This article has been published in a revised form in **Econometric Theory** 36(6) : 1064-1098 (2020), https://doi.org/10.1017/S0266466619000422 . This version is free to view and download for private research and study only.

Exact Local Whittle estimation in long memory time series with multiple poles

Josu Arteche [1]

Dept. of Econometrics and Statistics

University of the Basque Country UPV/EHU

Bilbao 48015

Spain

email: josu.arteche@ehu.eus

Revised: 12th November 2019

[1]Research supported by the Spanish Ministry of Science and Innovation and ERDF grant ECO2016-76884-P and UPV/EHU Econometrics Research Group, Basque Government grant IT1359-19. The author thanks the co-editor, three anonymous referees, Javier García-Enríquez, Rajendra Bhansali, Carlos Velasco, Javier Hualde and Tomás del Barrio for useful constructive comments.

Abstract

A generalization of the Exact Local Whittle estimator in Shimotsu and Phillips (2005) is proposed for jointly estimating all the memory parameters in general long memory time series that possibly display standard, seasonal and/or other cyclical strong persistence. Consistency and asymptotic normality are proven for stationary, non-stationary and non-invertible series, permitting straightforward standard inference of interesting hypotheses such as the existence of unit roots and equality of memory parameters at some or all seasonal frequencies, which can be used as a prior test for the application of seasonal differencing filters. The effects of unknown deterministic terms are also discussed. Finally, the finite sample performance is analysed in an extensive Monte Carlo exercise and an application to an U.S. Industrial Production index.

Running head: Multiple Exact Local Whittle Estimation

1 Introduction

Strong persistence is a common characteristic in many economic series and a great many papers have been devoted to analysing it, with special attention being paid to stochastic trends entailing spectral divergences at frequency zero (standard long memory). However, much less attention has been paid to the analysis of strong persistent cycles frequently found in economics, which generate spectral divergences at non-zero frequencies. In fact the *typical spectral shape of (seasonally unadjusted) economic variables* revealed in the seminal paper by Granger (1966) often show spectral poles not only at the origin but also at seasonal frequencies $2\pi h/S$, $h = 1, 2, .., \lfloor S \rfloor$ where S is the number of observations per basic period of time ($S = 12$ for monthly series, $S = 4$ for quarterly, etc.) and $\lfloor \cdot \rfloor$ denotes "the integer part of". More recently, Abadir et al. (2013) show that the autocorrelation function of macroeconomic series have a hyperbolic and sinusoidal decay that arises from the dynamics of a general equilibrium model with heterogeneous firms, which is consistent with the presence of spectral poles at non-null frequencies.

Conventionally, the most popular tool for modelling this behaviour is the seasonal difference operator $(1 - L^S)$ (see for example the popular Box-Jenkins airline model), which implies unit roots at the origin and at every seasonal frequency. These restrictions are relaxed in Hosking (1984), Porter-Hudak (1990) and Ray (1993) by using the seasonal fractional difference operator $(1-L^S)^d$, where the memory parameter d can be any real number, though the same degree of persistence is still imposed at every seasonal frequency and at the origin. This constraint is further relaxed in the fractional ARUMA model proposed by Giraitis and Leipus (1995) defined as $\prod_{h=1}^{H}(1 - 2\cos w_h L + L^2)^{d_h} X_t = u_t$ where u_t is a stationary and invertible ARMA process and the different w_h are frequencies in $[0, \pi]$. See also Chan and Wei (1988), Robinson (1994), Chan and Terrin (1995), Woodward et al. (1998) and Nielsen (2004).

The existence of several different memory parameters is thus a possibility to be considered. Equality of all memory parameters at the origin and seasonal frequencies should therefore be checked before using the seasonal fractional difference operator, and the equality of all of them to 1 should be tested before the traditional seasonal difference operator

$(1 - L^S)$ is applied. Statistical inference on these situations has generated a large body of research and several techniques have been proposed for testing for fractional seasonal roots, most of them based on Lagrange multiplier statistics that avoid the need to estimate the memory parameters. See for example Robinson (1994), Arteche (2002), Nielsen (2004) and Hassler et al. (2009), and the references therein. Estimation of memory parameters has however drawn less attention and efforts have been concentrated mainly on the standard long memory case at frequency zero. In that context, Whittle estimation has proven to be a valuable strategy in either its parametric (from Fox and Taqqu, 1986 to Shao, 2010) or semiparametric versions (see Robinson, 1995, Velasco, 1999, Shimotsu and Phillips, 2005, Shao and Wu, 2007, Abadir et al., 2007 and Shimotsu, 2010 among others). The latter make use of the fact that the spectral (pseudospectral in the nonstationary case) density function satisfies $f(\lambda) \sim C|\lambda|^{-2d}$ as $\lambda \to 0$, for C a positive constant, and thus only frequencies around zero are used to estimate d, which avoids the need to specify short memory components. All these papers have developed an extensive asymptotic theory on Whittle estimation in standard long memory series, covering non-invertible, invertible, stationary and non stationary series. However, the case of seasonal or cyclical long memory has been less analysed. The few relevant papers include Giraitis and Leipus (1995), who prove the consistency of the parametric Whittle estimator in stationary ARUMA processes with positive memory parameters (but give no results on the asymptotic distribution) and Arteche (2000) and Arteche and Robinson (2000), who extend the local Whittle criterion to estimate the memory parameter around a single stationary and invertible spectral pole, proving its consistency and asymptotic normality. Values of $d \notin (-1/2, 1/2)$ are however not covered by either of these proposals.

This paper complements those mentioned above by proposing a general semiparametric estimation technique for a finite number of memory parameters with different locations, allowing d to take any real value and thus covering not only stationary and invertible processes but also non-stationary and non-invertible ones. The novelty of the proposal is thus of interest for two main reasons: first, it permits joint estimation of all the memory parameters corresponding to persistent trends, cycles or seasonality in a local Whittle context, with no need to consider or specify any short memory component. Second, it permits estimation of non-stationary and non-invertible cycles, which has not been considered before, unlike the estimation at frequency zero. As a consequence, the proposed estimation technique permits simple implementation of Wald type tests of interesting hypothesis such as stationarity (memory parameters lower than 1/2) or unit roots at certain cyclical or seasonal frequencies.

The proposed estimator is an extension of the Exact Local Whittle estimator of Shimotsu and Phillips (2005) to cover long memory at non-zero frequencies, which requires three new challenges to be met with respect to the standard long memory case: first, spectral symmetry does not need to hold around a frequency $w \in (0, \pi)$, giving rise to some terms that entail the use of frequencies closer to w to avoid biasing effects. In other words, the bandwidth should increase at a lower rate with the sample size, so the optimal convergence rate is slower. This negative effect is however offset in finite samples by the use of frequencies on both sides of the location of the pole, thus taking advantage of the lack of symmetry of the periodogram. Second, for $w \in (0, \pi)$, $f(w + \lambda) \sim C(d)|\lambda|^{-2d}$ as $\lambda \to 0$, where $C(d)$ may depend not only on w but also on d such that it should be considered in the loss function for estimation of d. Otherwise a significant bias arises. Finally, the extension to positive frequencies opens up the possibility of multiple spectral poles (as with seasonal strong persistence) and a joint estimation of several memory parameters is called for.

The rest of the paper is organised as follows. Section 2 describes the kind of processes dealt with. They allow for spectral poles at any frequency between 0 and π inclusive. Section 3 introduces the estimator proposed for locally estimating all the memory parameters in those processes. Consistency and asymptotic normality are shown. Section 4 discusses the effects of unknown deterministic cycles, including an unknown mean as a particular case. Section 5 introduces an extensive Monte Carlo analysis showing the competitive finite sample performance and its applicability for unit roots testing. Section 6 shows an application to a U.S. Industrial Production Index. To save space, the proofs of the theorems and all the lemmas required for those proofs are relegated to the supplementary material, together with further results of the Monte Carlo analysis.

2 Multiple Generalised Fractionally Integrated Processes

We consider processes of the form

$$\Delta^H(L,d)X_t = u_t I(t \geq 1), \quad t = 0, 1, ..., \tag{1}$$

where $I(\cdot)$ is the indicator function, $d = (d_1, ..., d_H)'$, $\Delta^H(L,d) = \prod_{h=1}^H \Delta_h(L, d_h)$ for $\Delta_h(L, d_h) = (1 - 2\cos w_h L + L^2)^{\delta_h d_h}$ with $\delta_h = 0.5$ if $w_h = 0, \pi$ and $\delta_h = 1$ in any other case and u_t is stationary with zero mean and spectral density $f_u(w_h + \lambda) \sim G_h$ as $\lambda \to 0$ for $h = 1, 2, ..., H$. Note that $\Delta_h(L, d_h) = (1 - 2\cos w_h L + L^2)^{d_h}$ if $w_h \in (0, \pi)$, $\Delta_h(L, d_h) = (1 - L)^{d_h}$ if $w_h = 0$ and $\Delta_h(L, d_h) = (1 + L)^{d_h}$ if $w_h = \pi$. This model has been considered before by Chan and Wei (1988), Robinson (1994), Chan and Terrin (1995), Nielsen (2004) and Giraitis and Leipus (1995), who use the term "fractional ARUMA" to refer to the model in (1) with u_t a stationary and invertible ARMA. However, the model in (1) is much more general because it does not restrict u_t parametrically but it can be any short memory process or even stationary long memory with spectral poles at some frequencies other than the w_h (see Assumptions A.1 and A.3 below).

The definition of X_t in (1) places it in Type II long memory processes, which differ from Type I in the pre-sample treatment. Type II processes are based on a truncated application of the expansion of $\Delta^H(L, -d)$ up to $t = 0$, when the process is assumed to be initialised. In (1) the initial value X_0 is assumed to be known, specifically $X_0 = 0$. Such processes are then non-stationary and only asymptotically stationary for $d_h < 1/2$. Type I processes can be defined as weighted partial sums of stationary processes $Y_t = \prod_{h=1}^H \Delta_h(L, k_h - d_h)u_t$ for integer k_h such that$|k_h - d_h| < 1/2$. For $k_h = 1$ and $H = 1$ the weighted partial sum is implemented as $X_t = X_0 + \sum_{j=0}^{t-1} b_j Y_{t-j}$ for $b_0 = 1$, $b_1 = 2\cos w_h$, $b_j = 2\cos w_h b_{j-1} - b_{j-2}$ for $j > 1$ where the weights come from the expansion of $(1 - 2\cos w_h L + L^2)^{-1}$. For $H > 1$ such a weighted sum is applied for every $h = 1, ..., H$ and successive applications apply for larger values of k_h. For standard long memory, these two definitions could lead to different asymptotics. See for example Robinson (2005) and Shimotsu and Phillips (2006). Similar effects are expected here, though further analysis is necessary before rigorous conclusions can be offered.

The *multiple generalised fractional difference operator* $\Delta^H(L,d)$ is a product of a finite number of generalised fractional difference operators $\Delta_h(L,d_h)$, which can be expanded as $\Delta_h(L,d_h)=\sum_{k=0}^{\infty} c_k(w_h,d_h)L^k$ for

$$\begin{aligned} c_k(w_h,d_h) &= \sum_{j=0}^{\lfloor k/2 \rfloor} \frac{(-1)^j\Gamma(k-j-d_h)(2\cos w_h)^{k-2j}}{\Gamma(j+1)\Gamma(k-2j+1)\Gamma(-d_h)} \text{ for } w_h \in (0,\pi), \\ c_k(0,d_h) &= \frac{\Gamma(k-d_h)}{\Gamma(k+1)\Gamma(-d)} \\ c_k(\pi,d_h) &= (-1)^k c_k(0,d_h) \end{aligned}$$

where $\Gamma()$ is the gamma function. Using this result, $\Delta^H(L,d)=\sum_{k=0}^{\infty} d_k(d)L^k$ where the coefficients $d_k(d)$ can be derived as functions of different $c_k(w_h,d_h)$, $h=1,...,H$ from the relation $\Delta^H(L,d)=\prod_{h=1}^{H}\Delta_h(L,d_h)$. For example, for $H=2$, $d_k(d)=\sum_{k_1=0}^{k} c_{k_1}(w_1,d_1)c_{k-k_1}(w_2,d_2)$. Thus, for $t=1,2,...,n$, X_t in (1) can be written as

$$D_n(L,d)X_t = u_t I(t\geq 1) \quad \text{for} \quad D_n(L,d)=\sum_{k=0}^{n} d_k(d)L^k$$

and, by inversion of (1) $X_t = D_n(L,-d)u_t I(t\geq 1) = D_{t-1}(L,-d)u_t$.

Using Lemmas 2 and 3 in the supplement, the (pseudo)spectral density function of X_t can be approximated as

$$f_x(w_h+\lambda) \sim C_h(d)|\lambda|^{-2d_h} \tag{2}$$

as $\lambda \to 0$, where

$$C_h(d) = f_u(w_h)|2g_h|^{-2d_h}\prod_{\substack{l=1\\ l\neq h}}^{H} A_{l,h}^{-2d_l\delta_l} \tag{3}$$

for $\delta_l = 0.5$ if $w_l = 0,\pi$ and $\delta_l = 1$ in any other case, $A_{l,h} = |4\sin(0.5[w_h+w_l])\sin(0.5[w_h - w_l])|$ and $g_h = g(\sin w_h)$ where $g()$ is a function in $[0,1]$ defined as $g(x)=x$ if $x\in(0,1]$ and $g(0)=0.5$. Thus, the term $|2\sin w_h|^{-2d_h}$ only appears if $w_h \neq 0,\pi$. In the case of a single long memory component at such w_h then $C_h(d)=C_h(d_h)=f_u(w_h)|2\sin w_h|^{-2d_h}$ (see also Giraitis et al. 2001 for the stationary and invertible case). However, with more than one long memory components the term $\prod_{\substack{l=1\\ l\neq h}}^{H} A_{l,h}^{-2d_l\delta_l}$ arises and thus $f_x(w_h+\lambda)$ depends not only on the memory parameter at frequency w_h but also on the rest of memory parameters through the dependence of $C_h(d)$ on $d=(d_1,...,d_H)'$. This behaviour has been noted before by Giraitis and Leipus (1995) and Giraitis et al. (2001) in stationary series but it has never been extended to non-stationarity or used for semiparametric estimation of d. In fact, the constant $C_h(d)$ contains relevant information on the H memory parameters, not only on d_h, and ignoring that dependence may induce severe biasing effects.

3 Multiple Exact Local Whittle estimation

Consider first the case of a single pole ($H = 1$ in equation (1)) at a known frequency w_1. Extensions of the Local Whittle estimator in Robinson (1995) to the $w_1 \neq 0$ case have been proposed by Arteche (2000), Arteche and Robinson (2000) and Arteche and Velasco (2005) for invertible and stationary series with $-0.5 < d_1 < 0.5$. They deal with spectral density functions satisfying $f_x(w_1 + \lambda) \sim C|\lambda|^{-2d_1}$ as $\lambda \to 0$ for C a positive constant, without considering its possible dependence on d. Taking that dependence into account $f_x(w_1 + \lambda) \sim G_1|2g_1|^{-2d_1}|\lambda|^{-2d_1}$ as $\lambda \to 0$ (see (2) and (3)) for $G_1 = f_u(w_1)$ and thus the Local Whittle contrast function becomes

$$\sum_j \left\{ \log(G_1|2g_1|^{-2d_1}|\lambda_j|^{-2d_1}) + \frac{I_x(w_1 + \lambda_j)}{G_1|2g_1|^{-2d_1}|\lambda_j|^{-2d_1}} \right\},$$

for Fourier frequencies $\lambda_j = 2\pi j/n$, $j = 1, ..., m$, with m being the bandwidth, and $\sum_j = \sum_{j=\pm 1}^{\pm m_h}$ if $w_h \neq 0, \pi$, $\sum_j = \sum_{j=1}^{m_h}$ if $w_h = 0$ and $\sum_j = \sum_{j=-1}^{-m_h}$ if $w_h = \pi$ (accounting for spectral symmetry at 0 and π) and for a general series a_t, $t = 1, 2, ..., n$, $I_a(\lambda)$ is the periodogram evaluated at frequency λ defined as

$$I_a(\lambda) = |W_a(\lambda)|^2 \quad \text{for} \quad W_a(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} a_t e^{it\lambda},$$

where $W_a()$ is the discrete Fourier transform (DFT) of a_t.

Concentrating the constant G_1 out of the objective function the term $|2g_1|^{-2d_1}$ cancels out and the *Local Whittle (LW)* estimator of d_1 is obtained by minimising the function

$$\log \left(\frac{1}{2\delta_1 m} \sum_j \frac{I_x(w_1 + \lambda_j)}{|\lambda_j|^{2d_1}} \right) - \frac{d_1}{\delta_1 m} \sum_j \log|\lambda_j| \tag{4}$$

which for $w_1 = 0$ is the standard Local Whittle estimator and for $w_1 \neq 0$ is actually the estimator considered in Arteche (2000), Arteche and Robinson (2000) and Arteche and Velasco (2005). The consistency and asymptotic normality postulated in those papers remain valid because only constancy of $C(d)$ with λ is required. Note also that the estimator of d_1 based on the log periodogram regression proposed by Arteche and Robinson (2000) is similarly unaffected by the absence of $|2g_1|^{-2d_1}$ because its inclusion only implies adding a constant to the regressor such that it shifts from $-2\log|\lambda_j|$ to $-2\log|\lambda_j| - 2\log|2g_1|$ and therefore does not affect the least squares estimation of d.

In order to extend the range of values of d to cover highly non-stationary and non-invertible series, Shimotsu and Phillips (2005) propose the Exact Local Whittle estimator for $w_1 = 0$. We extend it now to cover the $w_1 \neq 0$ case. It starts from the local (around w_1) Whittle approximation of the negative Gaussian log likelihood of the innovations u_t (omitting constants), defined as

$$\frac{1}{2\delta_1 m}\sum_j \left\{\log G_1 + \frac{I_u(w_1+\lambda_j)}{G_1}\right\}. \tag{5}$$

In contrast to the original proposal in Shimotsu and Phillips (2005), this expression takes note of the lack of symmetry of the periodogram at $w_1 \neq 0, \pi$ and considers frequencies on both sides of such a w_1. For the sake of simplicity, the same number of frequencies m are considered on both sides of an $w_1 \neq 0$. Different bandwidths are also possible but that extension only complicates notation without affecting the results obtained hereafter.

The basic resource of the Exact Local Whittle estimation is to transform (5) to express it in terms of the observable data. To that end, an exact relationship between W_x and W_u is used such that, for any value of d, $I_u(\lambda) = |D_n(e^{i\lambda}, d_1)|^2 |v_x(\lambda)|^2$ where

$$v_x(\lambda) = W_x(\lambda) - \frac{1}{\sqrt{2\pi n}} D_n(e^{i\lambda}, d)^{-1}\tilde{X}_n(d)$$

for

$$\tilde{X}_n(d) = \sum_{p=0}^{n-1} \tilde{c}_p(\lambda, d) e^{-i\lambda p} X_{n-p} \quad , \quad \tilde{c}_p(\lambda, d) = \sum_{k=p+1}^{n} d_k(d) e^{ik\lambda}$$

(see Lemma 4 in the supplement). Thus, replacing $I_u(w_1+\lambda_j)$ in (5) with $|D_n(e^{i(w_1+\lambda_j)}, d_1)|^2$ $|v_x(w_1+\lambda_j)|^2$ and adding the Jacobian $\sum_j \log|D_n(e^{i(w_1+\lambda_j)}, d_1)|^{-2}$, the loss function becomes

$$\frac{1}{2\delta_1 m}\sum_j \left\{\log\left(G_1 |D_n(e^{i(w_1+\lambda_j)}, d_1)|^{-2}\right) + \frac{I_{\Delta_1^{d_1} x}(w_1+\lambda_j)}{G_1}\right\},$$

where $I_{\Delta_1^{d_1} x}(\lambda)$ is the periodogram of $D_{t-1}(L, d_1)X_t$ at frequency λ. Concentrating G_1 out of the objective function and using the approximation of $|D_n(e^{i(w_1+\lambda_j)}, d_1)|^{-2}$ in Lemma

3 in the supplement (for a single w_1), the *Exact Local Whittle (ELW)* estimator of d_1 is defined as $\hat{d}_1 = \arg\min_{d_1 \in [\Delta_{11}, \Delta_{12}]} R_1(d_1)$ for $-\infty < \Delta_{11} < \Delta_{12} < \infty$ for

$$R_1(d_1) = \log \hat{G}_1(d_1) - \frac{d_1}{\delta_1 m} \sum_j \log |\lambda_j| \quad \text{with} \quad \hat{G}_1(d_1) = \frac{1}{2\delta_1 m} \sum_j \frac{I_{\Delta_1^d x}(w_1 + \lambda_j)}{|2g_1|^{2d_1}} \tag{6}$$

where the dependence on $|2g_1|^{-2d_1}$ remains relevant. The term $|2g_1|^{2d_1}$ actually represents the main difference from the ELW estimator at frequency zero in Shimotsu and Phillips (2005). Note that the power transfer function of the generalised fractional difference operator at frequency $w_1 + \lambda$ is equal to $|1 - 2\cos w_1 e^{i(w_1+\lambda)} + e^{2i(w_1+\lambda)}|^{-2\delta_1 d_1} = |1 - e^{i(2w_1+\lambda)}|^{-2\delta_1 d_1} |1 - e^{i\lambda}|^{-2\delta_1 d_1}$. The last term in the product is the power transfer function of the standard fractional difference operator at frequency λ, $(1-L)^{-\delta_1 d_1}$, and thus approaches $\lambda^{-2\delta_1 d_1}$ as $\lambda \to 0$, while the first approaches $|2\sin w_1|^{2d_1}$ if $w_1 \in (0, \pi)$ or λ^{-d_1} for $w_1 = 0$. Thus, the (pseudo) spectral density function of X_t satisfies $f_x(w_1 + \lambda) \sim f_u(w_1)|2\sin w_1|^{-2d_1}|\lambda|^{-2d_1}$ as $\lambda \to 0$ for $w_1 \in (0, \pi)$, whereas in the standard case $w_1 = 0$ and $f_x(\lambda) \sim f_u(w_1)|\lambda|^{-2d_1}$.

Now consider X_t in (1) with possible spectral poles or valleys at H known frequencies satisfying $0 \leq w_1 < ... < w_h < .. < w_H \leq \pi$ and $|w_h - w_{h-1}| > \delta$ for a constant $\delta > 0$ and $h = 2, ..., H$. This last restriction implies that the poles are distinguishable and the distance between two consecutive poles is not affected by the sample size n. For example, in the persistent seasonal case $w_h = 2\pi h/S$ and $|w_h - w_{h-1}| = 2\pi/S$. Start once more with the local Whittle approximation of the negative log likelihood of the innovations u_t around the H spectral poles, which in this case is

$$\sum_{h=1}^{H} \frac{1}{2\delta_h m_h} \sum_j \left\{ \log G_h + \frac{I_{uhj}}{G_h} \right\} \tag{7}$$

where $I_{uhj} = I_u(w_h + \lambda_j)$ and m_h are possibly different bandwidths. As before, the different values of δ_h account for the spectral symmetry at 0 and π ($\delta_h = 0.5$) such that only one side of the spectral pole is used in the construction of the likelihood function, whereas in the rest of cases both sides are informative ($\delta_h = 1$). To make (7) data dependent consider again the exact relationship between W_x and W_u such that $I_u(\lambda) = |D_n(e^{i\lambda}, d)|^2 |v_x(\lambda)|^2$ (see Lemma 4 in the supplement). Replacing I_{uhj} in (7) with $|D_n(e^{i(w_h+\lambda_j)}, d)|^2 |v_x(w_h + \lambda_j)|^2$ and adding the Jacobian $\sum_{h=1}^{H} \sum_j \log |D_n(e^{i(w_h+\lambda_j)}, d)|^{-2}$, the Multiple Exact Local Whittle estimator of $d_0 = (d_{10}, d_{20}, ..., d_{H0})$ is obtained by minimising

$$\sum_{h=1}^{H} \frac{1}{2\delta_h m_h} \sum_j \left\{ \log G_h |D_n(e^{i(w_h+\lambda_j)}, d)|^{-2} + \frac{I_{\Delta_H^d x}(w_h+\lambda_j)}{G_h} \right\},$$

where $I_{\Delta_H^d x}(\lambda)$ is the periodogram at λ of $D_{t-1}(L,d)X_t$. Concentrating G_h, $h = 1, ..., H$, out of the objective function, and using the approximation of $|D_n(e^{i(w_h+\lambda_j)}, d)|^{-2}$ in Lemma 3, the *Multiple Exact Local Whittle (MELW)* estimator is defined as

$$\hat{d} = (\hat{d}_1, \hat{d}_2, ..., \hat{d}_H) = \arg \min_{d \in \prod_{h=1}^{H} [\Delta_{h1}, \Delta_{h2}]} R_H(d)$$

for $-\infty < \Delta_{h1} < \Delta_{h2} < \infty$, $h = 1, ..., H$, and

$$R_H(d) = \sum_{h=1}^{H} \left\{ \log \hat{G}_h(d) - \frac{d_h}{\delta_h m_h} \sum_j \log |\lambda_j| \right\} , \qquad (8)$$

$$\hat{G}_h(d) = \frac{1}{2\delta_h m_h} \sum_j \frac{I_{\Delta_H^d x}(w_h+\lambda_j)}{|2g_h|^{2d_h}} \prod_{\substack{l=1 \\ l\neq h}}^{H} A_{l,h}^{-2\delta_l d_l}.$$

Note that the MELW estimator for $H = 1$ is the individual ELW estimator obtained by minimising (6) and for $w_1 = 0$ is the ELW proposed by Shinotsu and Phillips (2005).

Consider the following set of conditions required for consistency of the MELW estimator. In what follows the subindex 0 is used as usual to represent the true values of the parameters.

Assumptions for consistency:

A1: For $h = 1, ..., H$, $f_u(w_h + \lambda) \sim G_{h0} \in (0, \infty)$ as $\lambda \to 0$.

A.2: In a neighbourhood $(-\delta, 0) \cup (0, \delta)$ of w_h, $f_u(\lambda)$ is differentiable and, as $\lambda \to 0^+$,

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} \log f_u(w_h + \lambda) \quad = \quad O(|\lambda|^{-1})$$

for $h = 1, ..., H$.

A.3: $u_t = B(L)\epsilon_t = \sum_{j=0}^{\infty} b_j \epsilon_{t-j}$ and $\sum_{j=0}^{\infty} b_j^2 < \infty$ where $E[\epsilon_t|F_{t-1}] = 0$, $E[\epsilon_t^2|F_{t-1}] = 1$ a.s. for $t = 0, \pm 1, \pm 2, ...$, F_t is the σ-field generated by ϵ_s, $s \leq t$, and there is a random variable ϵ such that $E\epsilon^2 < \infty$ and for all $\eta > 0$ and some $\kappa < 1$, $P(|\epsilon_t| > \eta) \leq \kappa P(|\epsilon| > \eta)$.

A.4: For $h = 1, ..., H$ and any $\gamma > 0$, as $n \to \infty$,

$$\frac{1}{m_h} + \frac{m_h (\log m_h)^{1/2}}{n} + \frac{\log n}{m_h^{\gamma}} \to 0.$$

A.5: For $h = 1, ..., H$, $\Delta_{h2} - \Delta_{h1} \leq \frac{9}{2}$.

Assumptions A.1 to A.3 are direct extensions of Assumptions 1 to 3 in Shimotsu and Phillips (2005) to the process defined in (1) (see also assumptions A.1-A.3 in Arteche, 2000). In particular, A.1 requires the spectral density function of u_t to be bounded and bounded away from zero at w_h, $h = 1, ..., H$, which is needed for identifiability of $d_1, ..., d_H$. A.3 imposes stationarity of u_t allowing for long memory such that there can be spectral poles in $f_x(\lambda)$ at frequencies other than the w_h in $\Delta^H(L, d)$. This makes for greater robustness in the MELW estimator because there is no need to consider the estimation of all the spectral poles as long as u_t is stationary. However, non-stationary poles are not allowed in u_t and if there are any the MELW is subject to a large bias and variance inflation suggesting inconsistency (see the Monte Carlo in Section 5). Finally, assumptions A.4 and A.5 are the same as in Shimotsu and Phillips (2005) but for every $h = 1, ..., H$. Note that the elements in d_0 can take any value and the only restriction is that they must belong to the set defined by Assumption A.5, which is broad enough to cover all the cases of interest in economic time series.

Theorem 1 *Let X_t be generated as in equation (1) with $d_0 = (d_{10}, d_{20}, ..., d_{H0}) \in \prod_{h=1}^{H}[\Delta_{h1}, \Delta_{h2}]$. Let assumptions A.1-A.5 be satisfied. Then $\hat{d} \xrightarrow{p} d_0$ as $n \to \infty$.*

The proof of Theorem 1 can be found in the supplementary material. It is based on proving the convergence in probability for different subsets of the parameter space as established in Assumption A.5 and the successive application of $\Delta_h(L, 1)$, $\Delta_h(L, -1)$ and Lemma 4 in the supplement to get exact relations between the periodograms of filtered and unfiltered series.

To obtain the asymptotic distribution, the required assumptions need to be strengthened as follows:

Assumptions for asymptotic normality:

B.1: For $h = 1, ..., H$ and $\beta_h \in (0, 1]$ (if $w_h \in (0, \pi)$), $\beta_h \in (0, 2]$ (if $w_h = \{0, \pi\}$),

$$f_u(w_h + \lambda) = G_{h0}(1 + O(|\lambda|^{\beta_h}) \text{ as } \lambda \to 0, \quad G_{h0} \in (0, \infty).$$

B.2: For $h = 1, ..., H$, in a neighbourhood $(-\delta, 0) \bigcup (0, \delta)$ of w_h, $B(e^{i\lambda})$ is differentiable and

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}B(e^{i(w_h+\lambda)}) = O(|\lambda|^{-1}) \text{ as } \lambda \to 0.$$

B.3: Assumption A.3 holds and for finite constants μ_3 and μ_4,

$$E(\epsilon_t^3|F_{t-1}) = \mu_3 \text{ and } E(\epsilon_t^4|F_{t-1}) = \mu_4, \ a.s., \ t = 0, \pm 1, ...$$

B.4: As $n \to \infty$ and for $h = 1, ..., H$,

$$\frac{1}{m_h} + \frac{m_h^{1+2\beta_h}\log^2 m_h}{n^{2\beta_h}} + \frac{\log n}{m_h^{\gamma}} \to 0 \quad \text{for any } \gamma > 0$$

and for any $q, r \in \{1, 2, ..., H\}$, $m_q m_r^{-1} = O(1)$.

B.5: If $w_h \in (0, \pi)$ for $h = 1, 2, ..., H$, then $0 < w_1 - \lambda_{m_1} < w_1 + \lambda_{m_1} < w_2 - \lambda_{m_2} < w_2 + \lambda_{m_2} < ... < w_H - \lambda_{m_H} < w_H + \lambda_{m_H} < \pi$. If $w_1 = 0$, $\lambda_{m_1} < w_2 - \lambda_{m_2}$ and if $w_H = \pi$, $w_{H-1} + \lambda_{m_{H-1}} < \pi - \lambda_{m_H}$.

Assumptions B.1-B.3 are analogous to Assumptions 1'-3' in Shimotsu and Phillips (2005). Assumption B.4 includes proportionality of bandwidths at different frequencies. It basically implies that no single band of frequencies around a spectral pole dominates the others. Assumption B.5 is imposed to avoid correlation between the bands of periodogram ordinates used in estimating memory parameters at different frequencies. If the estimation involves overlapping intervals of frequencies, that correlation will affect the variance in the asymptotic distribution in Theorem 2.

Theorem 2 *If X_t is generated by (1) with $d_0 \in \prod_{h=1}^{H}(\Delta_{h1}, \Delta_{h2})$ and assumptions B.1-B.5 and A.5 are satisfied, then as $n \to \infty$*

$$\Lambda_m(\hat{d} - d_0) \xrightarrow{d} N_H\left(0, \frac{1}{4}\mathbb{I}_H\right)$$

where $\Lambda_m = diag(\sqrt{2\delta_1 m_1}, ..., \sqrt{2\delta_H m_H})$ and $\mathbb{I}_H$ is the identity matrix.

Remark 1: The asymptotic distribution in Theorem 2 implies that the estimators of the different memory parameters at different frequencies are asymptotically independent. However, a joint estimation is needed to guarantee fulfilment of assumption A.3, which excludes the existence of non-stationary poles in the spectral density function of the filtered series $D_n(L, d_0)X_t$. As shown in the Monte Carlo in Section 5, ignoring the existence of other non-stationary spectral poles generates a significant bias and variance inflation that suggest inconsistency. However the filtered series u_t can include stationary long memory at some frequencies $w \neq w_h$, $h = 1, ..., H$.

Remark 2: Theorem 2 permits a very simple implementation of asymptotic inference to test for relations of interest between memory parameters at different frequencies. For example, the hypothesis of unit roots at frequencies w_{i_j}, for $j = 1, ..., k$ can be tested using the Wald statistic

$$4(R\hat{d} - r)'[R\Lambda_m^{-2}R']^{-1}(R\hat{d} - r) \overset{d,H_0}{\to} \chi^2_k$$

where R is a $k \times H$ matrix of zeros except for ones in the elements $j \times i_j$, $j = 1, ..., k$ and r is a $k \times 1$ vector of ones. See Section 5 for a finite sample analysis of this testing strategy.

Remark 3: As in other semiparametric estimators based on Whittle estimation, the variance in the asymptotic distribution tends to underestimate the true variance in finite samples (see the Monte Carlo analysis in Section 5). We propose instead a Hessian-based approximation of $4^{-1}\Lambda_m^{-2}$ defined as an $H \times H$ diagonal matrix with (s, s)-th element:

$$\left\{ 4\delta_s^2 \sum_j \left[Re\{\bar{J}_{nj}(w_s, w_s)\} - \frac{1}{2\delta_s m_s} \sum_k Re\{\bar{J}_{nk}(w_s, w_s)\} \right]^2 \right\}^{-1} \tag{9}$$

for $\bar{J}_{nj}(w_r, w_s) = J_n(e^{i(w_r+\lambda_j)}, w_s) + J_n(e^{i(w_r+\lambda_j)}, -w_s)$ and $J_n(L, w) = \sum_{k=1}^{n} \frac{e^{ikw}}{k} L^k$. Consistency of this approximation can be deduced from formula (43) in the proof of Theorem 2 in the supplementary material. The Monte Carlo in Section 5 shows that this is a much better approximation of the true finite sample variance than the expression in the asymptotic distribution, leading to more reliable inference.

Remark 4: The locations w_h of the poles are assumed to be known, as occurs for example in series with strong seasonality. In other cyclical contexts, they can be estimated as suggested by Hidalgo and Soulier (2004) or Hidalgo (2005). Hidalgo and Soulier (2004) propose to maximize the periodogram and show that the limiting distribution of a log-periodogram-based estimator of the memory parameter remains the same irrespective of whether the location of the pole is known or estimated. Hidalgo (2005) proposes instead to estimate w_h by maximizing an estimation of the memory parameter over a grid of locations. His simulations show that some semiparametric estimators of d can be significantly affected by replacing the unknown location of the pole by an estimated one.

It is not clear how a prior estimation of the w_h will affect the asymptotic properties of the MELW in our local estimation set up, because not only is the band of frequencies affected but also the filtered series $D_n(L,d)X_t$ would be defined with the w_h replaced by their estimates. Some light is shed via simulations in Section 5. Note however that the MELW estimator remains consistent and asymptotically normal with no need to account for all the (stationary) spectral poles and includes the case $d_h = 0$. The number and location of the poles can then be deduced by testing the hypothesis $d_h = 0$ for all the w_h where a spectral pole is expected a priori (seasonal frequencies, business cycles, trends, etc.).

Remark 5: A significant bias arises if the dependence of the constant $C(d)$ on d is ignored and $R_H(d)$ is naively defined as a shifted version of the ELW loss function proposed in Shimotsu and Phillips (2005). In that case d is estimated by minimising the function $R_H(d)$ in (8) with $\hat{G}_h(d)$ replaced by $\hat{G}_h(d) = (2\delta_h m_h)^{-1}\sum_j I_{\Delta_H^d x}(w_h + \lambda_j)$. In view of (26), (28) and the convergence of the score in the proof of Theorem 2 in the supplementary material, the bias can be approximated by $E(\hat{d} - d_0) \approx B$, where B is an $H \times 1$ vector with s-th element defined as $[B]_s = -0.5\left(\log|2g_s| + \sum_{h\neq s}\delta_h \log A_{s,h}\right)$. The Monte Carlo analysis in Section 5 shows that this is an accurate approximation of the true bias for large values of the bandwidth and can thus be used to adjust the estimates for bias reduction. However, this bias-adjusted estimator performs worse than the MELW, especially for small bandwidths, when this approximation of the bias fails.

Remark 6: The traditional local Whittle estimator can be similarly extended to cover estimation of multiple poles in stationary and invertible long memory series. In this context the contrast function is

$$\sum_{h=1}^{H}\frac{1}{2\delta_h m_h}\sum_j\left\{\log\left(G_h|2g_h|^{-2d_h}\prod_{\substack{l=1\\l\neq h}}^{H}A_{l,h}^{-2\delta_l d_l}\right) - 2d_h\log|\lambda_j| + \frac{I_x(w_h+\lambda_j)}{G_h|2g_h|^{-2d_h}\prod_{\substack{l=1\\l\neq h}}^{H}A_{l,h}^{-2\delta_l d_l}|\lambda_j|^{-2d_h}}\right\}.$$

Concentrating G_h out of the objective function, the local Whittle estimator of d is obtained by minimising

$$\sum_{h=1}^{H}\left\{\log\left[\frac{1}{2\delta_h m_h}\sum_j I_x(w_h+\lambda_j)|\lambda_j|^{2d_h}\right] - \frac{2d_h}{2\delta_h m_h}\sum_j\log|\lambda_j|\right\},$$

where all the dependence of the constant on d cancels out. Using the results in Robinson (1995) and Arteche (2000) it can be shown that this estimator is consistent with the same asymptotic distribution as that in Theorem 2 but only for stationary and invertible processes, which means that no gain is obtained over individual local Whittle estimation as long as disjoint sets of frequencies are used in the estimation of the different d_h, $h = 1, 2, ..., H$.

4 Unknown deterministic components

One important limitation of ELW estimation is the assumption of known mean or initial value. Shimotsu (2010) deals with this limitation in standard long memory series by estimating an unknown mean with the sample mean (if $d \in (-1/2, 3/4)$), the first observation (for $d > 0$) or a linear combination of both. When the range of frequencies considered is broadened from zero to the interval $[0, \pi]$, deterministic cycles are also a possibility. Consider for example a deterministic cycle at frequency $\bar{w}$

$$Y_t = \mu \cos(\bar{w}t) + X_t \ , \quad \Delta^H(L, d_0) X_t = u_t I(t \geq 1), \quad t = 0, 1, ..., n. \tag{10}$$

where μ is an unknown constant. Note that more general deterministic cyclical behaviours, in the form for example of deterministic seasonal dummies, can be expressed as linear combinations of cosine terms (see Arteche and Robinson, 1999). Note also that $\bar{w} = 0$ implies that the deterministic component is an unknown mean μ as analysed in Shimotsu (2010).

If the unknown μ is ignored, the unadjusted MELW estimator is obtained by minimising the objective function in (8) with $I_{\Delta_H^d x}$ replaced by $I_{\Delta_H^d y}$. The effects of the deterministic components can be analysed from the relation between DFTs: $W_{\Delta_H^d y}(w_h + \lambda_j) = \mu W_z(w_h + \lambda_j) + W_{\Delta_H^d x}(w_h + \lambda_j)$ for $z_t(d) = \Delta^H(L, d) \cos(\bar{w}t) I(t \geq 1)$. By Lemma 10 in the supplementary material $W_z(w_h + \lambda_j) = O\left(n^{1/2} j^{-1} \left[\lambda_j^{d_h} + n^{-\underline{d}} \log n\right]\right)$ if $w_h = \bar{w}$ and $O\left(n^{-1/2} \left[\lambda_j^{d_h} + n^{-\underline{d}} \log n\right]\right)$ if $w_h \neq \bar{w}$ for $\underline{d} = \min(d_1, ..., d_H)$. This result suggests that the deterministic component has a larger effect on the estimation of the memory parameter at the frequency of the deterministic cycle than at the rest of frequencies. It also suggests that the effect of the deterministic component is less adverse for $\underline{d} > 0$, which implies consideration of only positive memory at every frequency $w_1, .., w_H$. In fact, Shimotsu (2010) showed that the standard ELW estimator maintains its asymptotic properties when the unknown mean μ is replaced by any $O_p(1)$ term if $d > 0$ and the bandwidth increases fast enough (see Theorems 2a, 2b and his Remark 2), which implies asymptotic robustness of the ELW estimator to an unknown mean. A similar property is expected in the MELW estimator as long as the bandwidth is selected such that the DFT of $\Delta^H(L,d)X_t$ dominates the DFT of $z_t(d)$. However the finite sample performance is significantly affected by the selection of that $O_p(1)$ term. In order to reduce this effect, Shimotsu (2010) suggests to use Y_1 as an estimator of the unknown mean. In the multiple long memory processes considered here we suggest to deseasonalise by subtracting the first S observations, where $S = 2\pi/\varpi$ is the period of the deterministic cycle. The MELW estimator is then implemented in the deseasonalised series $Y_{Sk+s} - Y_s$ for $s = 1, 2, ..., S$ and $k = 0, 1, ..., n/S - 1$.

Alternatively μ can be estimated by ordinary least squares as

$$\hat{\mu} = \frac{\sum_{t=1}^{n} \cos(\varpi t)\, Y_t}{\sum_{t=1}^{n} \cos^2(\varpi t)}.$$

Note that $\hat{\mu}$ is the sample mean if $\varpi = 0$, as analysed in Shimotsu (2010). Consider for simplicity that ϖ is a Fourier frequency. Then $\sum_{t=1}^{n} \cos^2(\varpi t) = n/2$ and thus

$$\hat{\mu} = \frac{1}{n} \sum_{t=1}^{n} 2\cos(\varpi t)\, Y_t = \mu + \frac{1}{n} \sum_{t=1}^{n} 2\cos(\varpi t)\, X_t.$$

Lemma 1 *Consider Y_t in (10) and denote $\bar{d}_0 = \max\{d_{h0} + I(w_h = \varpi)\}_{h=1}^{H}$ where $I(w_h = \varpi) = 1$ if $w_h = \varpi$ and zero otherwise. If $\bar{d}_0 > 1/2$ then $Var(\hat{\mu}) = O(n^{2\bar{d}_0 - 3})$. Therefore $\hat{\mu}$ converges in mean square to μ as $n \to \infty$ if $1/2 < \bar{d}_0 < 3/2$.*

Note that if $\varpi = 0$ and $H = 1$ Lemma 1 implies the usual consistency of the sample mean because in that case $\bar{d}_0 = d_{10} + 1$ and then $\hat{\mu}$ converges at the usual rate $O_p(n^{d_{10} - 1/2})$. However, when $H > 1$ and some other persistent cycle exists, the memory parameter at those other frequencies can affect the convergence of the sample mean if they are larger than the memory parameter at the origin plus one. In general, for $\varpi \in [0, \pi]$ the convergence of $\hat{\mu}$ not only depends on the persistence at ϖ but it may also depend on the rest of memory parameters if they are large enough. Note also that consistency of $\hat{\mu}$ does not restrict X_t to be stationary as long as the source of non-stationarity is not at frequency ϖ.

Lemma 1 suggests that the MELW estimator with $\Delta_H^d X_t$ replaced by $\Delta_H^d(Y_t - \hat{\mu}\cos(\bar{w}t))$ in the objective function (8) is a good option for dealing with unknown deterministic terms as long as $1/2 < \bar{d}_0 < 3/2$, which implies that the memory parameter at $\bar{w}$ is in $(-1/2, 1/2)$. Considering the standard long memory case ($H = 1$, $\bar{w} = 0$), Shimotsu (2010) shows that the ELW with μ estimated by the sample mean is consistent for $d_{10} \in (-1/2, 1)$ and asymptotically normal if $d_{10} \in (-1/2, 3/4)$. The Monte Carlo in next section suggests that the MELW estimator can also retain the same properties as long as $\bar{d}_0 < 2$, except in the estimation of a negative memory parameter at $\bar{w}$ much lower than the largest one. Note that $\Delta_H^d(Y_t - \hat{\mu}\cos(\bar{w}t)) = \Delta_H^d X_t - (\hat{\mu} - \mu)\cos(\bar{w}t))$ with DFT equal to $W_{\Delta_H^d X}(\lambda) - (\hat{\mu} - \mu) W_{\cos(\bar{w}t)}(\lambda)$. Lemmas 1 and 10 in the supplement suggest that the second term will be negligible if the distance between highest and lowest memory parameter is controlled. For example, the Monte Carlo with two memory components shows that the performance of the estimation of d at $\bar{w}$ becomes significantly worse when it takes the value -0.4 and the other memory parameter is at least 1.5.

All things considered, an option for MELW estimation that deals with deterministic cycles always exists if all the memory parameters are positive (subtracting the first observations) or if $\bar{d}_0 \in (1/2, 2)$ (using $\hat{\mu}$). This implies that if there are some negative memory parameters and some others are larger than 2 then neither of these adjustments are valid options for MELW estimation (see also Tables 9 and 10 in next section). Note also that, attending to the poor performance of the tapered LW estimator in Section 5, a two step strategy as suggested by Shimotsu (2010) is not a possibility either.

5 Finite sample behaviour

This section provides a Monte Carlo analysis of the finite sample behaviour of the MELW estimator. It starts with the individual estimation of a single spectral pole where the performance of the ELW estimator is compared with other local estimators such as tapered versions of the LW and the naive extension of the ELW of Shimotsu and Phillips (2005) discussed in Remark 5. The case of unknown location of the spectral pole is also discussed. We deal next with the joint estimation of several memory parameters and the effects of deterministic terms. Finally, the use of the MELW estimator and its asymptotic distribution to test for common unit roots is analysed.

5.1 Single pole

Consider processes of the form

$$\Phi(L)(1-2\cos w_1 L+L^2)^d X_t = u_t I(t\geq 1), \quad t=0,1,...,n$$

where the u_t are independent standard normal and $w_1=\pi/4$. Six different values of d are considered $d\in\{-3.0,-1.5,0.4,0.8,1.5,3.0\}$ and two different polynomials $\Phi(L)=1$ and $\Phi(L)=1+1.06L-0.6L^2$. The second of these corresponds to a stationary $AR(2)$ with a spectral peak at $\pi/4$, which coincides with the location of the long memory pole and induces a positive bias in the estimation of d if a large bandwidth is used. The results with the $AR(2)$ are relegated to the supplementary material to save space.

The ELW estimator is obtained by minimising (6) with $w_1=\pi/4$. Its performance is compared with four different competitors:

- The original LW estimator as proposed by Arteche and Robinson (2000) obtained by minimising the objective function in equation (4). This estimator has the same asymptotic properties as the ELW for $-0.5<d<0.5$, and is expected to be consistent for $d\leq 1$.

- The *misspecified ELW* obtained by minimising (6) with $\hat{G}_h(d_h)=\sum_{j=\pm 1}^{\pm m} I_{\Delta_h^d x}(\pi/4+\lambda_j)(2m_h)^{-1}$. This is a naive extension of the ELW at frequency zero in Shimotsu and Phillips (2005) ignoring the term $|2\sin w_1|^{-2d}$. It coincides with the ELW only if $w_1=\{\pi/6,5\pi/6\}$, when $\sin w_1=0.5$.

- A tapered version of the LW estimator as suggested by Hurvich and Chen (2000) obtained by minimising (4) with the raw periodogram replaced by the tapered periodogram defined as

$$I_T(\lambda)=\frac{1}{2\pi\sum_{t=1}^n |h_t|^2}\left|\sum_{t=1}^n h_t x_t e^{it\lambda}\right|^2,$$

  for $h_t=0.5*[1-\exp\{i2\pi(t-0.5)/n\}]$. Consistency and asymptotic normality are proved in Hurvich and Chen (2000) for $-1.5<d<0.5$ and $w_1=0$.

- A tapered LW estimator as suggested by Velasco (1999) obtained as above but with the triangular Bartlett taper, $h_t = 1 - |n - 2t|/n$. Velasco shows its consistency and asymptotic normality for $-0.5 < d < 2$ and $w_1 = 0$.

Table 1: ELW, finite sample results, $\Phi(L) = 1$, $n = 512$

	$d = -3$								
	$m = 8$			$m = 32$			$m = 63$		
	bias	s.d.	MSE	bias	s.d.	MSE	bias	s.d.	MSE
LW	2.9990	0.0060	8.9942	2.9225	0.1132	8.5541	2.6926	0.1601	7.2760
ELW	-0.0179	0.2100	0.0444	0.0071	0.0767	0.0059	0.0393	0.0514	0.0042
ELW-sine	-0.3270	0.2029	0.1481	-0.1975	0.0822	0.0458	-0.1555	0.0555	0.0273
TLW (HC)	2.5744	0.2905	6.7119	1.4860	0.2053	2.2503	1.3175	0.1717	1.7654
TLW (V)	2.8763	0.1951	8.3110	1.8367	0.2103	3.4178	1.6120	0.1785	2.6304
	$d = -1.5$								
	$m = 8$			$m = 32$			$m = 63$		
	bias	s.d.	MSE	bias	s.d.	MSE	bias	s.d.	MSE
LW	1.3410	0.2134	1.8438	0.8916	0.2440	0.8546	0.7396	0.2035	0.5884
ELW	-0.0140	0.1934	0.0376	0.0077	0.0743	0.0056	0.0401	0.0499	0.0041
ELW-sine	-0.3251	0.1911	0.1422	-0.1968	0.0793	0.0450	-0.1546	0.0535	0.0267
TLW (HC)	0.1214	0.2337	0.0693	0.0889	0.0991	0.0177	0.1384	0.0696	0.0240
TLW (V)	0.1581	0.2395	0.0824	0.0744	0.0981	0.0151	0.1100	0.0703	0.0170
	$d = 0.4$								
	$m = 8$			$m = 32$			$m = 63$		
	bias	s.d.	MSE	bias	s.d.	MSE	bias	s.d.	MSE
LW	-0.0165	0.2152	0.0466	-0.0084	0.0807	0.0066	-0.0213	0.0524	0.0032
ELW	-0.0228	0.2012	0.0410	0.0031	0.0801	0.0064	0.0391	0.0520	0.0042
ELW-sine	-0.3322	0.1899	0.1464	-0.2024	0.0824	0.0478	-0.1581	0.0563	0.0282
TLW (HC)	0.0714	0.2222	0.0545	0.0205	0.0940	0.0093	-0.0013	0.0631	0.0040
TLW (V)	0.0558	0.2352	0.0584	0.0088	0.0981	0.0097	-0.0110	0.0661	0.0045

Note: Results for $(1 - 2\cos\frac{\pi}{4}L + L^2)^{d_0} X_t = u_t I(t \geq 1)$, $u_t \sim N(0,1)$. LW, ELW, ELW-sine, TLW (HC), TLW (V) denote the original Local Whittle estimator (Arteche and Robinson, 2000), the Exact Local Whittle, the misspecified ELW without the sine term and tapered versions of the Local Whittle estimator with the "efficient" taper in Hurvich and Chen (2000) and the triangular Barlett taper ($p = 2$ in Velasco, 1999) respectively.

Tables 1 and 2 show the Monte Carlo biases, standard deviations and mean square errors (MSE) of the five different estimators obtained with 1000 replications of series of length

$n = 512$ for $\Phi(L) = 1$. The sensitivity of the results to the selection of the bandwidth is analysed by considering three different values of $m = 8$, 32 and 63. The different objective functions are minimised by a grid search over the interval $[-6, 6]$ using the *optimise* function in R, so no initial value is needed for the optimisation procedure. As in Shimotsu and Phillips (2005), Assumption A.5 is violated, but this does not seem to affect the performance of the ELW estimator.

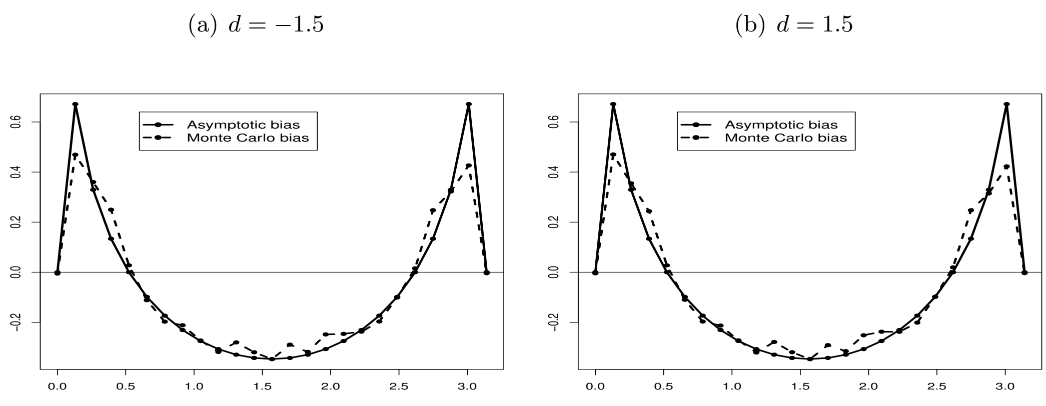
Table 2: ELW, finite sample results, $\Phi(L) = 1$, $n = 512$

	$d = 0.8$								
	$m = 8$			$m = 32$			$m = 63$		
	bias	s.d.	MSE	bias	s.d.	MSE	bias	s.d.	MSE
LW	0.0048	0.2098	0.0441	0.0183	0.0769	0.0062	-0.0141	0.0532	0.0030
ELW	-0.0304	0.2017	0.0416	0.0099	0.0744	0.0056	0.0423	0.0503	0.0043
ELW-sine	-0.3387	0.1894	0.1505	-0.1979	0.0793	0.0455	-0.1543	0.0531	0.0266
TLW (HC)	0.1821	0.2432	0.0923	0.0821	0.0947	0.0157	0.0344	0.0677	0.0058
TLW (V)	0.1658	0.2393	0.0848	0.0559	0.0935	0.0119	0.0063	0.0651	0.0043
	$d = 1.5$								
	$m = 8$			$m = 32$			$m = 63$		
	bias	s.d.	MSE	bias	s.d.	MSE	bias	s.d.	MSE
LW	-0.3985	0.1316	0.1762	-0.4397	0.0723	0.1985	-0.4659	0.0742	0.2225
ELW	-0.0268	0.2063	0.0433	0.0065	0.0778	0.0061	0.0396	0.0526	0.0043
ELW-sine	-0.3398	0.1969	0.1543	-0.2001	0.0830	0.0469	-0.1562	0.0557	0.0275
TLW (HC)	0.3587	0.2279	0.1806	0.2272	0.1172	0.0654	0.1476	0.0981	0.0314
TLW (V)	0.3527	0.2289	0.1768	0.1681	0.1068	0.0397	0.0736	0.0852	0.0127
	$d = 3$								
	$m = 8$			$m = 32$			$m = 63$		
	bias	s.d.	MSE	bias	s.d.	MSE	bias	s.d.	MSE
LW	-1.9781	0.0639	3.9169	-1.9908	0.0341	3.9645	-1.9953	0.0699	3.9862
ELW	-0.0248	0.1985	0.0400	0.0060	0.0768	0.0059	0.0414	0.0521	0.0044
ELW-sine	-0.3317	0.1901	0.1462	-0.2004	0.0811	0.0467	-0.1554	0.0554	0.0272
TLW (HC)	-0.9242	0.1252	0.8698	-0.9190	0.0655	0.8489	-0.9435	0.0791	0.8965
TLW (V)	-0.8870	0.0914	0.7952	-0.9667	0.0404	0.9361	-1.0051	0.0820	1.0169

Note: Results for $(1 - 2\cos\frac{\pi}{4}L + L^2)^{d_0} X_t = u_t I(t \geq 1)$, $u_t \sim N(0,1)$. LW, ELW, ELW-sine, TLW (HC), TLW (V) denote the original Local Whittle estimator (Arteche and Robinson, 2000), the Exact Local Whittle, the misspecified ELW without the sine term and tapered versions of the Local Whittle estimator with the "efficient" taper in Hurvich and Chen (2000) and the triangular Barlett taper ($p = 2$ in Velasco, 1999) respectively.

The lowest MSE is that of the ELW estimator in 16 out of the 18 cases considered. Only in two cases: $d = 0.4$ and $d = 0.8$ with $m = 63$, does the LW offers the lowest MSE, which are the cases where the LW is expected to perform best. But even for $d = 0.4$ and $d = 0.8$ the ELW has lower MSE for small values of $m = 8, 32$. When the AR(2) component is included, the supplement shows that its biasing effect is stronger for the ELW than for the LW, making the bias and MSE of the latter lower than those of the ELW. However, with $m = 8$ the ELW shows a significantly lower bias than the LW estimator even for $d = 0.4$ and $d = 0.8$ (see Tables I and II in the supplement). It is also noteworthy the bad performance of the tapered LW. The use of tapering for estimation of memory parameters in cyclical long memory process was suggested by Arteche and Velasco (2005) to avoid trimming in the asymmetric long memory case. We include them in this Monte Carlo to analyse its use in nonstationary cyclical long memory. Note however that their validity has been proven only for standard (at frequency zero) type I long memory processes and for certain values of d ($d \in (-1.5, 0.5)$ for Hurvich and Chen's and $d \in (-0.5, 2)$ for Velascos's proposal).

Figure 1: Bias of the misspecified ELW



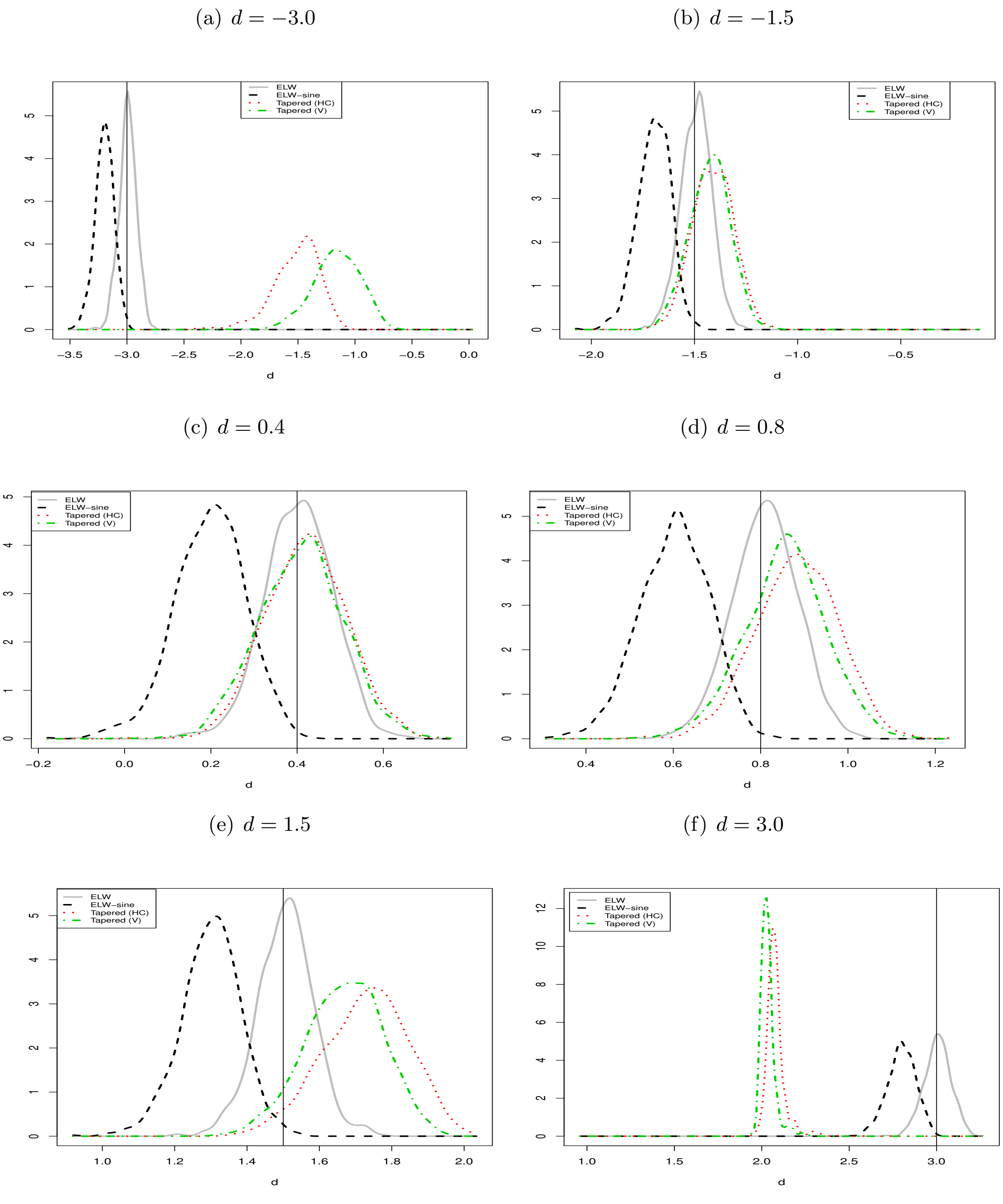
The misspecified ELW shows a negative bias that seems to remain stable with the values of d. To shed more light on this issue, Figure 1 shows its Monte Carlo bias as a function of the frequency w for $d = -1.5, 1.5$, $\Phi(L) = 1$ and $n = 512$. The bandwidth on each side of w is set as the minimum value between 32 and the number of Fourier frequencies between w and 0 (for the left interval) or π (for the right interval). The asymptotic approximation of the bias in Remark 5 is also plotted for the sake of comparison. In this case the approximate bias is $-0.5 \log |2 \sin w|$. Figure 1 shows that this is a good approximation, positive for $w \notin (\pi/6, 5\pi/5)$ and negative for $w \in (\pi/6, 5\pi/5)$, and illustrates the constancy of the bias with respect to d. The bias disappears for $w = \{0, \pi\}$ because there is no misspecification in those cases, and also for $w = \{\pi/6, 5\pi/6\}$ when $\sin(w) = 0.5$. In the cases considered in Tables 1 and 2, $-0.5 \log |2 \sin(\pi/4)| = -0.1733$ and a bias-corrected estimation can be defined by subtracting this value from the misspecified ELW estimator. As expected, this strategy is mainly beneficial with large values of m but for lower values it underestimates the true bias, which remains significant. In any case the ELW performs better.

More information on the performance of the estimators considered is offered in Figure 2, which shows Monte Carlo kernel estimates of the probability density functions of the ELW estimator and possible competitors: the misspecified ELW and the tapered LW estimators. They are all obtained with $m = 32$. The constant bias of the misspecified ELW is clearly apparent. The large bias of the tapered estimators for large (in magnitude) values of d is also noteworthy.

As mentioned in Remark 3, the variance obtained from the asymptotic distribution underestimates the true variance. The standard deviation in the asymptotic distribution in Theorem 2 is $(8m)^{-0.5}$, which for $m = 8$, 32 and 63, is 0.125, 0.062 and 0.044, clearly lower than the values obtained in Tables 1-2. For the same values of m, the standard deviations based on the Hessian-based approximation in (9) are 0.188, 0.071 and 0.046, which are closer to the true values as obtained in the Monte Carlo analysis. The benefits of this approximation can be observed in Table 3, which shows coverage frequencies and average widths of 95% confidence intervals of the form $\hat{d} \pm 1.96\sqrt{var_i}$ for $var_1 = 1/8m$ and var_2 obtained as described in equation (9). The use of the Hessian based approximation leads to coverage frequencies that are significantly closer to the nominal 95% confidence level, especially for low values of m.

Finally, the effect of an estimated w_1 is analysed by considering the maximizer of the periodogram, as proposed by Hidalgo (2005), and the maximizer of a log-periodogram based estimator of the memory parameter, as proposed by Hidalgo and Soulier (2004), over Fourier frequencies $\lambda_j = 2\pi j/n$, $j = 1, ..., [n/2]$. Figure 3 shows the Monte Carlo

Figure 2: Monte Carlo probability density functions, $m = 32$, $\Phi(L) = 1$



(a) $d = -3.0$ (b) $d = -1.5$ (c) $d = 0.4$ (d) $d = 0.8$ (e) $d = 1.5$ (f) $d = 3.0$

probability density functions obtained with 1000 replications of the ELW estimator with the true location $w_1 = \pi/4$, the location estimated by the maximizer of the periodogram and by the maximizer of the estimation of d. The estimation of w_1 using the proposal by Hidalgo (2005) has a distorting effect as advocated by Hidalgo, but this effect decreases as

Table 3: Coverage frequencies, $\Phi(L) = 1$, $n = 512$

	$d=-3.0$		$d=-1.5$		$d=0.4$		$d=0.8$		$d=1.5$		$d=3.0$	
m	Asy.	Hess.	Asy.	Hess.	Asy.	Hess.	Asy.	Hess.	Asy.	Hess.	Asy.	Hess.
8	0.770	0.918	0.822	0.943	0.787	0.935	0.790	0.930	0.778	0.931	0.807	0.941
	(0.490)	(0.752)	(0.490)	(0.752)	(0.490)	(0.752)	(0.490)	(0.752)	(0.490)	(0.752)	(0.490)	(0.752)
32	0.879	0.941	0.902	0.953	0.884	0.945	0.905	0.950	0.891	0.937	0.876	0.943
	(0.245)	(0.293)	(0.245)	(0.293)	(0.245)	(0.293)	(0.245)	(0.293)	(0.245)	(0.293)	(0.245)	(0.293)
63	0.804	0.863	0.830	0.880	0.805	0.867	0.806	0.866	0.814	0.863	0.809	0.862
	(0.175)	(0.197)	(0.175)	(0.197)	(0.175)	(0.197)	(0.175)	(0.197)	(0.175)	(0.197)	(0.175)	(0.197)

Note: Coverage frequencies and widths (in brackets) obtained with 95 % confidence intervals based on the asymptotic normal distribution (Asy.) of the ELW estimator and the Hessian-based approximation of the variance (Hess.). Results for $(1 - 2\cos\frac{\pi}{4}L + L^2)^{d_0} X_t = u_t I(t \geq 1)$, $u_t \sim N(0,1)$.

d gets larger. This is also the case with the maximizer of the periodogram, although here the performance is much better not only for large values of d, but also for lower values. In any case, it should be considered that consistency of both estimators of the location has only been proven for $d < 0.5$.

5.2 Multiple poles

Now consider the process

$$\Phi(L)(1-L)^{d_1}(1+L^2)^{d_2} X_t = u_t I(t \geq 1), \quad t = 0, 1, ..., n$$

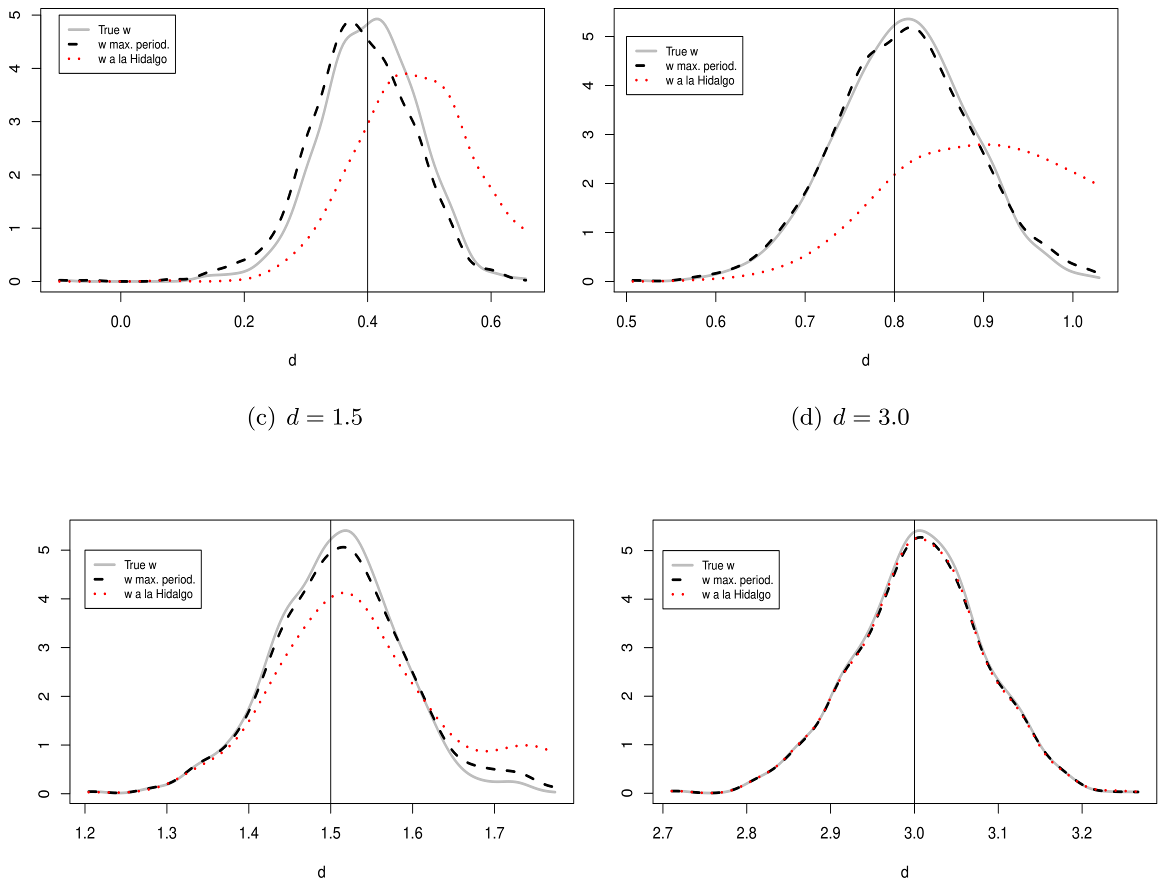
where the u_t are independent standard normal. Five different values of the memory parameters are considered $d_1, d_2 \in \{-1.5, 0.4, 0.8, 1.5, 3.0\}$ and two different polynomials $\Phi(L) = 1$ and $\Phi(L) = 1 - 0.6L^2$, the latter showing a spectral peak at $\pi/2$. Models of this kind are suitable for quarterly series with strong seasonality and stochastic trends. MELW estimates of d_1 and d_2 are obtained by minimising (8) for $H = 2$, $w_1 = 0$, $w_2 = \pi/2$ and $[\Delta_{h1}, \Delta_{h2}] = [-6, 6]$, $h = 1, 2$, as before, using the *sbplx* routine of the *nloptr* package in *R*.

Table 4 shows the Monte Carlo bias, standard deviation and MSE of the MELW estimation of d_1 obtained with 1000 replications of series of $n = 512$ observations with $\Phi(L) = 1$. For the sake of comparison the results obtained with the individual ELW of Shimotsu and

Figure 3: ELW pdf (estimated locations), $m = 32$, $n = 512$, $\Phi(L) = 1$

(a) $d = 0.4$ (b) $d = 0.8$

(c) $d = 1.5$ (d) $d = 3.0$



Note: Pdfs of ELW with known w_1 (continuous line), estimated by the maximizer of the periodogram (dashed line) and a la Hidalgo (2005) (dotted line). Results for $(1 - 2\cos\frac{\pi}{4}L + L^2)^{d_0} X_t = u_t I(t \geq 1)$, $u_t \sim N(0, 1)$.

Phillips (2005) are also included. The MELW has lower bias and standard deviation in every case. The performance of the ELW gets significantly worse as d_2 increases, especially when $d_2 > 0.5$, such that Assumption 3 in Shimotsu and Phillips (2005) is not satisfied. These results indicate that such an assumption may be necessary for consistency of the ELW and supports the use of the MELW estimator for joint estimation of all the memory parameters if strong seasonality is apparent.

A similar situation arises in the estimation of d_2 shown in Table 5. For the sake of comparison, the results obtained with the individual ELW estimator analysed in the previous

Table 4: MELW finite sample results for d_1, $\Phi(L) = 1$, $n = 512$, $m = 32$

	MELW					ELW				
$d_1 \backslash d_2$	-1.5	0.4	0.8	1.5	3.0	-1.5	0.4	0.8	1.5	3.0
-1.5	0.0020	-0.0036	0.0052	0.0040	0.0032	0.0263	-0.0274	-0.0931	-0.6955	-2.1937
	(0.1106)	(0.1096)	(0.1102)	(0.1076)	(0.1055)	(0.1132)	(0.1113)	(0.1163)	(0.3357)	(0.2854)
	[0.0122]	[0.0120]	[0.0122]	[0.0116]	[0.0111]	[0.0135]	[0.0131]	[0.0222]	[0.5964]	[4.8939]
0.4	0.0013	-0.0004	0.0036	0.0019	0.0053	0.0262	-0.0249	-0.0909	-0.7493	-2.2018
	(0.1072)	(0.1081)	(0.1104)	(0.1109)	(0.1107)	(0.1083)	(0.1095)	(0.1195)	(0.2380)	(0.2728)
	[0.0115]	[0.0117]	[0.0122]	[0.0123]	[0.0123]	[0.0124]	[0.0126]	[0.0226]	[0.6181]	[4.9224]
0.8	0.0021	0.0030	0.0041	0.0015	0.0006	0.0273	-0.0205	-0.0935	-0.7543	-2.2222
	(0.1087)	(0.1024)	(0.1045)	(0.1088)	(0.1113)	(0.1125)	(0.1034)	(0.1164)	(0.2481)	(0.2637)
	[0.0118]	[0.0105]	[0.0109]	[0.0118]	[0.0124]	[0.0134]	[0.0111]	[0.0223]	[0.6306]	[5.0076]
1.5	-0.0011	0.0031	0.0050	0.0003	0.0027	0.0232	-0.0211	-0.0879	-0.7240	-2.2173
	(0.1088)	(0.1117)	(0.1051)	(0.1087)	(0.1120)	(0.1110)	(0.1126)	(0.1187)	(0.2797)	(0.2510)
	[0.0118]	[0.0125]	[0.0111]	[0.0118]	[0.0126]	[0.0129]	[0.0131]	[0.0218]	[0.6024]	[4.9795]
3.0	0.0009	0.0081	-0.0061	0.0006	0.0002	0.0255	-0.0162	-0.0991	-0.6874	-2.2059
	(0.1094)	(0.1045)	(0.1110)	(0.1092)	(0.1088)	(0.1113)	(0.1053)	(0.1194)	(0.3502)	(0.2798)
	[0.0120]	[0.0110]	[0.0124]	[0.0119]	[0.0118]	[0.0130]	[0.0114]	[0.0241]	[0.5952]	[4.9442]

Note: Results for $(1-L)^{d_1}(1-2\cos\frac{\pi}{2}L+L^2)^{d_2}X_t = u_t I(t \geq 1)$, $u_t \sim N(0,1)$. The first number in each cell is the bias, the second is the standard deviation (in round brackets) and the third is the MSE (in square brackets).

subsection are also included. The performance of the MELW estimator is again significantly better, especially as d_1 gets larger and Assumption A.3 is not satisfied for the individual ELW.

Tables 6-7 show the results of the MELW estimation of d_2 when an AR(2) component with a spectral peak at $\pi/2$ is included (the estimation of d_1 is not affected by the AR(2) and the results are thus similar to those with $\Phi(L) = 1$) . Table 6 shows that a significant bias arises with $m = 32$. In fact, for low values of d_1 the individual ELW estimation performs better than the MELW estimator because the positive bias caused by the $AR(2)$ is offset by the negative bias induced by the spectral pole at the origin. However, for large values of d_1 the effect of that spectral pole predominates and the individual ELW behaves much worse than the MELW. The bias is controlled by reducing the bandwidth, as can be observed in Table 7, which shows the results with a lower $m = 8$.

Table 5: MELW finite sample results for d_2, $\Phi(L) = 1$, $n = 512$, $m = 32$

	MELW					ELW				
$d_1 \backslash d_2$	-1.5	0.4	0.8	1.5	3.0	-1.5	0.4	0.8	1.5	3.0
-1.5	0.0240	0.0237	0.0202	0.0235	0.0217	-0.0059	-0.0050	-0.0094	-0.0056	-0.0079
	(0.0767)	(0.0770)	(0.0738)	(0.0722)	(0.0748)	(0.0784)	(0.0786)	(0.0765)	(0.0748)	(0.0777)
	[0.0065]	[0.0065]	[0.0059]	[0.0058]	[0.0061]	[0.0062]	[0.0062]	[0.0059]	[0.0056]	[0.0061]
0.4	0.0216	0.0233	0.0288	0.0204	0.0200	-0.0087	-0.0071	-0.0014	-0.0098	-0.0106
	(0.0796)	(0.0730)	(0.0744)	(0.0769)	(0.0756)	(0.0793)	(0.0721)	(0.0740)	(0.0769)	(0.0750)
	[0.0068]	[0.0059]	[0.0064]	[0.0063]	[0.0061]	[0.0064]	[0.0053]	[0.0055]	[0.0060]	[0.0057]
0.8	0.0214	0.0207	0.0260	0.0222	0.0225	-0.0803	-0.0766	-0.0715	-0.0758	-0.0743
	(0.0789)	(0.0750)	(0.0770)	(0.0754)	(0.0776)	(0.0929)	(0.0883)	(0.0924)	(0.0904)	(0.0918)
	[0.0067]	[0.0061]	[0.0066]	[0.0062]	[0.0065]	[0.0151]	[0.0137]	[0.0136]	[0.0139]	[0.0139]
1.5	0.0219	0.0255	0.0256	0.0259	0.0217	-0.8251	-0.8151	-0.8256	-0.8255	-0.8117
	(0.0763)	(0.0787)	(0.0778)	(0.0797)	(0.0735)	(0.2281)	(0.2276)	(0.2192)	(0.2211)	(0.2202)
	[0.0063]	[0.0068]	[0.0067]	[0.0070]	[0.0059]	[0.7328]	[0.7162]	[0.7296]	[0.7303]	[0.7073]
3.0	0.0204	0.0216	0.0237	0.0201	0.0201	-2.4546	-2.4370	-2.4529	-2.4608	-2.4374
	(0.0748)	(0.0774)	(0.0741)	(0.0810)	(0.0752)	(0.2332)	(0.2415)	(0.2428)	(0.2279)	(0.2402)
	[0.0060]	[0.0065]	[0.0061]	[0.0070]	[0.0061]	[6.0792]	[5.9975]	[6.0758]	[6.1075]	[5.9988]

Note: Results for $(1-L)^{d_1}(1-2\cos\frac{\pi}{2}L+L^2)^{d_2}X_t = u_t I(t \geq 1)$, $u_t \sim N(0,1)$. The first number in each cell is the bias, the second is the standard deviation (in round brackets) and the third is the MSE (in square brackets).

5.3 Deterministic components

In order to analyse the effects of deterministic components consider processes of the form

$$Y_t = \mu \cos\left(\frac{\pi}{2}t\right) + X_t\ ,\ \ (1-L)^{d_1}(1+L^2)^{d_2}X_t = u_t I(t \geq 1),\ \ t = 0, 1, ..., n$$

where $\mu = 100$ and the u_t are independent standard normal such that Y_t contains stochastic persistence at frequencies 0 and $\pi/2$, together with a deterministic cycle of frequency $\pi/2$ or period 4. Table 8 shows that the presence of the deterministic cycle has an adverse effect especially on the estimation of the memory parameter d_2 at frequency $\pi/2$, which coincides with the frequency of the deterministic component. This is theoretically explained by Lemma 10 in the supplement, which shows a stronger effect of the DFT of the deterministic cycle at $\pi/2$. In fact, the estimation of d_1 seems to be unaffected when the value of d_1 is positive and much larger than d_2.

Tables 9 and 10 show the finite sample performance of the MELW estimation when

Table 6: MELW finite sample results for d_2, $\Phi(L) = 1 - 0.6L^2$, $n = 512$, $m = 32$

	MELW					ELW				
$d_1 \backslash d_2$	-1.5	0.4	0.8	1.5	3.0	-1.5	0.4	0.8	1.5	3.0
-1.5	0.2336	0.2335	0.2282	0.2299	0.2318	0.1960	0.1976	0.1913	0.1943	0.1949
	(0.0812)	(0.0819)	(0.0793)	(0.0773)	(0.0828)	(0.0819)	(0.0828)	(0.0810)	(0.0790)	(0.0850)
	[0.0612]	[0.0612]	[0.0583]	[0.0588]	[0.0606]	[0.0451]	[0.0459]	[0.0432]	[0.0440]	[0.0452]
0.4	0.2285	0.2356	0.2350	0.2320	0.2294	0.1925	0.1997	0.1987	0.1954	0.1939
	(0.0827)	(0.0778)	(0.0772)	(0.0802)	(0.0772)	(0.0819)	(0.0767)	(0.0757)	(0.0785)	(0.0751)
	[0.0591]	[0.0616]	[0.0612]	[0.0603]	[0.0586]	[0.0438]	[0.0458]	[0.0452]	[0.0444]	[0.0433]
0.8	0.2283	0.2342	0.2302	0.2321	0.2307	0.1622	0.1638	0.1632	0.1622	0.1599
	(0.0763)	(0.0781)	(0.0775)	(0.0799)	(0.0819)	(0.0761)	(0.0808)	(0.0779)	(0.0803)	(0.0830)
	[0.0579]	[0.0609]	[0.0590]	[0.0603]	[0.0599]	[0.0321]	[0.0334]	[0.0327]	[0.0328]	[0.0325]
1.5	0.2322	0.2337	0.2345	0.2309	0.2295	-0.5347	-0.5315	-0.5409	-0.5162	-0.5338
	(0.0834)	(0.0788)	(0.0791)	(0.0788)	(0.0769)	(0.2369)	(0.2389)	(0.2284)	(0.2301)	(0.2389)
	[0.0609]	[0.0608]	[0.0613]	[0.0595]	[0.0586]	[0.3420]	[0.3396]	[0.3448]	[0.3194]	[0.3420]
3.0	0.2304	0.2313	0.2269	0.2329	0.2353	-2.1849	-2.1777	-2.1725	-2.1720	-2.1849
	(0.0834)	(0.0767)	(0.0804)	(0.0798)	(0.0762)	(0.2382)	(0.2311)	(0.2427)	(0.2344)	(0.2289)
	[0.0600]	[0.0594]	[0.0580]	[0.0606]	[0.0612]	[4.8303]	[4.7956]	[4.7788]	[4.7727]	[4.8263]

Note: Results for $(1-L)^{d_1}(1-2\cos\frac{\pi}{2}L+L^2)^{d_2}X_t = u_t I(t \geq 1)$, $(1-0.6L^2)u_t = \epsilon_t$, $\epsilon_t \sim N(0,1)$. The first number in each cell is the bias, the second is the standard deviation (in round brackets) and the third is the MSE (in square brackets).

it is applied to the series filtered of deterministic components. Two different filters are considered: First, the series is deseasonalised as $Y_t - \hat{\mu}\cos(0.5\pi t)$ with $\hat{\mu}$ obtained by least squares of Y_t on $\cos(0.5\pi t)$. Second the series is filtered by substracting the first four observations (one cycle) as $Y_{4k+s} - Y_s$, $s = 1, 2, 3, 4$, $k = 0, 1, ..., n/4 - 1$. The first option is comparable to the estimation of the mean and the second one to subtracting the first observation as proposed by Shimotsu (2010) in ELW estimation in standard long memory.

Tables 9 shows results of the estimation of d_1. Using $\hat{\mu}$ is a good option when $\max(d_1, d_2 + 1) < 2$, which is comparable with the results obtained by Shimotsu (2010) who showed that the sample mean leads to a consistent and asymptotically normal ELW estimator of a memory parameter at the origin lower than one. On the other hand, subtracting the first four observations seems to be a good option for positive values of d_1. A similar situation can be observed in Table 10, which shows the results for the estimation of d_2. The only exception is

Table 7: MELW finite sample results for d_2, $\Phi(L) = 1 - 0.6L^2$, $n = 512$, $m = 8$

	MELW					ELW				
$d_1 \backslash d_2$	-1.5	0.4	0.8	1.5	3.0	-1.5	0.4	0.8	1.5	3.0
-1.5	0.0182	0.0138	0.0083	0.0251	0.0236	0.0065	0.0069	0.0019	0.0172	0.0167
	(0.2038)	(0.2084)	(0.2138)	(0.2131)	(0.2047)	(0.2019)	(0.1998)	(0.2062)	(0.2084)	(0.1981)
	[0.0419]	[0.0436]	[0.0458]	[0.0460]	[0.0425]	[0.0408]	[0.0400]	[0.0425]	[0.0437]	[0.0395]
0.4	0.0093	0.0256	0.0197	0.0173	0.0122	0.0005	0.0195	0.0070	0.0060	0.0077
	(0.2008)	(0.2015)	(0.2069)	(0.2120)	(0.2135)	(0.1989)	(0.1941)	(0.2008)	(0.2114)	(0.2081)
	[0.0404]	[0.0413]	[0.0432]	[0.0452]	[0.0458]	[0.0395]	[0.0381]	[0.0404]	[0.0447]	[0.0433]
0.8	0.0325	0.0082	0.0162	0.0176	0.0219	0.0057	-0.0165	-0.0016	-0.0130	-0.0043
	(0.2153)	(0.2044)	(0.1984)	(0.2141)	(0.2110)	(0.1966)	(0.1886)	(0.1892)	(0.1996)	(0.1971)
	[0.0474]	[0.0418]	[0.0396]	[0.0461]	[0.0450]	[0.0387]	[0.0358]	[0.0358]	[0.0400]	[0.0389]
1.5	0.0202	0.0240	0.0171	0.0179	0.0111	-0.5510	-0.5476	-0.5560	-0.5351	-0.5488
	(0.2011)	(0.2050)	(0.2078)	(0.2085)	(0.2122)	(0.2102)	(0.2138)	(0.2026)	(0.1989)	(0.2147)
	[0.0408]	[0.0426]	[0.0435]	[0.0438]	[0.0451]	[0.3478]	[0.3456]	[0.3502]	[0.3258]	[0.3473]
3.0	0.0053	0.0234	0.0124	0.0193	0.0250	-2.1433	-2.1355	-2.1307	-2.1299	-2.1432
	(0.2045)	(0.1999)	(0.2041)	(0.2218)	(0.2152)	(0.2398)	(0.2318)	(0.2442)	(0.2353)	(0.2304)
	[0.0418]	[0.0405]	[0.0418]	[0.0496]	[0.0470]	[4.6513]	[4.6139]	[4.5994]	[4.5920]	[4.6466]

Note: Results for $(1-L)^{d_1}(1 - 2\cos\frac{\pi}{2}L + L^2)^{d_2}X_t = u_t I(t \geq 1)$, $(1 - 0.6L^2)u_t = \epsilon_t$, $\epsilon_t \sim N(0,1)$. The first number in each cell is the bias, the second is the standard deviation (in round brackets) and the third is the MSE (in square brackets).

when $d_1 = 1.5$ and $d_2 = -0.4$, such that the difference between both parameters is so large that the strong pole at the origin exerts a positive bias in the estimation of the negative d_2.

5.4 Inference: Application for unit roots testing

The asymptotic distribution of the MELW estimator in Theorem 2 enables standard inference techniques to be applied to test hypotheses of interest on the different memory parameters. For example, the existence of common unit roots at seasonal frequencies and the origin should be tested before the seasonal difference operator $(1 - L^S)$ is applied. This section analyses the performance of a Wald type test based on the MELW estimator and its asymptotic distribution for such a hypothesis. Consider the process

$$(1-L)^{d_1}(1+L^2)^{d_2}(1+L)^{d_3}X_t = u_t I(t \geq 1), \quad t = 0, 1, ..., \tag{11}$$

Table 8: MELW estimation of d_1 and d_2 with deterministic components, $n = 512$, $m = 32$

	MELW of d_1					MELW of d_2				
$d_1 \backslash d_2$	-0.4	0.4	0.8	1.5	3.0	-0.4	0.4	0.8	1.5	3.0
-0.4	0.3514	0.4029	0.1418	0.0184	0.4497	0.7033	-0.3103	-0.3441	-0.5440	-0.2471
	(0.1926)	(0.0163)	(0.4420)	(0.5005)	(0.0116)	(0.4616)	(0.2875)	(0.5012)	(0.2762)	(0.0713)
	[0.1605]	[0.1626]	[0.2155]	[0.2508]	[0.2024]	[0.7078]	[0.1789]	[0.3696]	[0.3723]	[0.0661]
0.4	-0.2814	-0.2624	-0.3706	-0.3698	-0.3386	0.7324	0.2983	-0.7225	-0.4189	-0.3325
	(0.1646)	(0.1272)	(0.0805)	(0.1996)	(0.0310)	(0.4791)	(0.4747)	(0.2636)	(0.2008)	(0.0766)
	[0.1063]	[0.0850]	[0.1438]	[0.1766]	[0.1156]	[0.7659]	[0.3143]	[0.5915]	[0.2158]	[0.1164]
0.8	-0.0830	-0.1875	-0.5928	-0.3718	-0.5934	1.4052	0.6162	-0.7382	-0.4404	-0.3738
	(0.1408)	(0.1090)	(0.0908)	(0.1040)	(0.1045)	(0.1486)	(0.0990)	(0.2346)	(0.1895)	(0.0781)
	[0.0267]	[0.0470]	[0.3597]	[0.1491]	[0.3630]	[1.9967]	[0.3895]	[0.6000]	[0.2299]	[0.1459]
1.5	-0.1152	-0.0688	-0.7063	-0.2073	-0.4689	1.2874	0.6155	-0.6744	-0.4241	-0.4406
	(0.2982)	(0.1133)	(0.2594)	(0.1052)	(0.1093)	(0.3359)	(0.0111)	(0.3265)	(0.0695)	(0.0731)
	[0.1022]	[0.0176]	[0.5661]	[0.0540]	[0.2318]	[1.7702]	[0.3790]	[0.5614]	[0.1847]	[0.1995]
3.0	-0.0094	-0.0333	-0.1297	-0.0751	-0.1311	0.9550	0.5131	-0.7060	-0.5149	-0.6352
	(0.0858)	(0.1035)	(0.2116)	(0.1103)	(0.1116)	(0.4671)	(0.2081)	(0.2573)	(0.0907)	(0.0864)
	[0.0075]	[0.0118]	[0.0616]	[0.0178]	[0.0296]	[1.1301]	[0.3066]	[0.5646]	[0.2733]	[0.4109]

Note: Results for $Y_t = \mu \cos\left(\frac{\pi}{2}t\right) + X_t$, $\mu = 100$, $(1-L)^{d_1}(1 - 2\cos\frac{\pi}{2}L + L^2)^{d_2} X_t = u_t I(t \geq 1)$, $u_t \sim N(0,1)$. The first number in each cell is the bias, the second is the standard deviation (in round brackets) and the third is the MSE (in square brackets).

for $w_1 = 0$, $w_2 = \pi/2$ and $w_3 = \pi$. Unit roots at w_1, w_2 and w_3 imply $d_1 = d_2 = d_3 = 1$ such that X_t in (11) can be written as $(1 - L^4)X_t = u_t I(t \geq 1)$. To analyse the size and power of the Wald tests described in Remark 2, 1000 series of 512 and 1024 observations were generated with u_t standard normal, $d_1 = 1$ and $d_2, d_3 \in \{0.6, 0.8, 1.0, 1.2, 1.4\}$. Table 11 shows the rejection frequencies for a 5% significance level of the test in Remark 2 with $R = diag\{1,1,1\}$ and $r = (1,1,1)'$, using the variance in the asymptotic distribution and the Hessian based approximation in Remark 3 with $m = n/16$. The use of the variance in the asymptotic distribution leads to over-rejection, which is significantly corrected by the use of the Hessian approximation, although there is still some oversize. Power increases significantly with the sample size and is greater for positive departures from the null, mainly in the d_2 dimension.

Table 9: MELW-corrected estimation of d_1 with deterministic component $n = 512$ $m = 32$

	$\hat{\mu}\cos\left(\frac{\pi}{2}t\right)$					(Y_1, Y_2, Y_3, Y_4)				
$d_1 \backslash d_2$	-0.4	0.4	0.8	1.5	3.0	-0.4	0.4	0.8	1.5	3.0
-0.4	0.0020	-0.0034	0.0252	0.0870	0.4084	0.2755	0.3013	0.3183	0.3756	0.6384
	(0.1106)	(0.1102)	(0.1559)	(0.4613)	(0.0192)	(0.0977)	(0.0930)	(0.0946)	(0.1106)	(0.4465)
	[0.0122]	[0.0122]	[0.0250]	[0.2204]	[0.1671]	[0.0854]	[0.0994]	[0.1103]	[0.1533]	[0.6069]
0.4	0.0014	-0.0003	0.0023	-0.2526	-1.0304	0.0080	-0.0024	0.0005	-0.0041	-0.0062
	(0.1072)	(0.1081)	(0.1106)	(0.2277)	(0.4867)	(0.1104)	(0.1174)	(0.1211)	(0.1332)	(0.3018)
	[0.0115]	[0.0117]	[0.0122]	[0.1157]	[1.2987]	[0.0123]	[0.0138]	[0.0147]	[0.0178]	[0.0911]
0.8	0.0029	0.0030	0.0033	-0.2110	-1.2669	0.0099	0.0022	0.0045	0.0012	-0.0282
	(0.1087)	(0.1024)	(0.1046)	(0.1877)	(0.5041)	(0.1088)	(0.1028)	(0.1044)	(0.1076)	(0.1472)
	[0.0118]	[0.0105]	[0.0109]	[0.0798]	[1.8592]	[0.0119]	[0.0106]	[0.0109]	[0.0116]	[0.0225]
1.5	0.0047	0.0028	0.0040	-0.0881	-1.7748	0.0039	0.0047	0.0052	-0.0020	-0.0792
	(0.1084)	(0.1121)	(0.1051)	(0.1309)	(0.4818)	(0.1113)	(0.1116)	(0.1046)	(0.1073)	(0.1247)
	[0.0118]	[0.0126]	[0.0111]	[0.0249]	[3.3820]	[0.0124]	[0.0125]	[0.0110]	[0.0115]	[0.0218]
3.0	-0.2743	-0.2659	-0.2927	-0.2943	-1.3173	-0.0510	-0.0437	-0.1114	-0.1344	-0.2016
	(0.1924)	(0.1885)	(0.1910)	(0.1823)	(0.3342)	(0.1108)	(0.1023)	(0.1202)	(0.1284)	(0.1510)
	[0.1122]	[0.1062]	[0.1221]	[0.1198]	[1.8471]	[0.0149]	[0.0124]	[0.0269]	[0.0346]	[0.0634]

Note: Results for $Y_t = \mu\cos\left(\frac{\pi}{2}t\right) + X_t$, $\mu = 100$, $(1-L)^{d_1}(1-2\cos\frac{\pi}{2}L + L^2)^{d_2}X_t = u_t I(t \geq 1)$, $u_t \sim N(0,1)$. The first number in each cell is the bias, the second is the standard deviation (in round brackets) and the third is the MSE (in square brackets).

6 Empirical application: U.S. Industrial Production Index

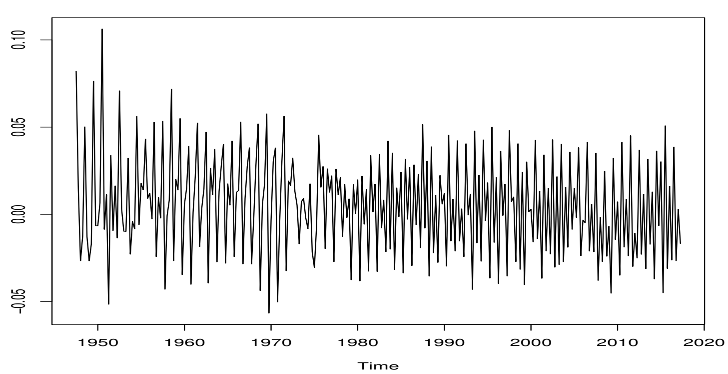
Seasonality is a common characteristic in many economic time series of higher than yearly frequency. Deterministic components, seasonal unit roots and weak dependent seasonal models have often been used with no consensus as to their suitability (see for example Beaulieu and Miron, 1993 or Hylleberg et al. 1993). This section analyses a quarterly series of the Industrial Production Index for non-durable consumer goods in the USA with base year 2012. We consider the growth rate obtained by first differencing the logarithm of the index, thus avoiding the potential distorting effects at frequency zero of a deterministic trend. The data, displayed in Figure 4, span the period from 1947Q3 to 2017Q2.

Table 10: MELW-corrected estimation of d_2 with deterministic component $n = 512$ $m = 32$

	$\hat{\mu}\cos\left(\frac{\pi}{2}t\right)$					(Y_1, Y_2, Y_3, Y_4)				
$d_1 \backslash d_2$	-0.4	0.4	0.8	1.5	3.0	-0.4	0.4	0.8	1.5	3.0
-0.4	0.0218	0.0251	0.0273	-0.3461	-1.6836	0.3094	0.0283	0.0247	0.0183	-0.0687
	(0.0787)	(0.0766)	(0.0743)	(0.2610)	(0.2065)	(0.0619)	(0.0819)	(0.0753)	(0.0776)	(0.1013)
	[0.0067]	[0.0065]	[0.0063]	[0.1879]	[2.8773]	[0.0995]	[0.0075]	[0.0063]	[0.0064]	[0.0150]
0.4	0.0209	0.0242	0.0376	-0.2281	-1.9956	0.3236	0.0230	0.0281	0.0196	-0.0252
	(0.0792)	(0.0726)	(0.0733)	(0.1421)	(0.3902)	(0.0568)	(0.0780)	(0.0749)	(0.0769)	(0.0842)
	[0.0067]	[0.0059]	[0.0068]	[0.0722]	[4.1349]	[0.1079]	[0.0066]	[0.0064]	[0.0063]	[0.0077]
0.8	0.0357	0.0218	0.0346	-0.2231	-1.9090	0.3380	0.0218	0.0260	0.0223	-0.0221
	(0.0760)	(0.0744)	(0.0762)	(0.1373)	(0.3614)	(0.0606)	(0.0829)	(0.0771)	(0.0753)	(0.0826)
	[0.0070]	[0.0060]	[0.0070]	[0.0686]	[3.7750]	[0.1179]	[0.0073]	[0.0066]	[0.0062]	[0.0073]
1.5	0.3387	0.0286	0.0337	-0.2265	-1.7538	0.3923	0.0276	0.0285	0.0222	-0.0597
	(0.0868)	(0.0821)	(0.0763)	(0.1401)	(0.2641)	(0.1138)	(0.1122)	(0.0777)	(0.0782)	(0.0961)
	[0.1223]	[0.0075]	[0.0070]	[0.0709]	[3.1456]	[0.1669]	[0.0134]	[0.0068]	[0.0066]	[0.0128]
3.0	1.3596	0.5814	0.1200	-0.5054	-1.6931	0.7369	-0.0107	0.0061	-0.1254	-0.2347
	(0.1641)	(0.0704)	(0.2509)	(0.0752)	(0.2150)	(0.3884)	(0.2984)	(0.1360)	(0.1217)	(0.1630)
	[1.8755]	[0.3430]	[0.0773]	[0.2611]	[2.9128]	[0.6938]	[0.0892]	[0.0185]	[0.0305]	[0.0816]

Note: Results for $Y_t = \mu\cos\left(\frac{\pi}{2}t\right) + X_t$, $\mu = 100$, $(1-L)^{d_1}(1-2\cos\frac{\pi}{2}L + L^2)^{d_2}X_t = u_t I(t \geq 1)$, $u_t \sim N(0,1)$. The first number in each cell is the bias, the second is the standard deviation (in round brackets) and the third is the MSE (in square brackets).

Figure 4: IPI returns, 1947Q3 to 2017Q2



We first test for the possibility of seasonal unit roots using conventional tools, in particular the HEGY test for unit roots (Hylleberg et al. 1990) and the test for seasonal stability in Canova and Hansen (1995). Table 12 shows the test statistics and p values of both strategies with the lags in HEGY selected by the BIC and including seasonal dum-

Table 11: Rejection frequencies for $H_0 : (d_1, d_2, d_3) = (0.5, 1.0, 0.5)$

		$n = 512$, $m = 32$					$n = 1024$, $m = 64$				
$d_2 \backslash d_3$		0.6	0.8	1.0	1.2	1.4	0.6	0.8	1.0	1.2	1.4
0.6	Asy.	1.000	0.999	0.996	0.999	1.000	1.000	1.000	1.000	1.000	1.000
	Hess.	1.000	0.986	0.978	0.996	1.000	1.000	1.000	1.000	1.000	1.000
0.8	Asy.	0.988	0.783	0.577	0.819	0.991	1.000	0.979	0.859	0.987	1.000
	Hess.	0.957	0.615	0.379	0.692	0.969	1.000	0.949	0.775	0.970	1.000
1.0	Asy.	0.961	0.557	0.190	0.599	0.964	1.000	0.747	0.172	0.830	1.000
	Hess.	0.909	0.358	0.087	0.400	0.913	0.999	0.650	0.102	0.747	0.999
1.2	Asy.	0.997	0.955	0.899	0.965	1.000	1.000	1.000	0.996	0.999	1.000
	Hess.	0.989	0.909	0.806	0.917	0.995	1.000	0.996	0.987	0.999	1.000
1.4	Asy.	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Hess.	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note: Rejection frequencies for $H_0 : (d_1, d_2, d_3) = (1, 1, 1)$ using the variance in the asymptotic distribution (Assy.) and the Hessian based approximation (Hess.). The generating process is $(1 - L)(1 + L^2)^{d_2}(1 + L)^{d_3} X_t = u_t I(t \geq 1)$, for $u_t \sim NID(0, 1)$.

mies. The Canova-Hansen test rejects stability at π and $\pi/2$, whereas HEGY rejects the existence of unit roots at the origin and seasonal frequencies, suggesting the possibility of seasonal long memory. Table 12 also shows MELW estimates of the memory parameters at the origin, seasonal $\pi/2$ and Nyquist π frequencies together with the 95% confidence intervals obtained with the Hessian approximation of the variance introduced in Remark 3. Adjusted MELW estimation for deterministic cycles using the residuals from least squares regression on $\cos(\bar{w}t)$, $\bar{w} = 0, \pi/2, \pi$ and subtracting the first four observations are also included.

Focusing on the MELW estimator unadjusted for deterministic components, there is clear evidence of a unit root at the yearly frequency $\pi/2$, long memory at the origin and weak dependence at the Nyquist frequency. When the series is adjusted from deterministic components, the main difference with the unadjusted MELW estimator is the larger estimate obtained at π, implying strong persistence also at that frequency. Considering the possibility of a common unit root, the Wald statistic rejects the hypothesis $(d_1, d_2, d_3) = (1, 1, 1)$ with large $p - values$, suggesting that the seasonal difference operator $(1 - L^4)$ should not

Table 12: Quarterly U.S. Industrial Production Index: 1947Q3 to 2017Q2

	$d_1(w_1=0)$	$d_2(w_2=\frac{\pi}{2})$	$d_3(w_3=\pi)$	Wald St.
		$m_1=m_2=m_3=15$		
MELW	0.468	0.982	0.075	38.372
	(0.131,0.805)	(0.742, 1.222)	(-0.262, 0.413)	
MELW($\hat{\mu}$)	0.371	0.789	0.403	28.315
MELW($Y_1, ..., Y_4$)	0.259	0.924	0.379	31.846
		Canova-Hansen		
$testst.$		2.223	1.049	
$p-value$		0.000	0.001	
		HEGY		
$testst.$	-5.528	16.987	-5.643	
$p-value$	0.000	0.000	0.000	

Note: MELW and ELW estimates, 95% confidence intervals (in round brackets) and Wald statistic for the unit root hypothesis $H_0 : (d_1, d_2, d_3) = (1, 1, 1)$.

be used. Note that the same asymptotic distribution is used here for the MELW adjusted for deterministic components with no theoretical justification. Taking into account the potential size and power distortions caused by short memory components in HEGY and Canova-Hansen tests (see Canova and Hansen, 1995 or Ghysels et al., 1994) the results with the MELW estimator can be considered as more reliable.

References

[1] Abadir, K.M., Distaso, W. and L. Giraitis (2007) Nonstationarity-extended Local Whittle Estimation. *Journal of Econometrics* 141, 1353-1384.

[2] Abadir, K.M., Caggiano, G. and G. Talmain (2013) Nelson–Plosser revisited: The ACF approach. *Journal of Econometrics* 175, 22-34.

[3] Arteche, J. (2000) Gaussian semiparametric estimation in seasonal/cyclical long memory time series. *Kybernetika* 36, 279-310.

[4] Arteche, J. (2002) Semiparametric robust tests on seasonal or cyclical long memory time series. *Journal of Time Series Analysis* 23, 251-286.

[5] Arteche, J. and P.M. Robinson (1999) Seasonal and cyclical long memory. In *Asymptotics, Nonparametrics and Time Series* (ed S. Ghosh). New York: Marcel Dekker, Inc., 115-148.

[6] Arteche, J. and P.M. Robinson (2000) Semiparametric Inference in Seasonal and Cyclical Long Memory Processes. *Journal of Time Series Analysis*, 21, 1-25.

[7] Arteche, J. and C. Velasco (2005) Trimming and tapering semiparametric estimates in asymmetric long memory time series. *Journal of Time Series Analysis* 26, 581-611.

[8] Beaulieu, J.J, and J.A. Miron (1993) Seasonal unit roots in aggregate U.S. data. *Journal of Econometrics* 55, 305-328.

[9] Canova, F. and B.E. Hansen (1995) Are Seasonal Patterns Constant over Time? A Test for Seasonal Stability. *Journal of Business and Economic Statistics* 13, 237-252.

[10] Chan, N.H. and Terrin, N. (1995) Inference for unstable long-memory processes with applications to fractional unit roots autorregressions. *Annals of Statistics* 23, 1662-1683.

[11] Chan, N.H. and Wei, C.Z. (1988) Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics* 16, 367-401.

[12] Fox, R. and M.S. Taqqu (1986) Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series. *Annals of Statistics* 14, 517-532.

[13] Ghysels, E., Lee, H.S. and J. Noh (1994) Testing for unit roots in seasonal time series: Some theoretical extensions and a Monte Carlo investigation. *Journal of Econometrics* 62, 415-442.

[14] Giraitis, L. and R. Leipus (1995) A generalized fractionally differencing approach in long-memory modeling. *Lithuanian Mathematical Journal*, 35, 53-65.

[15] Giraitis, L., Hidalgo, J. and P. M. Robinson (2001) Gaussian estimation of parametric spectral density with unknown pole. *The Annals of Statistics* 29, 987-1023.

[16] Granger, C.W.J. (1966) The Typical Spectral Shape of an Economic Variable. *Econometrica* 34, 150-161.

[17] Hassler, U., Rodrigues, P.M.M. and A. Rubia (2009) Testing for general fractional integration in the time domain. *Econometric Theory* 25, 1793-1828.

[18] Hidalgo, J. (2005) Semiparametric Estimation for Stationary Processes whose Spectra have an Unknown Pole. *Annals of Statistics* 33, 1843-1889.

[19] Hidalgo, J. and P. Soulier (2004) Estimation of the location and exponent of the spectral singularity of a long memory process. *Journal of Time Series Analysis*, 25, 55-81.

[20] Hosking, J. R. M. (1984) Modeling persistence in hydrological time series using fractional differencing. *Water Resources Research* 20, 1898–1908.

[21] Hylleberg, S., Engle, R.F., Granger, C.W.J. and B.S. Yoo (1990) Seasonal Integration and Cointegration. *Journal of Econometrics* 44, 215-238.

[22] Hylleberg, S., Jørgensen, C. and N.K. Sørensen (1993) Seasonality in macroeconomic time series. *Empirical Economics* 18, 321-335.

[23] Hurvich, C.M. and W.W. Chen (2000) An efficient taper for potentially overdifferenced long memory time series, *Journal of Time Series Analysis* 21, 155–80.

[24] Nielsen, M. Ø. (2004) Efficient likelihood inference in nonstationary univariate models. *Econometric Theory* 20, 116-146.

[25] Porter-Hudak, S. (1990) An application to the seasonally fractionally differenced model to the monetary aggregates. *Journal of the American Statistical Association* 85, 338-344.

[26] Ray B.K. (1993) Long-range forecasting of IBM product revenues using a seasonal fractionally differenced ARMA model. *International Journal of Forecasting* 9, 255-269.

[27] Robinson, P.M. (1994) Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association* 89, 1420-1436.

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

[29] Robinson, P.M. (2005) The distance between rival nonstationary fractional processes. *Journal of Econometrics*, 128, 283-300.

[30] Shao, X. (2010) Nonstationarity-extended Whittle estimation. *Econometric Theory* 26, 1060-1087.

[31] Shao, X. and W.B. Wu (2007) Local Whittle estimation of fractional integration for nonlinear processes. *Econometric Theory* 23, 899-929.

[32] Shimotsuu, K. (2010) Exact Local Whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory* 26, 501-540.

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

[34] Shimotsu, K. and P.C.B. Phillips (2006) Local Whittle estimation of fractional integration and some of its variants. *Journal of Econometrics* 103, 209–233.

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

[36] Woodward, W.A., Cheng, Q.C. and H.L. Gray (1998) A k-factor GARMA long memory model. *Journal of Time Series Analysis* 19, 485-504.