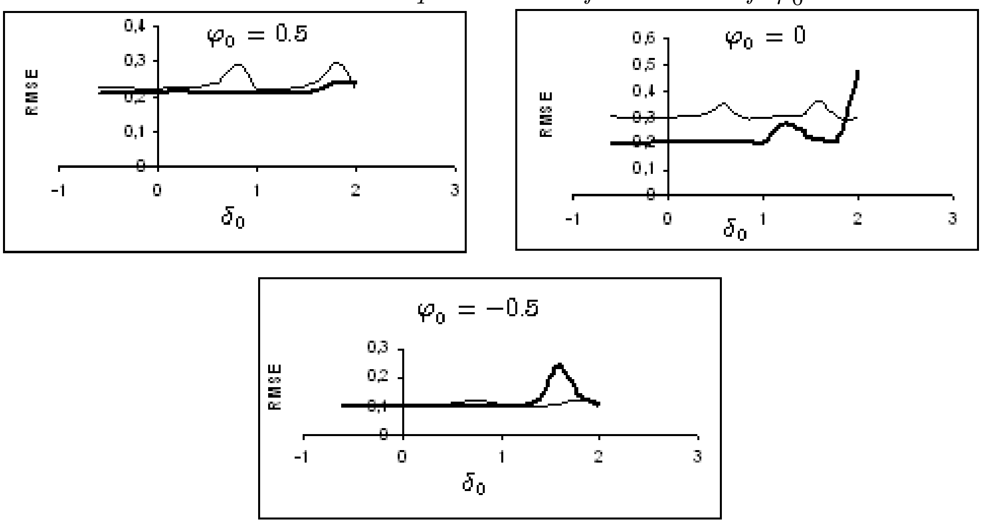
With the aim of providing a clearer picture of the pattern of estimates with respect to variations in $\boldsymbol{\tau}_0$, we plot in Figure 1 the Monte Carlo root mean square error of $\widehat{\delta}$, $\widehat{\delta}_W$, for $n = 128$, as a function of 15 values of δ_0 (those in the initial choice plus -0.2, 0.2, 0.8, 1.2, 1.4, 1.6, 1.8) for $\varphi_0 = -0.5, 0, 0.5$. In each of the plots, the thick line corresponds to results for $\widehat{\delta}$, whereas the thin one records results for $\widehat{\delta}_W$. As anticipated, the best results are for $\varphi_0 = -0.5$, $\widehat{\delta}$ being superior to $\widehat{\delta}_W$, except in the region between $\delta_0 = 1.4$ and 2. For $\widehat{\delta}_W$, results are worst for $\varphi_0 = 0$, when $\widehat{\delta}$ clearly dominates (except when $\delta_0 = 2$), whereas when $\varphi_0 = 0.5$, both estimates behave similarly, although $\widehat{\delta}$ is slightly superior overall.

FIGURE 1. *Root Mean Square Error of estimates of* δ_0



Finally, in Figure 2 we plot corresponding results for $\widehat{\varphi}$, $\widehat{\varphi}_W$ (recorded in the thick and thin lines, respectively). The pattern is in all cases almost identical to that in Figure 1.

FIGURE 2. *Root Mean Square Error of estimates of* φ_0



EMPIRICAL EXAMPLE

We now report an empirical application to US quarterly income and consumption data 1947Q1-1981Q2 ($n = 138$), which was previously analyzed by [3], for example. By means of traditional testing procedures [3] found evidence of a unit root in both series, and the semiparametric fractional approach of [5] tended to support this conclusion. Our analysis did not. We determined $\theta(s;\boldsymbol{\varphi})$ from the data, our approach permitting comparison among competing parametric models. This was achieved by first obtaining a preliminary estimate of δ_0, which was used to filter the series to have, approximately, short memory, and then employing the model choice procedure of [1] to select p_1 and p_2. For this purpose we cannot use a $\sqrt{n}-$consistent parametric estimate of δ_0 (for example, one based on a FARIMA$(0,\ \delta_0, 0)$) because under-specification of p_1 or p_2, or over-specification of both, results in inconsistent estimation of δ_0. Instead, we employed a semiparametric estimate of δ_0, which converges more slowly but does not require short memory specification and is thus more robust. In addition, we examine the issue of truncation, which is inherent to model (1.1), and arises because the model reflects the data start-time: given a sample x_t, $t = 1, ..., n$, the first observation of the filtered sequence $\Delta^d \{x_t \mathbb{1}(t > 0)\}$ equals the unfiltered x_1, the second is a linear combination of x_1, x_2, and so on. We check stability with respect to omitting from the analysis l initial observations of the filtered series.

We look first at the income series. We computed the local Whittle or semiparametric Gaussian estimate (see e.g. [6]) on first-differenced observations Δx_t, followed by adding back 1 (an alternative semiparametric estimate, which is valid also under nonstationarity, and thus avoids the initial first-differencing, was proposed and justified by [7]). In order to reflect possible sensitivity to choice of bandwidth m (the number of low Fourier frequencies employed) and because the choice of m only indirectly affects the final outcome, rather than employing an optimal, data-dependent m, we tried three different values, $m = 8$, 17, 34, obtaining estimates $\widetilde{\delta} = 1.107$, 1.017, 1.084, respectively. Using these $\widetilde{\delta}$, the filtered $\Delta^{\widetilde{\delta}} \{x_t \mathbb{1}(t > 0)\}$ were generated, and from their simple and partial correlograms we identified in the spirit of [1] the parametric model $\theta(s;\boldsymbol{\varphi})$. For the various estimates of δ_0, the methodology suggested that $\theta(s;\boldsymbol{\varphi}_0) = \left(1 - \varphi_0 s^{10}\right)^{-1}$ might be adequate. We report our estimates of δ_0, φ_0 in Table 5, along with t-ratios (denoted by t_δ, t_φ) corresponding to the null hypotheses $H_0 : \delta_0 = 1$, $H_0 : \varphi_0 = 0$, where denominators are corresponding elements of the 2-dimensional square matrix $\sum_{t=10+l}^{n} \partial\varepsilon_t(\widehat{\boldsymbol{\tau}})/\partial\boldsymbol{\tau}\,(\partial\varepsilon_t(\widehat{\boldsymbol{\tau}})/\partial\boldsymbol{\tau})' \,/ \sum_{t=10+l}^{n} \varepsilon_t^2(\widehat{\boldsymbol{\tau}})$, where $\widehat{\boldsymbol{\tau}} = (\widehat{\delta}, \widehat{\varphi})'$. For $l > 2$, the corresponding null hypotheses are in all cases rejected at 1% significance level, thus casting doubt on the unit root hypothesis.

TABLE 5. *Parameter estimates for the income series*

l	1	2	3	4	5	6	7	8	9	10
$\widehat{\delta}$	1.12	1.14	1.15	1.15	1.15	1.15	1.15	1.15	1.15	1.16
t_δ	2.29	2.54	2.62	2.62	2.62	2.67	2.66	2.64	2.66	2.68
$\widehat{\varphi}$	.204	.257	.236	.235	.235	.233	.247	.249	.245	.242
t_φ	2.55	3.43	3.02	3.01	3.01	3.01	3.18	3.13	3.07	3.03

For the consumption series results are provided in Table 6. As before, we computed three different $\widetilde{\delta} = 0.855,\ 0.976,\ 1.127$, for $m = 8, 17, 34$, respectively. We again identified $\theta(s; \boldsymbol{\varphi})$ based on the corresponding residuals, but now the greater variation of the $\widetilde{\delta}$, leads to two different specifications, namely $\theta(s; \boldsymbol{\varphi}_0) = (1 - \varphi_0^{(1)} s)^{-1}$ (suggested by $\widetilde{\delta} = 0.855$) and $\theta(s; \boldsymbol{\varphi}_0) = (1 - \varphi_0^{(8)} s^8)^{-1}$ (suggested by $\widetilde{\delta} = 0.976, 1.127$). Given the discrepancy, we let the two short run models compete in our parametric specification, setting $\theta(s; \boldsymbol{\varphi}_0) = (1 - \varphi_0^{(1)} s - \varphi_0^{(8)} s^8)^{-1}$, obtaining parametric estimates $\widehat{\delta}$, $\widehat{\varphi}^{(1)}$, $\widehat{\varphi}^{(8)}$. As before, t-ratios for identical null hypotheses are provided, supporting clearly the specification with $\varphi_0^{(1)} = 0$, a unit root being again strongly rejected.

TABLE 6. *Parameter estimates for the consumption series*

l	1	2	3	4	5	6	7	8	9	10
$\widehat{\delta}$	1.07	1.10	1.11	1.11	1.12	1.12	1.15	1.14	1.15	1.15
t_δ	2.33	2.74	2.78	2.76	2.70	2.63	2.79	2.65	2.72	2.63
$\widehat{\varphi}^{(1)}$	-.016	-.054	-.068	-.072	-.074	-.075	-.092	-.064	-.054	-.041
$t_{\varphi^{(1)}}$	-.178	-.600	-.750	-.785	-.807	-.809	-1.01	-.674	-.570	-.423
$\widehat{\varphi}^{(8)}$	-.164	-.196	-.213	-.220	-.221	-.223	-.225	-.220	-.240	-.233
$t_{\varphi^{(8)}}$	-2.06	-2.50	-2.62	-2.67	-2.68	-2.64	-2.77	-2.70	-3.01	-2.89

PROOFS OF LEMMAS IN SECTION 5

PROOF OF LEMMA 1. Clearly,

$$c_j(\boldsymbol{\tau}) = \sum_{k=0}^{j} \phi_k(\boldsymbol{\varphi})\, a_{j-k}, \tag{0.1}$$

writing $a_j = a_j(\delta_0 - \delta)$, so that for any $\delta \in \mathcal{I}$, by Stirling's approximation

$$\begin{aligned}\sup_{\boldsymbol{\varphi}\in\Psi}|c_j(\boldsymbol{\tau})| &\leq K\sum_{k=0}^{j-1}(j-k)^{\delta_0-\delta-1}\sup_{\boldsymbol{\varphi}\in\Psi}|\phi_k(\boldsymbol{\varphi})| \\ &\leq K\sum_{k=0}^{[j/2]}(j-k)^{\delta_0-\delta-1}\sup_{\boldsymbol{\varphi}\in\Psi}|\phi_k(\boldsymbol{\varphi})| \\ &\quad +K\sum_{k=[j/2]}^{j-1}(j-k)^{\delta_0-\delta-1}\sup_{\boldsymbol{\varphi}\in\Psi}|\phi_k(\boldsymbol{\varphi})|. \qquad (0.2)\end{aligned}$$

(0.2) is bounded by

$$Kj^{\delta_0-\delta-1}\sum_{k=1}^{\infty}k^{-1-\varsigma}+Kj^{-1-\varsigma}\sum_{k=[j/2]}^{j-1}(j-k)^{\delta_0-\delta-1}=O\left(j^{\max(\delta_0-\delta-1,-1-\varsigma)}\right),$$

because $\varsigma > 1/2$ and the second sum is $O\left(j^{\delta_0-\delta}\right)$ if $\delta < \delta_0$, $O(\log j)$ if $\delta = \delta_0$, and $O(1)$ if $\delta > \delta_0$. The proof of (5.2) is almost identical on noting

$$c_{j+1}-c_j=\phi_{j+1}(\boldsymbol{\varphi})+\sum_{k=1}^{j}\phi_k(\boldsymbol{\varphi})(a_{j+1-k}-a_{j-k}),\ a_{j+1}-a_j=O\left(j^{\delta_0-\delta-2}\right).$$

PROOF OF LEMMA 2. From (5.1), (0.1)

$$\varepsilon_t(\boldsymbol{\tau}^*)=\sum_{j=0}^{t-1}a_j\sum_{k=0}^{t-j-1}\phi_k(\boldsymbol{\varphi}_0)u_{t-j-k}=\sum_{j=0}^{t-1}a_j\varepsilon_{t-j}+v_t(\delta),$$

where

$$v_t(\delta)=-\sum_{j=0}^{t-1}a_j\sum_{k=t-j}^{\infty}\phi_k(\boldsymbol{\varphi}_0)u_{t-j-k}. \qquad (0.3)$$

Thus

$$\sup_{\delta_0-\kappa\leq\delta<\delta_0-\frac{1}{2}+\eta}|v_t(\delta)|\leq K\sum_{j=1}^{t}j^{\kappa-1}\left|\sum_{k=t-j}^{\infty}\phi_k(\boldsymbol{\varphi}_0)u_{t-j-k}\right|.$$

Now

$$Var\left(\sum_{k=t-j}^{\infty}\phi_k(\boldsymbol{\varphi}_0)u_{t-j-k}\right)\leq K\sum_{k=t-j}^{\infty}\phi_k^2(\boldsymbol{\varphi}_0)\leq K(t-j)^{-1-2\varsigma}.$$

Thus

$$\sup_{\delta_0-\kappa\leq\delta<\delta_0-\frac{1}{2}+\eta}|v_t(\delta)| = O_p\left(\sum_{j=1}^{t-1} j^{\kappa-1}(t-j)^{-1/2-\varsigma}\right) = O_p\left(t^{\kappa-1}\right),$$

as in the proof of Lemma 1, noting that $1+\varsigma>3/2$. Finally, by (0.3)

$$v_t(\delta_0) = -\sum_{k=t}^{\infty}\phi_k(\boldsymbol{\varphi}_0)\,u_{t-j-k} = O_p\left(t^{-1/2-\varsigma}\right),$$

by previous arguments.

PROOF OF LEMMA 3. Since $\varepsilon_t(\boldsymbol{\tau}) = \xi(L;\boldsymbol{\varphi})\,\varepsilon_t(\boldsymbol{\tau}^*)$, following similar steps as in [2] (p.346),

$$w_{\varepsilon(\boldsymbol{\tau})}(\lambda) = \xi_{n-1}\left(e^{i\lambda};\boldsymbol{\varphi}\right) w_{\varepsilon(\boldsymbol{\tau}^*)}(\lambda) + U_n(\lambda;\boldsymbol{\tau}),$$

where $\xi_{n-1}(z;\boldsymbol{\varphi}) = \sum_{j=0}^{n-1}\xi_j(\boldsymbol{\varphi})\,z^j$ and

$$U_n(\lambda;\boldsymbol{\tau}) = -n^{-\frac{1}{2}}\sum_{k=1}^{n-1}\xi_k(\boldsymbol{\varphi})\,e^{ik\lambda}\sum_{t=n-k+1}^{n}\varepsilon_t(\boldsymbol{\tau}^*)\,e^{it\lambda},$$

so that (5.3) holds with

$$\begin{aligned} V_n(\boldsymbol{\tau}) \;=\; & \sum_{j=1}^{n}\left(\left|\xi_{n-1}\left(e^{i\lambda_j};\boldsymbol{\varphi}\right)\right|^2 - \left|\xi\left(e^{i\lambda_j};\boldsymbol{\varphi}\right)\right|^2\right) I_{\varepsilon(\boldsymbol{\tau}^*)}(\lambda_j) + \sum_{j=1}^{n}|U_n(\lambda_j;\boldsymbol{\tau})|^2 \\ & +2\,\mathrm{Re}\left\{\sum_{j=1}^{n}\xi_{n-1}\left(e^{i\lambda_j};\boldsymbol{\varphi}\right) w_{\varepsilon(\boldsymbol{\tau}^*)}(\lambda_j)\,U_n(-\lambda_j;\boldsymbol{\tau})\right\}. \end{aligned} \tag{0.4}$$

The third term of (0.4) is

$$\begin{aligned} & -\frac{2}{n}\sum_{k=0}^{n-1}\sum_{t=1}^{n}\sum_{l=1}^{n-1}\sum_{s=n-l+1}^{n}\xi_k(\boldsymbol{\varphi})\,\xi_l(\boldsymbol{\varphi})\,\varepsilon_t(\boldsymbol{\tau}^*)\,\varepsilon_s(\boldsymbol{\tau}^*)\sum_{j=1}^{n}e^{i(k+t-l-s)\lambda_j} \\ = \; & -2\sum_{k=1}^{n-1}\sum_{l=1}^{n-1}\xi_k(\boldsymbol{\varphi})\,\xi_l(\boldsymbol{\varphi})\sum_{s=n-l+1}^{n+\min(k-l,0)}\varepsilon_{s+l-k}(\boldsymbol{\tau}^*)\,\varepsilon_s(\boldsymbol{\tau}^*), \end{aligned}$$

where by Lemma 2

$$\varepsilon_s(\boldsymbol{\tau}^*) = \sum_{j=0}^{s-1} a_j\varepsilon_{s-j} + v_s(\delta). \tag{0.5}$$

By summation by parts, for $s \geq 2$, the first term on the right of (0.5) is

$$a_{s-1}\sum_{j=0}^{s-1}\varepsilon_{s-j}-\sum_{j=0}^{s-2}\left(a_{j+1}-a_j\right)\sum_{k=0}^{j}\varepsilon_{s-k},$$

so that

$$\begin{aligned} E\sup_{\delta_0-\kappa\leq\delta<\delta_0-\frac{1}{2}+\eta}\left|\sum_{j=0}^{s-1}a_j\varepsilon_{s-j}\right| &\leq \sup_{\delta_0-\kappa\leq\delta<\delta_0-\frac{1}{2}+\eta}\left|a_{s-1}\right|E\left|\sum_{j=0}^{s-1}\varepsilon_{s-j}\right| \\ &+\sum_{j=0}^{s-2}\sup_{\delta_0-\kappa\leq\delta<\delta_0-\frac{1}{2}+\eta}\left|a_{j+1}-a_j\right|E\left|\sum_{k=0}^{j}\varepsilon_{s-k}\right|. \end{aligned}$$

It can be readily shown that $Var\left(\sum_{j=0}^{s-1}\varepsilon_{s-j}\right)=O\left(s\right)$, whereas, uniformly in j, $Var\left(\sum_{k=0}^{j}\varepsilon_{s-k}\right)=O\left(j\right)$, so that

$$\begin{aligned} E\sup_{\delta_0-\kappa\leq\delta<\delta_0-\frac{1}{2}+\eta}\left|\sum_{j=0}^{s-1}a_j\varepsilon_{s-j}\right| &\leq K\sup_{\delta_0-\kappa\leq\delta<\delta_0-\frac{1}{2}+\eta}\left(s^{\delta-\frac{1}{2}}+\sum_{j=1}^{s-2}j^{\delta-\frac{3}{2}}\right) \\ &\leq K\left(\log s\mathbb{1}\left(\kappa=1/2\right)+s^{\kappa-1/2}\mathbb{1}\left(\kappa>1/2\right)\right), \qquad (0.6) \end{aligned}$$

whereas by Lemma 2

$$E\sup_{\delta_0-\kappa\leq\delta<\delta_0-\frac{1}{2}+\eta}\left|v_s\left(\delta\right)\right|\leq Ks^{\kappa-1}.$$

Then since

$$\begin{aligned} &E\sup_{\delta_0-\kappa\leq\delta<\delta_0-\frac{1}{2}+\eta}\left|\sum_{s=n-l+1}^{n+\min(k-l,0)}\varepsilon_{s+l-k}\left(\boldsymbol{\tau}^*\right)\varepsilon_s\left(\boldsymbol{\tau}^*\right)\right| \\ \leq\ & Kl\left(\log n\mathbb{1}\left(\kappa=1/2\right)+n^{\kappa-1/2}\mathbb{1}\left(\kappa>1/2\right)\right)^2+Kl, \end{aligned}$$

we have

$$\begin{aligned} &E\sup_{\substack{\delta_0-\kappa\leq\delta<\delta_0-\frac{1}{2}+\eta \\ \boldsymbol{\varphi}\in\Psi}}\left|\sum_{k=1}^{n-1}\sum_{l=1}^{n-1}\xi_k\left(\boldsymbol{\varphi}\right)\xi_l\left(\boldsymbol{\varphi}\right)\sum_{s=n-l+1}^{n+\min(k-l,0)}\varepsilon_{s+l-k}\left(\boldsymbol{\tau}^*\right)\varepsilon_s\left(\boldsymbol{\tau}^*\right)\right| \\ \leq\ & K\left(\log^2 n\mathbb{1}\left(\kappa=1/2\right)+n^{2\kappa-1}\mathbb{1}\left(\kappa>1/2\right)\right)\sup_{\boldsymbol{\varphi}\in\Psi}\sum_{k=0}^{\infty}\left|\xi_k\left(\boldsymbol{\varphi}\right)\right|\sum_{l=0}^{\infty}l\left|\xi_l\left(\boldsymbol{\varphi}\right)\right| \\ \leq\ & K\left(\log^2 n\mathbb{1}\left(\kappa=1/2\right)+n^{2\kappa-1}\mathbb{1}\left(\kappa>1/2\right)\right). \end{aligned}$$

Following similar steps to previous ones, it is immediate to show that

$$\sup_{\substack{\delta_0-\kappa\leq\delta<\delta_0-\frac{1}{2}+\eta\\ \boldsymbol{\varphi}\in\Psi}}\sum_{j=1}^{n}\left|U_n\left(\lambda_j;\boldsymbol{\tau}\right)\right|^2=O_p\left(\log^2 n\mathbb{1}\left(\kappa=1/2\right)+n^{2\kappa-1}\mathbb{1}\left(\kappa>1/2\right)\right).$$

Finally

$$\sup_{\substack{\delta_0-\kappa\leq\delta<\delta_0-\frac{1}{2}+\eta\\ \boldsymbol{\varphi}\in\Psi}}\left|\sum_{j=1}^{n}\left(\left|\xi_{n-1}\left(e^{i\lambda_j}\right)\right|^2-\left|\xi\left(e^{i\lambda_j}\right)\right|^2\right)I_{\varepsilon(\boldsymbol{\tau}^*)}\left(\lambda_j\right)\right|$$

$$\leq\sup_{\substack{\lambda\in[-\pi,\pi]\\ \boldsymbol{\varphi}\in\Psi}}\left|\left|\xi_{n-1}\left(e^{i\lambda}\right)\right|^2-\left|\xi\left(e^{i\lambda}\right)\right|^2\right|\sup_{\delta_0-\kappa\leq\delta<\delta_0-\frac{1}{2}+\eta}\sum_{t=1}^{n}\varepsilon_t^2\left(\boldsymbol{\tau}^*\right).$$

By previous results

$$\sup_{\delta_0-\kappa\leq\delta<\delta_0-\frac{1}{2}+\eta}\sum_{t=1}^{n}\varepsilon_t^2\left(\boldsymbol{\tau}^*\right)=O_p\left(n\log^2 n\mathbb{1}\left(\kappa=1/2\right)+n^{2\kappa}\mathbb{1}\left(\kappa>1/2\right)\right),$$

and noting that

$$\sup_{\substack{\lambda\in[-\pi,\pi]\\ \boldsymbol{\varphi}\in\Psi}}\left|\left|\xi_{n-1}\left(e^{i\lambda};\boldsymbol{\varphi}\right)\right|^2-\left|\xi\left(e^{i\lambda};\boldsymbol{\varphi}\right)\right|^2\right|=O\left(n^{-\varsigma}\right)=o\left(1\right),$$

the first term on the right of (0.4) is of smaller order, to conclude the proof.

PROOF OF LEMMA 4. The proof is very similar to that of Lemma 1. The only point worth mentioning is the calculation of the order of magnitude of $\partial\left(a_{j+1}\left(c\right)-a_j\left(c\right)\right)/\partial c$ and $\partial^2\left(a_{j+1}\left(c\right)-a_j\left(c\right)\right)/\partial c^2$. First, $\partial\left(a_{j+1}\left(c\right)-a_j\left(c\right)\right)/\partial c$ is

$$\psi\left(j+1+c\right)a_{j+1}\left(c\right)-\psi\left(j+c\right)a_j\left(c\right)-\psi\left(c\right)\left(a_{j+1}\left(c\right)-a_j\left(c\right)\right).\tag{0.7}$$

The third term in (0.7) is $O\left(j^{c-2}\right)$, whereas since

$$\left|\psi\left(j+1+c\right)-\psi\left(j+c\right)\right|\leq K\left|\psi'\left(j+c\right)\right|\leq K\left(j+1\right)^{-1},$$

and

$$\left|\psi\left(j+c\right)\right|\leq K\log\left(j+1\right),$$

then

$$\left(a_{j+1}\left(c\right)-a_j\left(c\right)\right)\psi\left(j+1+c\right)+a_j\left(c\right)\left(\psi\left(j+1+c\right)-\psi\left(j+c\right)\right)$$

is $O\left(j^{c-2}\log j\right)$. Thus (0.7) is $O\left(j^{c-2}\log j\right)$. Second, it can be shown that

$$\begin{aligned} & \frac{\partial^2}{\partial c^2}\left(a_{j+1}\left(c\right)-a_j\left(c\right)\right) \\ = \;& \psi\left(j+1+c\right)\frac{\partial a_{j+1}\left(c\right)}{\partial c}-\psi\left(j+c\right)\frac{\partial a_j\left(c\right)}{\partial c}+o\left(j^{c-2}\log^2 j\right) \\ = \;& O\left(j^{c-2}\log^2 j\right), \end{aligned}$$

by similar steps to those in the treatment of (0.7).

References.

[1] Box, G.E.P. and Jenkins, G.M. (1971). *Time series analysis. Forecasting and control.* Holden-Day, San Francisco.

[2] Brockwell, P.J. and Davis, R.A. (1990). *Time series: theory and methods.* Springer-Verlag, New York.

[3] Engle, R.F. and Granger, C.W.J. (1987). Cointegration and error correction: representation, estimation and testing. *Econometrica* **55**, 251-276.

[4] Hualde, J. and Robinson P.M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models.

[5] Marinucci, D. and Robinson P.M. (2001). Semiparametric fractional cointegration analysis. *Journal of Econometrics* **105**, 225-247.

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

[7] Shimotsu, K. and Phillips, P.C.B. (2005). Exact local Whittle estimation of fractional integration. *Annals of Statistics* **33**, 1890-1933.

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

J. Hualde
Departamento de Economía
Universidad Pública de Navarra
Campus Arrosadía
31006 Pamplona
Spain
E-mail: javier.hualde@unavarra.es

P.M. Robinson
Department of Economics
London School of Economics
Houghton Street
London WC2A 2AE
United Kingdom
E-mail: P.M.Robinson@lse.ac.uk