Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2008-28

Local polynomial Whittle estimation covering non-stationary fractional processes

Frank S. Nielsen

School of Economics and Management
University of Aarhus
Building 1322, DK-8000 Aarhus C
Denmark

Local polynomial Whittle estimation covering non-stationary fractional processes*

Frank S. Nielsen†
Aarhus University and CREATES

March 23, 2009

Abstract

This paper extends the local polynomial Whittle estimator of Andrews & Sun (2004) to fractionally integrated processes covering both stationary and non-stationary regions. We utilize the notion of the extended discrete Fourier transform and periodogram to extend the local polynomial Whittle estimator to the non-stationary region. We further, approximate the short-run component of the spectrum by a polynomial instead of a constant in a shrinking neighborhood of zero, and thereby alleviate some of the bias that the local Whittle estimator is prone to. This bias reduction comes at a cost as the variance is inflated by a multiplicative constant. We show consistency and asymptotic normality for $d \in (-1/2, \infty)$, and if the spectral density of the short-run component is infinitely smooth near frequency zero, we obtain a rate of convergence arbitrarily close to the parametric rate. A simulation study illustrates the performance of the proposed estimator compared to the classical local Whittle estimator and the local polynomial Whittle estimator. The empirical justification of the proposed estimator is shown through an analysis of credit spreads.

Keywords: Bias reduction, fractional integration, local polynomial, local Whittle estimation, long memory.

JEL Classification: C22

*I am grateful to Niels Haldrup, Javier Hualde, and Morten Ørregaard Nielsen for valuable suggestions and comments. This work was partly done while I was visiting Cornell University, their hospitality is gratefully acknowledged. I would also like to thank Thomas Stephansen for proofreading the manuscript. I greatly acknowledge financial support from the Danish Social Sciences Researh Council (grant no. FSE275-05-0199) and Center for Research in Econometric Analysis of Time Series (CREATES), funded by the Danish National Research Foundation.

†CREATES, School of Economics and Management, Aarhus University, Building 1322, DK-8000 Aarhus C, Denmark. email: `fnielsen@creates.au.dk`.

1 Introduction

We are interested in semiparametric frequency-domain estimation based on the local approximation

$$f(\lambda) \sim \varphi(\lambda)\lambda^{-2d} \text{ as } \lambda \to 0^{+}, \tag{1}$$

where $\varphi(0) \in (0, \infty)$ and the symbol "$\sim$" means that the ratio of the left and right hand sides tends to one in the limit. $\varphi(\lambda)$ is an even, positive, continuous function on $[-\pi, \pi)$ which can be thought of as the spectral density of the short-memory component of the series of interest. Semiparametric based estimators have been popular for a long time as it is believed that the loss of efficiency with respect to the parametric estimators entailed by the local specifications may be offset by a possible greater robustness. This robustness stems from avoiding the inconsistency in estimating the long-run dynamics that may be caused by a misspecification of short-run dynamics.

Under stationarity and modeling $\varphi(\lambda)$ in (1) by a constant $G \in (0, \infty)$, a common semiparametric estimator is the local Whittle (LW) estimator proposed by Künsch (1987). Robinson (1995*a*) shows its consistency and asymptotic normality for $d \in (-1/2, 1/2)$. Velasco (1999*a*) extended Robinson's (1995*a*) results to show that the estimator is consistent for $d \in (-1/2, 1)$ and asymptotically normally distributed for $d \in (-1/2, 3/4)$, given that the fractional process is of Type I, see Marinucci & Robinson (1999) and Robinson (2005). Phillips & Shimotsu (2004) show that the LW estimator is consistent for $d \in (1/2, 1]$ and has a nonnormal limit distribution for $d \in (3/4, 1)$, and a mixed normal limit distribution for $d = 1$. When $d > 1$ the LW estimator converges to unity in probability and therefore is inconsistent, given that the fractional process is of Type II, Phillips & Shimotsu (2004). This convergence in probability to unity when $d > 1$ also holds for log periodogram estimators as shown in simulations studies by Hurvich & Ray (1995) and Velasco (1999*b*), and theoretically by Kim & Phillips (2006). That is, in general the LW (or log periodogram) estimator is not a good general purpose estimator when d takes on values in the non-stationary region beyond 3/4. The asymptotic theory is discontinuous at $d \in \{3/4, 1\}$ and the estimator is not consistent for $d > 1$. Several methods are available to avoid the problems when entering the non-stationary region. A simple one is to first difference the series before using the semiparametric estimator and then add one to the estimate. This method runs into problems if the series of interest is trend stationary, Shimotsu & Phillips (2005) and Shimotsu (2006). Tapering the data is another method often implemented and suggested, see Velasco (1999*a*) and Hurvich & Chen (2000).

Shimotsu & Phillips (2005) introduce what they call an exact local Whittle estimator[1] which is consistent and has the same $N(0, 1/4)$ limit distribution for all values of d if the $I(d)$ series is generated by a linear sequence and the range of the estimator is not wider than 9/2.[2] Instead of using fractional differencing of the data, Abadir, Distaso & Giraitis (2007) use a different approach first noted by Phillips (1999). They extend the discrete Fourier transform to the non-stationary case and use this in whitening of the periodogram. Abadir et al. (2007) show that when the $I(d)$ series is generated by a

[1] Shimotsu (2006) extends this to a feasible exact local Whittle estimator when introducing an unknown mean and trend.

[2] The assumption concerning the width of the admissible parameter space is needed to ensure that the difference in the criteria function is uniformly bounded away from zero, see Shimotsu & Phillips (2005).

linear sequence the extended discrete Fourier transform and periodogram have the same asymptotic behavior for $d \in (-3/2, \infty)$.

Our main interest in this paper is to analyze a general purpose estimator where the limiting distribution holds in the non-stationary case and when there is short-run contamination. To achieve bias reduction when there is contamination by short-run dynamics, we follow Andrews & Sun (2004) and model the spectral density of the short-memory component $\varphi(\lambda)$ as a finite and even polynomial instead of a constant near frequency zero. In extending the local polynomial Whittle (LPW) estimator of Andrews & Sun (2004) to the non-stationary region, we use the notion of the extended discrete Fourier transform and periodogram as in Abadir et al. (2007). We call the new estimator the extended local polynomial Whittle (ExtLPW) estimator. In establishing consistency and asymptotic normality for the estimator $\hat{d}$ we follow the method set out by Andrews & Sun (2004). Given that the generating process is linear, the same central limit theorem argument as in the stationary case $|d| < \frac{1}{2}$ derived by Robinson (1995a) holds; although, not for $d_0 = \left\{\frac{1}{2}, \frac{3}{2}, ...\right\}$. We establish consistency and asymptotic normality for $d_0 \in (-1/2, \infty)$. Furthermore, if $\varphi(\lambda)$ is infinitely smooth near frequency zero, the rate of convergence can become arbitrary close to the parametric rate. The simulations reveal that our proposed estimator is superior when considering possible short-run contamination and non-stationary values of d. We also include an analysis of credit spreads that demonstrates the usefulness of the estimator.

The remainder of the paper is structured as follows: Section 2 gives a short introduction to the LPW estimator of Andrews & Sun (2004). Section 3 expands the usual stationary framework to the non-stationary framework thereby defining the ExtLPW estimator. Section 4 states the assumptions needed for showing consistency and asymptotic normality. Section 5 introduces the theorem for consistency and asymptotic normality. Section 6 presents the results from a small simulation study. Section 7 provides an empirical investigation of potential long memory properties of treasury yield and yields on corporate bonds, spreads over treasury and spreads between corporate yields. Section 8 concludes. Lemmas, and proofs to Theorem 1 and Lemma 1-2 are situated in the Appendix.

2 The local polynomial Whittle approach

Define at the jth frequency $\lambda_j = \frac{2\pi j}{n}$ for $1 \leq j \leq m$, the discrete Fourier transform (DFT) and periodogram of X_t as

$$w(\lambda_j) = \frac{1}{\sqrt{(2\pi n)}} \sum_{t=1}^{n} X_t \exp(it\lambda_j) \tag{2}$$

$$I(\lambda_j) = |\omega(\lambda_j)|^2 . \tag{3}$$

Following Andrews & Sun (2004) the (negative) local polynomial Whittle log-likelihood is

$$U_n(d, G, \theta) = \frac{1}{m} \sum_{j=1}^{m} \left[\log(G\lambda_j^{-2d} \exp(-P_r(\lambda_j, \theta))) + \frac{I(\lambda_j)}{G\lambda_j^{-2d} \exp(-P_r(\lambda_j, \theta))} \right], \tag{4}$$

where $P_r(\lambda_j, \theta) = \sum_{\tau=1}^{r} \theta_\tau \lambda_j^{2\tau}$ and defines the closed interval of admissible estimates to be $\mathcal{D} = [\Delta_1, \Delta_2] \subset [-1/2, 1/2]$ and $m = o(n)$ is the bandwidth choice, i.e. the number of periodogram ordinates to be used in the estimation. Then concentrating $U_n(d, G, \theta)$ with respect to G we can write the likelihood function as

$$L_n(d,\theta) = \log \hat{G}(d,\theta) - \frac{1}{m}\sum_{j=1}^{m} P_r(d,\theta) - \frac{2d}{m}\sum_{j=1}^{m}\log\lambda_j + 1, \quad (5)$$

$$\hat{G}(d,\theta) = \frac{1}{m}\sum_{j=1}^{m} I(\lambda_j)\exp\left(P_r(\lambda_j,\theta)\right)\lambda_j^{2d}. \quad (6)$$

Thus, Andrews & Sun (2004) propose to minimize (5) over the admissible set $(d, \theta) \in \mathcal{D} \times \Theta$

$$(\hat{d}_{AS}, \hat{\theta}_{AS}) = \underset{(d,\theta)\in\mathcal{D}\times\Theta}{\arg\min}\, L_n(d,\theta), \quad (7)$$

where Θ is compact and convex set in $\mathbb{R}^r$. As shown by Andrews & Sun (2004) the asymptotic variance of $\hat{d}_{AS}$ is inflated by a multiplicative constant.

It should be noted that it is not necessary to correct for an unknown mean of $\{X_t\}$ as we only compute the DFT at the frequencies $\lambda_j = \frac{2\pi j}{n}$ for $j = 1, ..., m$ where $m = o(n)$, rendering the log-likelihood local to frequency zero. This general result only holds for stationary values of d. Assuming an unknown mean of the generating process when we are in the non-stationary region is the same as saying that the data generating process is free of linear trends, in the usual setup of e.g. Robinson (1995*a*) and Andrews & Sun (2004).

The difference between the objective function defined in Robinson (1995*a*) and Andrews & Sun (2004) is how we approximate $\varphi(\lambda)$ as $\lambda \to 0$ by $\log G - P_r(\lambda, \theta)$ where the polynomial term $P_r(\lambda, \theta)$ vanishes for $\theta = 0$.

Given Assumptions 1-4 in Andrews & Sun (2004) and utilizing their Lemma 1 and Lemma 2, the estimates $(\hat{d}_{AS}, \hat{\theta}_{AS})$ are equal to the solution to the first-order conditions with probability that goes to one as $n \to \infty$. This solution is consistent and asymptotically normal, Andrews & Sun (2004, pp. 572). The asymptotic bias is of order $O(\lambda^{\min\{s,2+2r\}})$, where s is measure of the smoothness of the spectral density near frequency zero, for the LPW estimator and $O(\lambda^2)$ for the LW estimator. That the asymptotic bias of the LPW estimator is of order $O(\lambda^{\min\{s,2+2r\}})$ follows from Assumption 4 and it is clearly seen that if $r = 0$ and by Assumption 2 that $s > 2r$ the asymptotic bias reduces to that of the classical LW estimator, i.e. $O(\lambda^2)$. In this paper, we only consider long memory processes with potential short-run contamination, but if $\{X_t\}$ is a perturbed fractional process, the orders will be smaller and dependent on d. Nonetheless, the LPW estimator will still be consistent at the expense of lower convergence rate and higher asymptotic bias, see Arteche (2004) for the LW case. In the perturbed case, the asymptotic bias of not modeling the spectral density appropriately will be of order $O(\lambda^{2d})$, Hurvich & Ray (2003), Arteche (2004) and Hurvich, Moulines & Soulier (2005).

3 The extended local polynomial Whittle estimator

We define a fractional integrated process as one that is stationary or exhibits some weak dependence after the application of the fractional filter, $(1 - L)^d$. We often distinguish between two ways of expressing a fractional process as a function of weakly dependent innovations, i.e. Type I and Type II processes, see Marinucci & Robinson (1999). As we want to stay in the framework of Abadir et al.

(2007) and Andrews & Sun (2004), we work in the setting of defining the fractional process as a Type I process. Because we are not only interested in the stationary region, it is not enough just to expand the filter $(1-L)^d$ and express it as an infinite order moving average of the innovations which results in a stationary process when $d < 1/2$. When we move into the non-stationary region, i.e. $d \geq 1/2$, this procedure breaks down because the infinite order moving average of the innovations does not converge. This is circumvented by modeling the process as the partial sum of the component $I(d-p)$ process for some $p \in \mathbb{Z}$ and expanding $(1-L)^{p-d}$ in terms of the innovations. This results in a stationary integer differenced series. The disadvantage is that it introduces discontinuities at $d = 1/2, 3/2, ..p - 1/2$, where $p \in \mathbb{Z}$. The Type II process of fractional integration is designed to cover a wider range of d and thereby circumvent some of the problems concerning the Type I process, see Robinson (1994), Phillips (1999), Tanaka (1999), Marinucci & Robinson (1999), and Robinson (2005). For the derivation of the ExtLPW estimator, we define the fractional process as a Type I process. More specifically, we define the $I(d)$ process as in Abadir & Taylor (1999)

Definition 1 *For $d = p + d_u$, where $p \in \mathbb{Z}$ and $d_u \in (-1/2, 1/2)$, we say that $\{X_t\}$ is an $I(d)$ process, i.e. $X_t \sim I(d)$, if*

$$(1-L)^p X_t = u_t, \quad t = 1-p, 2-p, ..., \tag{8}$$

where $\{u_t\}$ is a second order stationary sequence with spectral density

$$f_u(\lambda) = G_0 |\lambda|^{-2d_u} + o\left(|\lambda|^{-2d_u}\right), \tag{9}$$

as $\lambda \to 0$, where $G_0 \in (0, \infty)$.

Define the extended DFT and the extended periodogram of a time series $\{X_t\}$ evaluated at the Fourier frequencies $\lambda_j = \frac{2\pi j}{n}$, where $j = 1, ..., n$, by

$$w(\lambda_j, d) = w_x(\lambda_j) + c(\lambda_j, d), \tag{10}$$

$$I(\lambda_j, d) = |w(\lambda_j, d)|^2, \tag{11}$$

where $w_x(\lambda_j)$ is the usual DFT defined as

$$w_x(\lambda_j) = \frac{1}{\sqrt{(2\pi n)}} \sum_{t=1}^{n} X_t \exp(it\lambda_j), \tag{12}$$

and the correction term $c(\lambda_j, d)$ takes on constant values on the intervals $d \in \mathcal{D}_p := [p - 1/2, p + 1/2)$, $p \in \mathbb{N}$ and is defined by

$$c(\lambda_j, d) = \begin{cases} 0 \text{ if } d \in \mathcal{D}_0 = [-1/2, 1/2) \\ \exp(i\lambda_j) \sum_{\ell=1}^{p} (1 - \exp(i\lambda_j))^{-\ell} Z_\ell \text{ if } d \in \mathcal{D}_p \quad \text{for } p = 1, 2, ..., \end{cases} \tag{13}$$

where

$$Z_0 = w_x(0) = \frac{1}{\sqrt{(2\pi n)}} \sum_{t=1}^{n} X_t, \tag{14}$$

$$Z_\ell = \frac{1}{\sqrt{(2\pi n)}} \left\{(1-L)^{\ell-1} X_n - (1-L)^{\ell-1} X_0\right\}, \quad \ell = 1, 2, ..., p. \tag{15}$$

In the computation of the step function $c(\lambda_j, d)$, we have to enumerate the data depending on what subspace of $\mathcal{D} = [d_1, d_2]$ we are interested in. This is apparent from looking at (15), for example when $p = 2$. That is, $X_{-i+1}, X_{-i+2}, ..., X_n$ where $i = (0 \vee \lfloor d_2 - 1/2 \rfloor)$. The usual DFT, (12) is always computed using the enumeration $\{X_t\}_{t=1}^{n}$.

This notion of the extended DFT allows us to estimate the usual LPW estimator in the context of non-stationary values for d by minimizing the criteria function defined as (5) over the admissible parameter space. The extension of the DFT to the non-stationary case is based on the work of Phillips (1999), Lahiri (2003), Dalla, Giraitis & Hidalgo (2006) and Abadir et al. (2007). Define the pseudo spectral density of the sequence $\{X_t\} \sim I(d_0)$, where $d_0 = p_0 + d_u$ and $d_u \in (-1/2, 1/2)$ as

$$f(\lambda) = |1 - \exp(i\lambda)|^{-p_0} f_u(\lambda), \quad |\lambda| \leq \pi. \tag{16}$$

From this definition it is clear that

$$f(\lambda) \sim G_0 |\lambda|^{-2d_0} \text{ as } \lambda \to 0^+. \tag{17}$$

Then following Abadir et al. (2007, Lemma 4.4), Definition 1, and (10), the extended DFT has the property that

$$w(\lambda_j, d_0) = (1 - \exp(i\lambda_j))^{-p_0} \omega_u(\lambda_j), \quad j = 1, ..., n, \tag{18}$$

where $\omega_u(\lambda_j)$ is the DFT of the stationary sequence $\{u_t\}$. From Abadir et al. (2007, Lemma 4.4(i)), it follows that

$$\begin{aligned} w_x(\lambda_j) &= (1 - \exp(i\lambda_j))^{-p_0} w_{\Delta^{p_0} x}(\lambda_j) - \exp(i\lambda_j) \sum_{r=1}^{p_0} (1 - \exp(i\lambda_j))^{-r} w_{\Delta^r x} && (19) \\ &= (1 - \exp(i\lambda_j))^{-p_0} w_u(\lambda_j) - \exp(i\lambda_j) \sum_{r=1}^{p_0} (1 - \exp(i\lambda_j))^{-r} w_{\Delta^r x}, && (20) \end{aligned}$$

where the second equality follows from Definition 1. Then the definition in (10) follows trivially. Denote the rescaled extended DFT by

$$v_j = v(\lambda_j, d_0) = \frac{w(\lambda_j, d_0)}{\varphi(\lambda_j)^{1/2} \lambda_j^{-d_0}}, \quad 1 \leq j \leq m. \tag{21}$$

Given that the generating process is linear, equation (18) and Lemma 2 show that the asymptotic behavior of the rescaled extended DFT and periodogram is the same for all $d_0 \in (-1/2, \infty)$. Furthermore, given consistency, $\hat{d} \xrightarrow{p} d_0$ and the definition of the extended DFT, we get

$$w\left(\lambda_j, \hat{d}\right) \xrightarrow{p} w(\lambda_j, d_0). \tag{22}$$

This follows because $c(\lambda_j, d)$ is a step function and therefore constant on the intervals $d \in (p - 1/2, p + 1/2)$ for $p \in \mathbb{N}$. This considerably eases the estimation as we are left with the same estimation procedure as in the stationary case.

If the process is stationary the ExtLPW estimator is identical to the LPW estimator of Andrews & Sun (2004). Similarly to the estimators in Robinson (1995*a*), Andrews & Sun (2004), and Abadir et al. (2007) this estimator is based on the whitening principle of the periodogram. That is, similarly to the stationary case, Robinson (1995a) and Andrews & Sun (2004), the ExtLPW estimator is based on the behavior of the random variables

$$\eta_j = \eta\left(\lambda_j\right) = \frac{I_u(\lambda_j)}{f_u(\lambda_j)}, \ 1 \leq j \leq m. \tag{23}$$

Then given the spectral density of the second order stationary sequence $\{u_t\}$,(9), the first moment is given by

$$E\left[\eta_j\right] = 1 + o(1) + O(j^{-1}\log j) \ \ \forall\ 1 \leq j \leq m \text{ as } n \to \infty. \tag{24}$$

Additionally, under regularity assumptions, see Lahiri (2003) and Abadir et al. (2007), the random variables also satisfy

$$var\left[\eta_j\right] \leq C \ \ \forall\ 1 \leq j \leq m, \tag{25}$$

where C is a positive finite constant and

$$cov\left[\eta_j, \eta_s\right] \to 0 \text{ for } j, s \to \infty \text{ and } j \neq s. \tag{26}$$

In the proof to Lemma 4.6 in Abadir et al. (2007), the above equations are proven.

Then given the equations (24), (25) and (26), the random variables η_j satisfy a weak law of large numbers (WLLN), i.e.

$$\frac{1}{m}\sum_{j=1}^{m}\eta_j \xrightarrow{p} 1, \quad \text{as } n \to \infty. \tag{27}$$

Given additional assumptions, this result is sufficient to ensure consistency of the estimator $\hat{d}$. See further discussions on this later. The WLLN for the random variables η_j is equivalent to a WLLN for the random variables $|v_j|^2$, i.e.

$$\frac{1}{m}\sum_{j=1}^{m}|v_j|^2 \xrightarrow{p} 1, \quad \text{as } n \to \infty. \tag{28}$$

Then given the nature of the spectral density (9) and (18)

$$|v_j|^2 = \eta_j\left(1 + o(1)\right) \ \forall\ 1 \leq j \leq m \text{ as } n \to \infty. \tag{29}$$

Furthermore, given equation (24)

$$E\left[|v_j|^2\right] \leq C \ \forall\ 1 \leq j \leq m. \tag{30}$$

For a more thorough walkthrough of the extended DFT, see Phillips (1999), Lahiri (2003), Dalla et al. (2006), and Abadir et al. (2007). We further note that the variables v_j and η_j are invariant with respect to the mean of $\{u_t\}$.

4 Assumptions

In this section, we introduce the assumptions needed to establish consistency and asymptotic normality of the proposed estimator.

Assumption 1 $\mathcal{D}\times\Theta$ is a compact and convex subset of $\mathbb{R}^{r+1}$ and d_0 and θ_0 lie in the interior of $\mathcal{D} = [d_1, d_2] \subset [-1/2, \infty]$ where $d_0 \neq p_0 - 1/2$, $p_0 \in \mathbb{N}$ and Θ, respectively.

Assumption 1 is a combination of similar assumptions given in Andrews & Sun (2004) and Abadir et al. (2007). Yet, our lower bound is more restrictive than in Abadir et al. (2007), because we need to restrict $E[X_t] = 0$ to facilitate $d_1 = -3/2$. Therefore, we only consider invertible processes, i.e. $d > -1/2$. Furthermore, the assumption restricts the parameters of interest to be in the interior of a compact and convex set. If $\hat{d}$ lies on the boundary of the parameter space, we conjecture that the estimator will be consistent,[3] but it may not be asymptotically normal. As noted by Newey & McFadden (1986, p. 2144), it is sufficient that the estimator is in the relative interior of the parameter space, allowing for equality restrictions to be imposed on the parameters of interest.

Assumption 2 The spectral density of the stationary sequence $\{u_t\}$ is

$$f_u(\lambda) = \varphi(\lambda)\,|\lambda|^{-2d_u} + o\left(|\lambda|^{-2d_u}\right) \quad \text{as } \lambda \to 0^+, \tag{31}$$

where $\varphi(\lambda)$ is continuous at $\lambda = 0$, $\varphi(0) \in (0, \infty)$, and $d_u \in (-1/2, 1/2)$.

Assumption 2 is a result of using the basic semiparametric setup from Definition 1.

Assumption 3 Let $\varphi(\lambda)$ be smooth of order s at $\lambda = 0$, where $s > 2r$ and $r \in \mathbb{N}\backslash\{0\}$, $s \geq 1$. That is, in a neighborhood of $\lambda = 0$, $\varphi(\lambda)$ is $\lfloor s \rfloor$ times continuously differentiable with $\lfloor s \rfloor -$ derivative, $\varphi^{(\lfloor s \rfloor)}$, satisfying a Hölder condition of order $s - \lfloor s \rfloor$ at zero, i.e. $\left|\varphi^{(\lfloor s \rfloor)}(\lambda) - \varphi^{(\lfloor s \rfloor)}(0)\right| \leq C\,|\lambda|^{s-\lfloor s \rfloor}$ for a constant $C < \infty$.

The assumption imposes a regularity condition on the function $\varphi(\lambda)$ that characterizes the semi-parametric setup, Andrews & Sun (2004), i.e. $\varphi(\lambda)$ has a Taylor expansion around $\lambda = 0$

$$\varphi(\lambda) = \sum_{k=0}^{\lfloor s/2 \rfloor} \theta_k \lambda^{2k} + O(\lambda^s) \tag{32}$$

$$= P_r(\lambda, \theta) + O(\lambda^s), \text{ as } \lambda \to 0^+, \tag{33}$$

where $\theta_0 = \varphi(0)$ and

$$\theta_k = -\frac{1}{(2k)!}\frac{d^k}{d\lambda^k}\varphi(\lambda)\bigg|_{\lambda=0}. \tag{34}$$

In general, Assumption 3 holds for general ARFIMA processes for all finite s.

As noted by Andrews & Sun (2004), if $r = 0$ and Assumption 3 holds with $s = 2$, then Assumption A1' of Robinson (1995*a*) holds with $\beta = 2$.

Assumption 4 (a) $\{X_t\}$ is generated by the linear sequence $\{u_t\}$

$$u_t = A(L)\varepsilon_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} a_j^2 < \infty, \tag{35}$$

[3] See e.g. the proof of Hurvich et al. (2005) Theorem 3.1, and their discussion of bounding d away from zero. It is not a trivial question, as we in some sense need d to be bounded away from the boundary because the convergence of the log-likelihood is not uniform on $\mathcal{D} \times \mathbb{R}^r$.

where

$$\begin{aligned} E\left[\varepsilon_t|\Im_{t-1}\right] &= 0, \quad E\left[\varepsilon_t^2|\Im_{t-1}\right] = 1 \quad \text{a.s.}, & (36)\\ E\left[\varepsilon_t^3|\Im_{t-1}\right] &= \sigma^3 \quad \text{a.s.}, & (37)\\ E\left[\varepsilon_t^4|\Im_{t-1}\right] &= \sigma^4 \quad \text{a.s. } \forall\, t = 0, \pm 1, \pm 2, ..., & (38) \end{aligned}$$

and $\Im_{t-1}$ is the $\sigma-$ field generated by $\{\varepsilon_s : s < t\}$. (b) There exists a random variable ε with $E\varepsilon^2 < \infty$ such that for all $\upsilon > 0$ and some generic constant $K > 0$, $\Pr(|\varepsilon_t| > \upsilon) < K \Pr(|\varepsilon| > \upsilon)$. (c) In some neighborhood $(0, \delta)$ of the origin $\alpha(\lambda)$ is differentiable

$$\frac{d}{d\lambda}\alpha(\lambda) = O\left(|\alpha(\lambda)|/\lambda\right) \quad \text{as } \lambda \to 0^+, \tag{39}$$

where $\alpha(\lambda) = \sum_{j=0}^{\infty} a_j \exp(ij\lambda)$.

Assumption 4 says that $\{u_t\}$ is a linear sequence with martingale difference innovations. That is, $\{\varepsilon_t\}$ is adapted to the filtration $\{\Im_t\}$. Furthermore, Assumption 4 does not rule out non-Gaussian processes. It should be possible to relax the linearity assumption, see the consideration regarding non-linearity of $\{u_t\}$ in Abadir et al. (2007).

Assumption 5 $\frac{m^{2r+1/2}}{n^{2r}} \to \infty$ and $\frac{m^{\phi+1/2}}{n^{\phi}} \to 0$ as $n \to \infty$, where $\phi = \min\{s, 2+2r\}$.

Assumption 5 is the same as Assumption 4 in Andrews & Sun (2004). The assumptions are needed to show simultaneous consistency of $(\hat{d}, \hat{\theta})$ and asymptotic normality. Note that the first condition imposes a lower bound on the growth of m which ensures simultaneous consistency of $\hat{d}$ and $\hat{\theta}$ by ensuring that the scaling matrix used to normalize the score and Hessian satisfies a regularity condition that is necessary for consistency of $(\hat{d}, \hat{\theta})$ which will be clarified later on. The second condition is to ensure that the normalized score in distribution converges to a zero mean Gaussian process which is required to show asymptotic normality of the estimators $(\hat{d}, \hat{\theta})$. Andrews & Sun (2004) instead work with $\lim_{n\to\infty} \frac{m^{\phi+1/2}}{n^{\phi}} = A \in (0, \infty)$ where $\phi = \min\{s, 2+2r\}$. They set the divergence rate of m such that they can derive the asymptotic bias and asymptotic mean squared error of $\hat{d}$. We choose a bandwidth m that diverges at a slower rate. Note that the two conditions never exclude each other as $s > 2r$ which follows from Assumption 3.

Assumption 6 For $m = o(n)$ the renormalized periodogram, $\eta'_j\ \forall\ 1 \le j \le m$, satisfies a WLLN

$$\frac{1}{m}\sum_{j=1}^{m} \eta'_j \xrightarrow{p} 1, \text{ as } m, n \to \infty, \tag{40}$$

where $\eta'_j = \frac{I_u(\lambda_j)}{\varphi(\lambda_j)\lambda_j^{-2d_u}} \quad \forall 1 \le j \le m$.

Assumption 6 is equivalent to Assumption B in Abadir et al. (2007) and states that if Assumption 2, 3 and equation (18) hold then

$$\eta'_j = \eta_j\left(1 + o(1)\right)\ \forall\ 1 \le j \le m \text{ as } n \to \infty. \tag{41}$$

Furthermore, (24) implies that

$$E\left[\eta'_j\right] \le C\ \forall\ 1 \le j \le m \text{ as } n \to \infty. \tag{42}$$

5 Consistency and asymptotic normality

Theorem 1 states consistency and asymptotic normality of the proposed ExtLPW estimator.

Theorem 1 *Let $\{X_t\}$ be generated by (8) and assume that Assumptions 1 through 6 hold. Then, as $n \to \infty$, $\hat{d}$ and $\hat{\theta}$ are both consistent and*

$$B_n \begin{pmatrix} \hat{d} - d_0 \\ \hat{\theta} - \theta_0 \end{pmatrix} \xrightarrow{d} N\left(0, \Omega_r^{-1}\right), \tag{43}$$

where B_n being the $(r+1) \times (r+1)$ diagonal matrix with jth diagonal element defined as

$$[B_n]_{11} = m^{1/2} \tag{44}$$

$$[B_n]_{jj} = \left(\frac{2\pi m}{n}\right)^{2j-2} m^{1/2} \text{ for } j = 2, 3, ..., r+1. \tag{45}$$

And Ω_r is the $(r+1) \times (r+1)$ covariance matrix defined as

$$\Omega_r = \begin{pmatrix} 4 & 2\mu_r' \\ 2\mu_r & \Gamma_r \end{pmatrix}. \tag{46}$$

μ_r is a $r \times 1$ vector with kth element $\mu_{r,k}$ and Γ_r is an $r \times r$ matrix with (i,k)th element $[\Gamma_r]_{i,k}$,

$$\mu_{r,k} = \frac{2k}{(2k+1)^2} \text{ for } k = 1, ..., r, \tag{47}$$

$$[\Gamma_r]_{i,k} = \frac{4ik}{(2i+2k+1)(2i+1)(2k+1)} \text{ for } i, k = 1, ..., r. \tag{48}$$

We could have shown $\log^5 m-$consistency (irrespective of $\theta \in \Theta$) using Robinson's (1995*a*) pp. 1642-1643 proof of the $\log^3 m-$consistency of $\hat{d}\,(r=0)$ and adjusting it to account for our weaker Assumption 5 compared to his Assumption 4'. But as this would mainly be a theoretical addition it is left out. Theorem 1 utilizes the F.O.C. approach (because of the multidimensionality of the parameter space) as in Andrews & Sun (2004). Therefore, Theorem 1 jointly delivers consistency and asymptotic normality of $(\hat{d}, \hat{\theta})$. More specifically, since the ExtLPW likelihood ((5) where the periodogram is defined by its extended DFT, i.e. (18)) is a continuous function on a compact set the ExtLPW estimator exists. From Lemma 1 (in the Appendix) we know by Lemma 1 of Andrews & Sun (2004) that there exists a solution to the first order conditions with probability tending to one, and that the solution satisfies the convergence result in Theorem 1, see also Lemmas 1 and 2 of Andrews & Sun (2004). If the (negative) likelihood function is strictly convex and twice differentiable then the solution to the first order conditions is unique and minimizes the log-likelihood and hence equals the ExtLPW estimator.

By the formula for a partitioned inverse Theorem 1, in consequence, implies that,

$$\Omega_r^{-1} = \begin{pmatrix} (4 - 2\mu_r'\Gamma_r^{-1}2\mu_r)^{-1} & -4 * 2\mu_r'\left(\Gamma_r - 2\mu_r 4^{-1} 2\mu_r'\right)^{-1} \\ -\Gamma_r^{-1}2\mu_r\left(4 - 2\mu_r'\Gamma_r^{-1}2\mu_r\right)^{-1} & \left(\Gamma_r - 2\mu_r 4^{-1} 2\mu_r'\right) \end{pmatrix} \tag{49}$$

$$= \begin{pmatrix} c_r/4 & -\frac{c_r}{2}\mu_r'\Gamma_r^{-1} \\ -\frac{c_r}{2}\Gamma_r^{-1}\mu_r & \Gamma_r^{-1} + c_r\Gamma_r^{-1}\mu_r\mu_r'\Gamma_r^{-1} \end{pmatrix}, \tag{50}$$

where $c_r = \left(1 - \mu_r' \Gamma_r^{-1} \mu_r\right)^{-1}$ for $r > 0$ and $c_0 = 1$.

A few remarks are in order. First of all, the asymptotic variance of $\sqrt{m}\left(\hat{d} - d_0\right)$ is free of nuisance parameters and equal to $c_r/4$. Secondly, in light of Assumption 5 the estimator given by ExtLPW for $r > 0$ allows one to choose a bandwidth m much larger than in the classical LW approach, resulting in an estimator that has asymptotic normality with a faster rate of convergence, as a function of the sample size n. The cost of introducing a polynomial is inflation of the asymptotic variance by a multiplicative constant, i.e. $c_0 = 1, c_1 = 9/4, c_2 = 3.52, \ldots$, see Andrews & Sun (2004).

Consistency of $\hat{d}$ provides no information about θ_0 as the concentrated log-likelihood becomes flat as a function of $\hat{\theta}$ as $n \to \infty$. The rate at which it becomes flat furthermore differs for each element of $\hat{\theta}$.

As discussed earlier, our model setup does not consider volatility processes, e.g. in the sense of long memory signal-plus-noise models as in Hurvich et al. (2005) and Frederiksen, Nielsen & Nielsen (2008), among others. Introducing a perturbation in our framework would indeed bias our estimator in the same manner as the classical fractional integration estimators (LW and log-periodogram), i.e. the leading bias term is of order $O\left(\lambda^{2d}\right)$ implying a slower convergence rate compared to the leading bias term of the pure long memory process of $O\left(\lambda^{2}\right)$, Hurvich & Ray (2003), Arteche (2004), Hurvich et al. (2005), and Frederiksen et al. (2008).

6 Simulation study

6.1 Setup

This sections concerns the finite sample performance of the proposed estimator. We generate $I(d)$ processes according to Definition 1 by using the circulant embedding method as described in Davies & Harte (1987), i.e. as a stationary Type I fractionally integrated process in the terminology of Marinucci & Robinson (1999), see also Beran (1994, pp. 215-217). Non-stationary processes are then defined as $[d]$ fold partial sums of stationary $I\left(d - [d]\right)$ processes. $[d]$ is defined as the integer closest to d. Furthermore, when $d - [d] = 1/2$, $[d]$ is equal to $d + 1/2$. $\{u_t\}$ is contaminated by autoregressive (AR) and moving average (MA) roots. That is, we consider the following data generating process (DGP)

$$(1 - \alpha L)(1 - L)^d (X_t) = (1 + \beta L) u_t,$$

where $u_t \overset{i.i.d.}{\sim} N(0,1)$, $\alpha \in \{-0.8, -0.5, 0, 0.5, 0.8\}$ and $\beta \in \{-0.8, -0.5, 0, 0.5, 0.8\}$. We set the fractional parameter of interest equal to $d = \{-0.3, 0, 0.3, 0.7, 1, 1.3, 1.7, 2\}$. Sample size is set equal to $n \in \{128, 512, 1024\}$ and bandwidth $m = \lfloor n^a \rfloor$ where $a \in \{0.5, 0.65, 0.8\}$. The bias and root mean squared error (RMSE) were computed using 1000 replications. Simulations were done in Matlab v7.2. The optimization procedure was implemented using the unconstrained minimization procedure in Matlab where we used the BFGS algorithm. We tried different procedures to find the optimum, among others evaluating the first-order conditions and thereby finding the corresponding roots. All the different approaches yielded similar results and we therefore elected to use the BFGS algorithm as it is easy to implement and fairly fast computationalwise.

We compare our derived estimators to the local Whittle (LW) estimator of Robinson (1995a), local polynomial Whittle (LPW) estimator of Andrews & Sun (2004), and extended local Whittle (ExtLW) estimator of Abadir et al. (2007). Regarding the parameterization of the polynomial, we set $r = \{1, 2\}$. As initial values we set the memory parameter equal to the log-periodogram estimate of Geweke & Porter-Hudak (1983) and the polynomial terms were all set equal to 1.

Tables 1-5 present results from the small simulation study. We only display a subset of the results. Attention is restricted to the cases with no short-run dynamics with $n = 512$, Table 1, with moving-average short-run dynamics, $\beta \in \{-0.8, -0.5\}$ with $n = 512$, Table 2-3, and finally with autoregressive short-run contamination, $\alpha \in \{0.5, 0.8\}$ with $n = 512$, Table 4-5.

6.2 Simulation results

When the DGP is not contaminated by any short-run dynamics, it is be preferable to use a larger bandwidth. If there is short-run contamination, the opposite is the case, excluding of course the bias reducing methods, i.e. LPW estimators. Furthermore, bias decreases as a function of sample size.

In Table 1, results without short-run contamination are presented. For the stationary region all the estimators are seemingly unbiased, and we clearly see that the extended estimators are in a statistical sense equal to their non-extended counterparts. The RMSE shows that the fractional parameter is estimated quite accurately, and we notice that the estimators using a polynomial to reduce potential bias has a larger RMSE than the estimators using a constant in a shrinking neighborhood of zero. Moving on to the non-stationary region, i.e. $d \geq 1/2$, we see that the bias of LW and LPW increases quite considerably, especially when d is larger than 1. This is to be expected as the LW estimator is not consistent for $d > 1$ and the LW estimator is biased towards unity, thereby confirming the results of Phillips & Shimotsu (2004). This result for the LW estimator also seems to hold for the LPW estimator. Clearly, in the case where $d > 1$, the extended estimators are the best, and with the ExtLW estimator being the best in a bias sense, as there is no short-run contamination. For the extended estimators, regardless of which region we are in, RMSE indicates that the fractional parameter of interest is estimated accurately. Additionally, the RMSE does not vary much in the given range of d.

Looking at the case where we introduce short-run contamination, Tables 2-5, we generally find that the estimators are biased and the bias increases as the contamination of the signal increases. This is expected as the low frequencies (long-run in the time domain) are contaminated by the higher frequencies (short-run in the time domain) of the spectral density. The bias is highest when introducing positive AR noise and negative MA noise. When $\alpha = 0.8$ and $\beta = -0.8$ we clearly see the advantage of using an estimator that approximates this short-run contamination. Furthermore, when looking at more moderate negative MA noise and positive AR noise, $\beta = -0.5$ and $\alpha = 0.5$, respectively, it is not preferable to use a lower bandwidth for the LPW (only in the stationary case) and ExtLPW estimators, as for the other estimators. That is, the LPW and ExtLPW estimators are very robust to MA and AR contamination because of the way they approximate the spectral density of the short-run noise by a polynomial. Hence, it is possible to choose a higher bandwidth without increasing the bias which is an important result especially when looking at shorter time series.

To sum up, it is important to approximate the short-run component of the local approximation by a polynomial function instead of merely a constant, especially when there is a high degree of persistence, since the polynomial estimators are clearly less biased than the LW and ExtLW. This is especially important in shorter time series as the bias can be extreme when there is short-run contamination. As shown by Andrews & Sun (2004) and in this paper, the improved bias comes at a cost of increasing the variance by a multiplicative constant. When looking at the non-stationary region, it is important to use the extended versions especially when $d \geq 1$ as there is considerable bias gains from using these extensions.

For the ExtLW estimator where $\alpha = \{-0.5, 0, 0.5\}$ Abadir et al. (2007) arrive at similar results. Furthermore, the simulation results for the ExtLW estimator from Abadir et al. (2007) when $n = 500$ and $m = \left[n^{.65}\right]$ are similar to the results obtained in Shimotsu & Phillips (2005) for their exact local Whittle estimator (ELW). Shimotsu & Phillips (2005) compare their ELW estimator to two different types of tapered estimators (the tapering proposed in Velasco (1999*a*) and Hurvich & Chen (2000)), and conclude that their estimator is the best general purpose estimator when compared to the tapered version of the LW estimator. Therefore, we conclude that our proposed estimator also outperforms the tapered LW estimators especially in the presence of short-run contamination.

7 Application to credit spreads

In this section, we investigate potential long memory properties of treasury yield and yields on corporate bonds, spreads over treasury and spreads between corporate yields, as previously examined by Ratta & Urga (2005). Both in structural models (Merton (1974), Black & Cox (1976), Das (1995), Longstaff & Schwartz (1995), Hull & White (1995), Leland & Toft (1996), among others) and reduced-form models (Ramaswamy & Sundaresan (1986), Jarrow & Turnbull (1995), Das & Tufano (1995), Duffie & Huang (1996), Jarrow, Lando & Turnbull (1997), Madan & Unal (1996), Duffie & Singleton (1999), among others), credit spreads play an integral role in pricing of risky debt and credit derivatives. Neither approach considers that the process driving the data generating process might be poorly approximated by considering the classical $I(0)/I(1)$ setup, as discussed in Ratta & Urga (2005), see references therein. The objective of Ratta & Urga (2005) was to fill a gap in the credit spread literature, i.e. to investigate if credit spreads exhibit potential fractional integration and if there are some long-run relations that can be explained through fractional cointegration. They use the log-periodogram estimator of Geweke & Porter-Hudak (1983) and the LW estimator analyzed by Robinson (1995*a*). As both of these estimators are severely biased in the presence of short-run contamination, see Nielsen & Frederiksen (2005) for a simulation study, and there is no asymptotic theory for $d \geq 3/4$, we suggest using more up-to-date semiparametric estimators that potentially mitigate the bias introduced by short-run contamination and where the distributional theory holds for $d \geq 3/4$. The usual way to reduce bias for the log-periodogram and the LW estimator is to select a smaller bandwidth thereby sacrificing efficiency in the form of a larger variance.

The data considered here consists of daily observations for the 30 year historical US Treasury Constant Maturity Yields and Moody' Aaa and Baa.[4] For a more thorough description of the data and our reason for using rating-specific indices, see Ratta & Urga (2005). Our data covers the period

[4]Ratta & Urga (2005) look at two other ratings besides the two that we consider, i.e., Aa and A. The reason we only look at Aaa and Baa is that these data series are downloadable from the Federal Reserve at http://www.federalreserve.gov/releases/h15/data.htm

2nd of January 1986 through 15th of February 2002 for a total of 4,034 observations. The 30-year Treasury constant maturity series was discontinued on February 18th, 2002, and reintroduced on February 9th, 2006. We could have used the 20-year Treasury constant maturity series and used a correction factor delivered by the U.S. Treasury, but we choose to focus on the shorter sample period.

As opposed to Ratta & Urga (2005) we opt to log transform[5] the series before considering further analysis. Therefore, spreads, i.e. spreads over treasury (sAaaTreas, sBaaTreas) and spreads between corporate yields (sBaaAaa), are defined as the difference between the logs of the respective series. Time series plot of the individual series and the spreads are shown in Figure 1. There are signs of heteroskedasticity, volatility clustering and potential structural breaks. Granger & Terasvirta (1999) show that the number of regime switches affects the long memory parameter. Diebold & Inoue (2001), Granger & Hyung (2004) and Haldrup & Nielsen (2007) discuss that if series display breaks, particular in their deterministic components, these processes will give the impression of persistence. That is, we can mistakenly conclude that a process displays long memory, where in fact it is due to a structural break in the series. Therefore, we split the full sample in four even subperiods. The results from looking at subperiods were comparable to the whole sample period, and therefore omitted. Additionally, we implemented a test for spurious long memory where we temporally aggregated the data and compared the long memory estimates through a Wald type test for identical memory across aggregation, see Ohanissian, Russell & Tsay (2008) and Frederiksen & Nielsen (2008). We could not reject that the memory parameters are identical. Hence, we conjecture that the estimated long memory is not spurious in the sense that it is generated by structural breaks, e.g. a non-stationary level shift in mean DGP. Looking at first differences of the respective series, they seem stationary (when looking at the autocorrelation diagrams which are omitted). Especially, the spread series look as if they have been overdifferenced, i.e. the introduction of moving average behavior in the autocorrelation diagram.

Insert Figure 1 about here

Figures 2-7 display the semiparametric results for the LW, LPW, ExtLW and ExtLPW estimators, for different bandwidth ranges.

Insert Figures 2-7 about here

Generally, results for the fractional integration estimates show that the estimators that do not model the short-run components by a polynomial have a tendency to decrease as a function of the bandwidth, at least for sufficiently large bandwidth. This is of course reasonable considering the theoretical properties of these estimators.

The logs of Aaa, Baa and Treasury yields are in the non-stationary area with the long memory parameter estimated in the proximity of a unit root. As the asymptotic theory does not hold for the LPW estimator when $d \geq 1/2$ and for the LW estimator when $d \geq 3/4$, we primarily rely on the extended estimators. In general, we cannot reject that the log yields contain a unit root.

Looking at the spreads over treasury (sAaaTreas, sBaaTreas) and spreads between corporate yields (sBaaAaa), the estimated long memory is clearly in the non-stationary region regardless of the chosen

[5] Log transforming the data is also preferred in the sense that it better captures the non-linear relationship between yields and ratings, Manzoni (2002).

bandwidth and estimator. The LW and ExtLW are for larger bandwidth choices significantly different from $d = 1$, whereas we cannot reject the presence of a unit root for the polynomial estimators.

Like Ratta & Urga (2005), we have also applied a parametric ARFIMA-GARCH.[6] The results confirm the findings obtained from the semiparametric analysis, so they are omitted. If indeed the true generating process is modeled by GARCH innovations this does not affect the asymptotic theory of the semiparametric estimates as shown in a simulation study by Nielsen & Frederiksen (2005), so, in that respect, it is not unreasonable that the conclusions are the same.

Overall, it cannot be rejected that the log yields of Aaa, Baa and Treasury bonds contain a unit root. However, the results are more mixed when looking at spreads, depending on the estimator and bandwidth choice. Therefore, as in Ratta & Urga (2005), we can reject the reduced-form modeling of Das & Tufano (1995), Jarrow et al. (1997) and Duffie & Singleton (1999). This explicitly implies the data generating process of the risk-free process, and hence also credit spreads, follows a short-memory process, i.e. $I(0)$. The relevance of modeling yields in a more flexible fractional cointegration setup should be considered and is at least a relevant alternative to the classical $I(1)/I(0)$ terminology.

8 Concluding remarks

In this paper, we propose a semiparametric estimator that circumvents the relatively slow convergence and finite sample bias of the classical local Whittle estimator when there is short-run contamination (e.g. autoregressive and/or moving average roots) and non-stationarity. The bias reduction is obtained by approximating the spectrum of the short memory component by a polynomial instead of a constant in a shrinking neighborhood of frequency zero. In addition, the notion of extended DFT and periodogram is used to extend the estimator to cover non-stationary values of the fractional integration parameter d. We show consistency and asymptotic normality of the estimator. A simulation study confirms the asymptotic results. The adequacy of the estimator is shown through an empirical analysis of credit spreads.

As a final note, we could also have opted to expand the work of Andrews & Sun (2004) by utilizing the work of Shimotsu & Phillips (2005). However, we conjecture that such an estimator would in fact be consistent and asymptotically normal in the same manner as the exact local Whittle estimator of Shimotsu & Phillips (2005). Robinson (2005) showed that the expected squared deviation between the DFT of the Type I and the Type II model is of order $O\left(n^{-1}\right)$. Therefore, we conjecture that the derived results also hold for Type II fractional processes.

9 Appendix of proofs and lemmas

The proof to Theorem 1 relies heavily on Abadir et al. (2007) and Andrews & Sun (2004).

The Appendix section is structured as follows: In the first section the proof to Theorem 1 is given. Section 2 presents technical lemmas adapted from Andrews & Sun (2004) and Abadir et al. (2007).

[6] We also estimated other GARCH specifications, e.g., IGARCH and FIGARCH. These other specifications seem indeed to be justified in the sense that $\alpha + \beta \approx 1$ in the GARCH(1,1) specification. A further analysis is beyond the scope of this paper.

9.1 Proof of Theorem 1

Proof. Set

$$
\begin{aligned}
D_m(\eta) &= \left\{ d \in [d_1, d_2] : \left(\log^5 m\right) |d - d_0| < \eta \right\} \text{ for } \eta > 0, && (51)\\
g_j &= \lambda_j^{-2d} G \exp(-P_r(\lambda_j, \theta)) && (52)
\end{aligned}
$$

As in Andrews & Sun (2004) denote the score and the Hessian of the scaled objective function as $S_n(d,\theta) = m\nabla L_n(d,\theta)$ and $H_n(d,\theta) = m\nabla^2 L_n(d,\theta)$, respectively.

$$
\begin{aligned}
S_n(d,\theta) &= \hat{G}^{-1}(d,\theta) \sum_{j=1}^{m} \left(I_j(d) \exp\left(P_r(\lambda_j,\theta)\right) \lambda_j^{2d} - m^{-1} \sum_{k=1}^{m} I_k(d) \exp\left(P_r(\lambda_k,\theta)\right) \lambda_k^{2d} \right) && (53)\\
&\quad \times \left(2\log j, \lambda_j^2, ..., \lambda_j^{2r}\right)' \\
&= \hat{G}^{-1}(d,\theta) \sum_{j=1}^{m} \left(\frac{GI_j(d)}{g_j(d,\theta)} - m^{-1} \sum_{k=1}^{m} \frac{GI_k(d)}{g_k(d,\theta)} \right) X_j. && (54)
\end{aligned}
$$

$$
\begin{aligned}
H_n(d,\theta) &= \hat{G}^{-2}(d,\theta) \left(\begin{array}{c} \hat{G}(d,\theta) \sum_{j=1}^{m} I_j(d) \exp\left(P_r(\lambda_j,\theta)\right) \lambda_j^{2d} \left(2\log j, \lambda_j^2, ..., \lambda_j^{2r}\right)' \left(2\log j, \lambda_j^2, ..., \lambda_j^{2r}\right) \\ -m \left(m^{-1} \sum_{j=1}^{m} I_j(d) \exp\left(P_r(\lambda_j,\theta)\right) \lambda_j^{2d} \left(2\log j, \lambda_j^2, ..., \lambda_j^{2r}\right)' \right) \\ \times \left(m^{-1} \sum_{j=1}^{m} I_j(d) \exp\left(P_r(\lambda_j,\theta)\right) \lambda_j^{2d} \left(2\log j, \lambda_j^2, ..., \lambda_j^{2r}\right)' \right)' \end{array} \right) \\
&= \hat{G}^{-2}(d,\theta) \left(\begin{array}{c} \hat{G}(d,\theta) \sum_{j=1}^{m} \frac{GI_j(d)}{g_j(d,\theta)} X_j' X_j - m \left(m^{-1} \sum_{j=1}^{m} \frac{GI_j(d)}{g_j(d,\theta)} X_j \right) \\ \times \left(m^{-1} \sum_{j=1}^{m} \frac{GI_j(d)}{g_j(d,\theta)} X_j \right)' \end{array} \right), && (55)
\end{aligned}
$$

where $X_j = \left(2\log j, \lambda_j^2, ..., \lambda_j^{2r}\right)'$. Define the deterministic scaling matrix B_n equal to the $(r+1)\times(r+1)$ diagonal matrix with jth diagonal element defined as

$$
\begin{aligned}
[B_n]_{11} &= m^{1/2} && (56)\\
[B_n]_{jj} &= \left(\frac{2\pi m}{n}\right)^{2j-2} m^{1/2} \text{ for } j = 2, 3, ..., r+1. && (57)
\end{aligned}
$$

Since $L_n(d,\theta)$ is a continuous function defined on a compact set the estimator exists. Strict convexity of the (negative) log-likelihood, $L_n(.)$, implies that the estimator is unique. Then, by strict convexity and twice continuous differentiability of $L_n(.)$, implies that if a solution to the F.O.C. exists with probability tending to one, which essentially follows by Andrews & Sun (2004, Lemma 1), then it is unique and minimizes the objective function. Now we can use Lemma 1 to verify that the conditions in Lemma 1 of Andrews & Sun (2004) hold. Andrews & Sun (2004, Lemma 1(i)) holds by Assumption 5. Andrews & Sun (2004, Lemma 1(ii)) holds by Lemma 1(e) and the second condition in Assumption 5. Andrews & Sun (2004, Lemma 1(iii)) holds by Lemma 1(a) and Lemma 1(b) and the positive definiteness of Ω_r. Andrews & Sun (2004, Lemma 1(iv)) holds by Lemma 1(c) and Lemma 1(d) as it ensures for some sequence η_n that goes sufficiently slowly to zero, $C_n \to \infty$, holds.[7]. Thus, what remains is to show strict convexity. We know that if for all leading principal minors $D_l(d,\theta) > 0$, $l = 1, 2, ..., 1+r+1$ and $\forall\ (d,\theta) \in D \times \Theta \subset [d_1, d_2] \times \mathbb{R}^{r+1}$ then it follows that (negative) $L_n(d,\theta)$ is

[7] Andrews & Sun (2004), give an example of such a sequence, i.e., setting $\eta_n = \log^{-1} m$.

strictly convex on $D \times \Theta \subset [d_1, d_2] \times \mathbb{R}^{r+1}$. Andrews & Sun (2004) prove this by noticing that for any $c \in \mathbb{R}^{r+1} \backslash \{0\}$

$$\begin{aligned} c' H_n(d,\theta) c \hat{G}^2(d,\theta) m^{-1} &= \hat{G}(d,\theta) m^{-1} \sum_{j=1}^{m} \frac{G I_j(d)}{g_j(d,\theta)} (c' X_j)^2 - \left(m^{-1} \sum_{j=1}^{m} \frac{G I_j(d)}{g_j(d,\theta)} c' X_j \right)^2 \\ &= a'a \cdot b'b - (a'b)^2 > 0, \end{aligned} \tag{58}$$

where a and b are vectors of order m with $a_j = \left(m^{-1} \frac{G I_j(d)}{g_j(d,\theta)} \right)^{1/2}$ and $b_j = \left(m^{-1} \frac{G I_j(d)}{g_j(d,\theta)} \right)^{1/2} c' X_j(d,\theta)$ for $j = 1, ..., m$ and the inequality holds by the Cauchy-Schwarz inequality. ■

9.2 Lemmas

Lemma 1 is the same as Lemma 2 in Andrews & Sun (2004). Part (e) is lacking the bias term as we impose a weaker form of divergence of the bandwidth m in Assumption 5 than Andrews & Sun (2004) do. Otherwise, the proof follows from Andrews & Sun (2004) with modifications to allow for $d \geq 1/2$. These modifications follow Abadir et al. (2007) and there notion of the extended DFT and periodogram. Lemma 2 is adapted from Lemma 4.6 in Abadir et al. (2007) and deals with the asymptotic properties of the renormalized DFT's and hence generalizes Theorem 2 of Robinson (1995*b*, Theorem 2) as done in Abadir et al. (2007) to suit the non-stationary case. Furthermore, we will use Lemma 4.2 and Lemma 4.4 of Abadir et al. (2007) extensively. In short, Lemma 4.2 gives relevant bounds for proving consistency of the estimator $\hat{d}$. Lemma 4.4 gives the algebraic relation between the DFT of the series $\{X_t\}$ and the differenced series $\{\Delta^p X_t\}$.

Lemma 1 *Under Assumptions 1-6, as $n \to \infty$, we have*

(a) $B_n^{-1} J_n B_n^{-1} \xrightarrow{p} \Omega_r$

(b) $\left\| B_n^{-1} \left(H_n(d_0,\theta_0) - J_n \right) B_n^{-1} \right\| = o_p(1)$

(c) $\sup_{\theta \in \Theta} \left\| B_n^{-1} \left(H_n(d_0,\theta) - H_n(d_0,\theta_0) \right) B_n^{-1} \right\| = o_p(1)$

(d) $\sup_{(d,\theta) \in D_m(\eta_n) \times \Theta} \left\| B_n^{-1} \left(H_n(d,\theta) - H_n(d_0,\theta) \right) B_n^{-1} \right\| = o_p(1)$ *for all sequences of constants* $\{\eta_n\}_{n \geq 1}$ *for which* $\eta_n = o(1)$

(e) $B_n^{-1} S_n(d_0,\theta_0) \xrightarrow{d} N(0, \Omega_r)$,

Proof of (a). Follows by approximating sums by integrals, see Andrews & Sun (2004, pp. 597), where they refer to Andrews & Guggenberger (2003) and Lemma 2(a), (h), and (i). ■

Proof of (b). As in Andrews & Sun (2004) write with the only difference that our extended periodogram of a time series $\{X_t\}$ depends not only on the Fourier frequencies, but also on the value of d, i.e. $I_j(d) = I(\lambda_j, d) = |\omega(\lambda_j, d)|^2$

$$\hat{G}_{a,b}(d,\theta) = m^{-1} \sum_{j=1}^{m} I_j(d) \exp\left(P_r(\lambda_j,\theta) \right) \lambda_j^{2d} (2 \log j)^a \left(\frac{j}{m} \right)^{2b} \tag{59}$$

$$J_{a,b} = G_0 m^{-1} \sum_{j=1}^{m} (2 \log j)^a \left(\frac{j}{m} \right)^{2b}, \tag{60}$$

for $a=0,1,2,\ b=0,...,r$ and where we in defining $J_{a,b}$ have substituted $I_j(d)\exp(P_r(\lambda_j,\theta))\lambda_j^{2d}$ with G_0. As in (A.7) in Andrews & Sun (2004), the $(1,1),(1,k)$ and (k,i) element of $B_n^{-1}H_nB_n^{-1}$ and $B_n^{-1}J_nB_n^{-1}$ for $k,i=2,...,r+1$ are then given by

$$\begin{aligned}&\hat{G}_{0,0}^{-2}\left(\hat{G}_{0,0}\hat{G}_{2,0}-\hat{G}_{1,0}^{2}\right),\\&\hat{G}_{0,0}^{-2}\left(\hat{G}_{0,0}\hat{G}_{1,k-1}-\hat{G}_{1,0}\hat{G}_{0,k-1}\right),\\&\hat{G}_{0,0}^{-2}\left(\hat{G}_{0,0}\hat{G}_{1,k+i-2}-\hat{G}_{0,k-1}\hat{G}_{0,i-1}\right),\end{aligned}\tag{61}$$

where for $B_n^{-1}J_nB_n^{-1}$ just substitute $\hat{G}_{a,b}(d,\theta)$ with $J_{a,b}$. To prove Lemma 1(b), it then suffices to show that

$$\Delta_{a,b}=\left|\frac{\hat{G}_{a,b}(d_0,\theta_0)}{G_0}-\frac{J_{a,b}}{G_0}\right|=o_p\left(\log^{-2}m\right),\tag{62}$$

$\forall a=0,1,2$ and $b=0,1,...,r$. Write

$$\begin{aligned}\Delta_{a,b}&=\left|\frac{m^{-1}\sum_{j=1}^{m}I_j(d_0)\exp(P_r(\lambda_j,\theta_0))\lambda_j^{2d_0}(2\log j)^a\left(\frac{j}{m}\right)^{2b}}{G_0}-\frac{G_0m^{-1}\sum_{j=1}^{m}(2\log j)^a\left(\frac{j}{m}\right)^{2b}}{G_0}\right|\\&=\left|m^{-1}\sum_{j=1}^{m}\frac{I_j(d_0)\exp(P_r(\lambda_j,\theta_0))\lambda_j^{2d_0}(2\log j)^a\left(\frac{j}{m}\right)^{2b}}{g_j\exp(P_r(\lambda_j,\theta_0))\lambda_j^{2d_0}}-m^{-1}\sum_{j=1}^{m}(2\log j)^a\left(\frac{j}{m}\right)^{2b}\right|\\&=\left|m^{-1}\sum_{j=1}^{m}\frac{I_j(d_0)}{g_j}(2\log j)^a\left(\frac{j}{m}\right)^{2b}-m^{-1}\sum_{j=1}^{m}(2\log j)^a\left(\frac{j}{m}\right)^{2b}\right|\\&=\left|m^{-1}\sum_{j=1}^{m}\left(\frac{I_j(d_0)}{g_j}-1\right)(2\log j)^a\left(\frac{j}{m}\right)^{2b}\right|\\&\leq\left|m^{-1}\sum_{j=1}^{m-1}\left[(2\log k)^a\left(\frac{k}{m}\right)^{2b}-(2\log(k+1))^a\left(\frac{k+1}{m}\right)^{2b}\right]\sum_{j=1}^{k}\left(\frac{I_j(d_0)}{g_j}-1\right)\right|\\&\quad+\left|(2\log m)^a m^{-1}\sum_{j=1}^{m}\left(\frac{I_j(d_0)}{g_j}-1\right)\right|\\&:=\nu_{1,a,m}+\nu_{2,a,m},\end{aligned}$$

where the inequality follows from using summation by parts. Furthermore, under Assumption 1, i.e. $d_0\neq p_0-1/2$ for $p_0\in\mathbb{Z}$ and the definition of the extended DFT, the assumption of linearity of the generating process (Assumption 4(a)), and together with Abadir et al. (2007, Lemma 4.4) and Lemma 2, implies that the behavior of the extended DFT and periodogram are the same for all $d\in(-1/2,\infty)$. Therefore, the results from Andrews & Sun (2004) also hold in our case. That is, the proof of (62) follows by collecting the terms (A.11)-(A.13) in Andrews & Sun (2004, pp. 598-599) and using Assumption 5

$$\nu_{1,a,m}+\nu_{2,a,m}=O_p\left((\log^a m)\,m^{-1/2}+(\log^a m)\,m^{\phi}n^{-\phi}\right)=o_p\left(\log^{-2}m\right).$$

■

Proof of (c). By (62) and $J_{a,b} = O(\log^a m)$, we get that

$$\hat{G}_{a,b}(d_0, \theta_0) = O_p(\log^a m), \tag{63}$$

for $a = 0, 1, 2$ and $b = 0, ..., r$ and

$$\hat{G}_{0,0}(d_0, \theta_0) = G_0 + o_p\left(\log^{-2} m\right), \tag{64}$$

where $G_0 > 0$. Then given that we can write the elements for $B_n^{-1} H_n B_n^{-1}$ as in (61) and the above results hold, it suffices to show that

$$\sup_{\theta\in\Theta}\left|\hat{G}_{a,b}(d_0, \theta) - \hat{G}_{a,b}(d_0, \theta_0)\right| = o_p\left(\log^{-2} m\right), \tag{65}$$

$\forall a = 0, 1, 2$ and $b = 0, ..., r$. Write the left-hand side of (65) as

$$\begin{aligned}
&\sup_{\theta\in\Theta}\left|\begin{array}{c} m^{-1}\sum_{j=1}^{m} I_j(d_0)\exp\left(P_r(\lambda_j, \theta)\right)\lambda_j^{2d_0}(2\log j)^a\left(\frac{j}{m}\right)^{2b} \\ -m^{-1}\sum_{j=1}^{m} I_j(d_0)\exp\left(P_r(\lambda_j, \theta_0)\right)\lambda_j^{2d_0}(2\log j)^a\left(\frac{j}{m}\right)^{2b}\end{array}\right| \\
=\;& \sup_{\theta\in\Theta}\left|m^{-1}\sum_{j=1}^{m} I_j(d_0)\left[\exp\left(P_r(\lambda_j, \theta)\right) - \exp\left(P_r(\lambda_j, \theta_0)\right)\right]\lambda_j^{2d_0}(2\log j)^a\left(\frac{j}{m}\right)^{2b}\right| \\
\leq\;& \sup_{\theta\in\Theta}\sup_{k=1,...,m}\left|\exp\left(P_r(\lambda_k, \theta)\right) - \exp\left(P_r(\lambda_k, \theta_0)\right) - 1\right| \\
&\times m^{-1}\sum_{j=1}^{m} I_j(d_0)\exp\left(P_r(\lambda_j, \theta_0)\right)\lambda_j^{2d_0}(2\log j)^a \\
=\;& O\left(\lambda_m^2\right)\hat{G}_{a,0}(d_0, \theta_0) \\
=\;& O_p\left((m/n)^2(\log^a m)\right) \\
=\;& o_p\left(\log^{-2} m\right).
\end{aligned} \tag{66}$$

The second equality holds by a mean-value expansion using the compactness of Θ, the third equation holds by (62) and $J_{a,b} = O(\log^a m)$. The last equality holds by Assumption 5. ■

Proof of (d). Given the same arguments as in Andrews & Sun (2004), we note that, (i) utilizing equations (62) and (65) we have that $\hat{G}_{a,b}(d_0, \theta) = J_{a,b} + o_p\left(\log^{-2} m\right)$ (ii) $J_{a,b} = O(\log^a m)$, (iii) $J_{0,0}J_{2,0} - J_{1,0}^2 = O(1)$ by replacing sums by integrals and noting that the part of $J_{0,0}J_{2,0}$ that is $O\left(\log^2 m\right)$ cancels with an identical term in $J_{1,0}^2$, (iv) $J_{0,0}J_{1,k-1} - J_{1,0}J_{0,k-1} = O(1)$ by the same argument as in (iii), and (v) $J_{0,0} = G_0 > 0$. Then from (i)-(v) and equation (61) it suffices to show

$$\sup_{(d,\theta)\in D_m(\eta_n)\times\Theta}\left|\hat{G}_{a,b}(d, \theta) - \hat{G}_{a,b}(d_0, \theta)\right| = o_p\left(\log^{-2} m\right). \tag{67}$$

Replacing λ_j^{2d} with j^{2d} in $\hat{G}_{a,b}(d, \theta)$, and thereby defining $\hat{E}_{a,b}(d, \theta)$, equation (61) also holds for $\hat{G}_{a,b}(d, \theta)$ replaced by $\hat{E}_{a,b}(d, \theta)$. Hence, it suffices in proving Lemma 1(d) that

$$Z_{a,b} = \sup_{(d,\theta)\in D_m(\eta_n)\times\Theta}\left|\hat{E}_{a,b}(d, \theta) - \hat{E}_{a,b}(d_0, \theta)\right| = o_p\left(n^{2d_0}\log^{-2} m\right), \tag{68}$$

$\forall a = 0, 1, 2$ and $b = 0, ..., r$. Then from Andrews & Sun (2004, pp. 600), $\exists C < \infty$ and for $(d, \theta) \in D_m(\eta_n) \times \Theta$

$$
\begin{aligned}
Z_{a,b} &= \sup_{(d,\theta) \in D_m(\eta_n) \times \Theta} \left| m^{-1} \sum_{j=1}^{m} I_j(d_0) \exp\left(P_r(\lambda_j, \theta)\right) (2 \log j)^a \left(\frac{j}{m}\right)^{2b} j^{2d_0} \left(j^{2(d-d_0)} - 1\right) \right| \\
&\leq C \sup_{d \in D_m(\eta_n)} m^{-1} \sum_{j=1}^{m} I_j(d_0) (\log j)^a j^{2d_0} \left| j^{2(d-d_0)} - 1 \right| + o_p(1) \\
&\leq 2C \exp\left(2\eta_n \log^{-4} m\right) \sup_{d \in D_m(\eta_n)} m^{-1} \sum_{j=1}^{m} I_j(d_0) (\log j)^{a+1} j^{2d_0} |d - d_0| \\
&\leq \eta_n \left(\log^{-2} m\right) 2C \exp\left(2\eta_n \log^{-4} m\right) m^{-1} \sum_{j=1}^{m} I_j(d_0) \lambda_j^{2d_0} \left(\frac{2\pi}{n}\right)^{-2d_0}.
\end{aligned}
$$

The first inequality follows from using $\sup_{0 \leq \pi \leq 2\pi, \theta \in \Theta} \sup_{j=1,...,m} \exp\left(P_r(\lambda_j, \theta)\right) < \infty$ because Θ is compact. The second inequality stems from noting

$$
\frac{\left| j^{2(d-d_0)} - 1 \right|}{|d - d_0|} \leq 2m^{2|d-d_0|} \log j \leq 2m^{2\eta_n \log^{-5} m} \log j = 2 \exp\left(2\eta_n \log^{-4} m\right) \log j
$$

for $d \in D_m(\eta_n)$ by a mean-value expansion where we use that $m^{\log^{-1} m} = e$. The third inequality uses $d \in D_m(\eta_n)$. Then from equations (62) and (65) we have $m^{-1} \sum_{j=1}^{m} I_j(d_0) \lambda_j^{2d_0} = \hat{G}_{0,0}(d_0, 0) = G_0 + o_p\left(\log^{-2} m\right)$. Hence, (68) follows. ■

Proof of (e). By using (62), and setting $a = b = 0$ we get $\hat{G}(d_0, \theta_0) = G_0\left(1 + o_p\left(\log^{-2} m\right)\right)$, and therefore the normalized score can be written as

$$
\begin{aligned}
B_n^{-1} S_n(d_0, \theta_0) &= \hat{G}^{-1}(d_0, \theta_0) m^{-1/2} \sum_{j=1}^{m} \left(\frac{I_j(d_0)}{g_j(d_0, \theta_0)} - m^{-1} \sum_{k=1}^{m} \frac{I_k(d_0)}{g_k(d_0, \theta_0)} \right) \tilde{X}_j \\
&= (1 + o_p(1)) m^{-1/2} \sum_{j=1}^{m} \left(\frac{I_j(d_0)}{g_j(d_0, \theta_0)} - 1 \right) \left(\tilde{X}_j - m^{-1} \sum_{k=1}^{m} \tilde{X}_k \right), \qquad (69)
\end{aligned}
$$

where

$$
\tilde{X}_j = \left(\log j, (j/m)^2, ..., (j/m)^{2r}\right)'. \qquad (70)
$$

Therefore, omitting the small order terms write the RHS of (69) as Andrews & Sun (2004, pp. 601), $T_{1,n}+T_{2,n}+T_{3,n}+T_{4,n}$, where

$$\begin{aligned} T_{1,n} &= m^{-1/2}\sum_{j=1}^{m}\left(\frac{I_j\left(d_0\right)}{g_j\left(d_0,\theta_0\right)}-2\pi I_{\varepsilon}\left(\lambda_j\right)-E\left(\frac{I_j\left(d_0\right)}{g_j\left(d_0,\theta_0\right)}-2\pi I_{\varepsilon}\left(\lambda_j\right)\right)\right) \quad (71)\\ &\quad\times\left(\tilde{X}_j-m^{-1}\sum_{k=1}^{m}\tilde{X}_k\right), \\ T_{2,n} &= m^{-1/2}\sum_{j=1}^{m}\left(\frac{E[I_j\left(d_0\right)]}{f_j(d_0)}-1\right)\frac{f_j\left(d_0\right)}{g_j(d_0,\theta_0)}\left(\tilde{X}_j-m^{-1}\sum_{k=1}^{m}\tilde{X}_k\right), \quad (72)\\ T_{3,n} &= m^{-1/2}\sum_{j=1}^{m}\left(2\pi I_{\varepsilon}\left(\lambda_j\right)-1\right)\left(\tilde{X}_j-m^{-1}\sum_{k=1}^{m}\tilde{X}_k\right), \quad (73)\\ T_{4,n} &= m^{-1/2}\sum_{j=1}^{m}\left(\frac{f_j\left(d_0\right)}{g_j(d_0,\theta_0)}-1\right)\left(\tilde{X}_j-m^{-1}\sum_{k=1}^{m}\tilde{X}_k\right), \quad (74)\end{aligned}$$

using that $E\left(2\pi I_{\varepsilon}\left(\lambda_j\right)\right)=1$. Next, we need to show that $T_{1,n}$ and $T_{4,n}$ are $o_p(1)$, $T_{2,n}=o\left(1\right)$, and $T_{3,n}\xrightarrow{d}N\left(0,\Omega_r\right)$. To show, $T_{1,n}=o_p(1)$, use summation by parts

$$\begin{aligned} T_{1,n} &= m^{-1/2}\sum_{k=1}^{m-1}\left(\tilde{X}_k-\tilde{X}_{k+1}\right)\sum_{j=1}^{k}\left(\frac{I_j\left(d_0\right)}{g_j\left(d_0,\theta_0\right)}-2\pi I_{\varepsilon}\left(\lambda_j\right)-E\left(\frac{I_j\left(d_0\right)}{g_j\left(d_0,\theta_0\right)}-2\pi I_{\varepsilon}\left(\lambda_j\right)\right)\right)\\ &\quad+\left(\tilde{X}_m-m^{-1}\sum_{k=1}^{m}\tilde{X}_k\right)m^{-1/2}\sum_{j=1}^{m}\left(\frac{I_j\left(d_0\right)}{g_j\left(d_0,\theta_0\right)}-2\pi I_{\varepsilon}\left(\lambda_j\right)-E\left(\frac{I_j\left(d_0\right)}{g_j\left(d_0,\theta_0\right)}-2\pi I_{\varepsilon}\left(\lambda_j\right)\right)\right)\\ &= m^{-1/2}\sum_{k=1}^{m-1}O\left(k^{-1}\right)O_p\left(k^{1/3}\log^{2/3}k+k^{\phi+1/2}n^{-\phi}+k^{1/2}n^{-1/4}\right)\\ &\quad+O(1)m^{-1/2}O_p\left(m^{1/3}\log^{2/3}m+m^{\phi+1/2}n^{-\phi}+m^{1/2}n^{-1/4}\right)\\ &= O_p\left(m^{-1/6}\log^{2/3}m+(m/n)^{\phi}+n^{-1/4}\right)\\ &= o_p(1), \quad (75)\end{aligned}$$

which follows from noting that $\tilde{X}_k-\tilde{X}_{k+1}=O\left(k^{-1}\right)$ uniformly over $k=1,...,m$ and $\tilde{X}_m-m^{-1}\sum_{k=1}^{m}\tilde{X}_k=O(1)$ which follows from approximating sums by integrals, see Andrews & Sun (2004, pp. 602). A remark is in order. Remember that under the assumption of linearity of the generating process, Assumption 4(a), together with Abadir et al. (2007, Lemma 4.4) and Lemma 2(ii), says that the behavior is the same for all $d\in(-1/2,\infty)$. Therefore, the results from Andrews & Sun (2004) also hold in our case. Since d_0 belongs to the interior of the admissible parameter space, $T_{1,n}=o_p(1)$. To prove that $T_{2,n}=o(1)$, we again utilize Assumption 4(a) together with Abadir et al. (2007, Lemma 4.4) and Lemma 2(ii) that enables us to use the result that

$$E\left(\frac{I_j\left(d_0\right)}{f_j\left(d_0\right)}\right)=1+O\left(j^{-1}\log j\right), \quad (76)$$

where $o(1) \to 0$ uniformly over $1 \leq j \leq m$ as $n \to \infty$. Then using (76), $T_{2,n}$ is bounded by

$$\begin{aligned}
T_{2,n} &= m^{-1/2} \sum_{j=1}^{m} O\left(j^{-1} \log j\right) O(1) \left(\tilde{X}_j - m^{-1} \sum_{k=1}^{m} \tilde{X}_k \right) \qquad (77) \\
&= O\left(m^{-1/2} \log m \sum_{j=1}^{m} j^{-1} \log j \right) \\
&= O\left(m^{-1/2} \log^3 m \right),
\end{aligned}$$

where we have used that $\tilde{X}_j - m^{-1} \sum_{k=1}^{m} \tilde{X}_k = O(\log m)$ uniformly in $1 \leq j \leq m$. Therefore, $T_{2,n} = o(1)$. Next, we need to show that $\forall \beta \neq 0$ $\beta' T_{3,n} \xrightarrow{d} N(0, \beta' \Omega_r \beta)$. That is, we need to verify that for $n \to \infty$

$$m^{-1} \sum_{j=1}^{m} \varsigma_j^2 \to \beta' \Omega_r \beta, \qquad (78)$$

where $\varsigma_j = \beta_j' \left(\tilde{X}_j - m^{-1} \sum_{k=1}^{m} \tilde{X}_k \right)$ and $\Omega_r = \begin{pmatrix} 4 & 2\mu_r' \\ 2\mu_r & \Gamma_r \end{pmatrix}$ which follows from Lemma 1(a), Lemma 1(d) and finally noting that $|\varsigma_j - \varsigma_{j+1}| \leq \|\beta\| \left\| \tilde{X}_j - \tilde{X}_{j+1} \right\| \leq C j^{-1}$ for some constant $C > 0$ independent of j. Finally, we need to show that $T_{4,n} = o_p(1)$. This follows from summation by parts and $\frac{f_j(d_0)}{g_j(d_0,\theta_0)} - 1 = O\left((j/n)^{\phi}\right)$ uniformly on $1 \leq j \leq m$, Frederiksen et al. (2008). This implies

$$\begin{aligned}
T_{4,n} &= m^{-1/2} \sum_{k=1}^{m-1} \left(\tilde{X}_k - \tilde{X}_{k+1} \right) \sum_{j=1}^{k} \left(\frac{f_j(d_0)}{g_j(d_0, \theta_0)} - 1 \right) \\
&\quad + \left(\tilde{X}_m - m^{-1} \sum_{k=1}^{m} \tilde{X}_k \right) m^{-1/2} \sum_{j=1}^{m} \left(\frac{f_j(d_0)}{g_j(d_0, \theta_0)} - 1 \right) \\
&= m^{-1/2} \sum_{k=1}^{m-1} O\left(k^{-1}\right) \sum_{j=1}^{k} O\left((j/n)^{\phi}\right) + O(1) m^{-1/2} \sum_{j=1}^{m} O\left((j/n)^{\phi}\right) \\
&= O\left(m^{-1/2+\phi} n^{-\phi} \right) \\
&= o_p(1), \qquad (79)
\end{aligned}$$

where the last equality holds by Assumption 5. ■

Lemma 2 *Assume that the sequence $\{v_j\}$ is given as in (21). The following holds uniformly in $1 \leq k < j \leq m = o(n)$, as $n \to \infty$. (i) If f_u satisfies Assumption 2, then*

$$E\left[|w_u(\lambda_j)|^2 / f_u(\lambda_j) \right] = 1 + o(1) + O\left(j^{-1} \log j\right), \qquad (80)$$

where $o(1) \to 0$ uniformly in $1 \leq j \leq m$, as $n \to \infty$, and

$$\begin{aligned}
|E[v_j \overline{v}_j]| + |E[v_j v_k]| &= O\left(\frac{\log j}{j - k} \right) + O\left(\frac{\log j}{k^{|d|} j^{1-|d|}} \right), \qquad (81) \\
|E[v_j v_j]| &= O\left(\frac{\log j}{j} \right). \qquad (82)
\end{aligned}$$

(ii) If f_u satisfies Assumption 2 and 4(c), then

$$E\left[\left|w_u(\lambda_j)\right|^2 / f_u(\lambda_j)\right] = 1 + O\left(j^{-1}\log j\right), \tag{83}$$

and

$$\begin{aligned} \left|E\left[v_j \overline{v}_j\right]\right| + \left|E\left[v_j v_k\right]\right| &= O\left(\frac{\log j}{k^{|d|} j^{1-|d|}}\right), && (84) \\ \left|E\left[v_j v_j\right]\right| &= O\left(\frac{\log j}{j}\right). && (85) \end{aligned}$$

Proof. Follows from Abadir et al. (2007) and their proof to Lemma 4.6, given Assumption 3 and by interchanging b_0 by $G_0 \exp\left(-P_r(\lambda_j, \theta)\right)$. Let $d_0 = p_0 + d_u$. Then for $d_0 = d_u$ equations (80)-(82) and (83)-(85) follow from Robinson (1995*b*) and his proof of Theorem 2 pp. 1060. For $p_0 \in \mathbb{N}\backslash\{0\}$ and the property of the extended DFT and the rescaled extended DFT, (18) and (21), respectively, it follows

$$v_j = \left(\frac{1-\exp(i\lambda_j)}{\lambda_j}\right)^{-p_0} \frac{w_u(\lambda_j)}{\varphi(\lambda)^{1/2}\lambda_j^{-d_u}}. \tag{86}$$

As $\left|\frac{1-\exp(i\lambda_j)}{\lambda_j}\right|^{-p_0} \leq C$ uniformly in $1 \leq j \leq m$ (81)-(82) and (84)-(85) also hold for $p_0 \in \mathbb{N}\backslash\{0\}$. ∎

References

Abadir, K. M., Distaso, W. & Giraitis, L. (2007), 'Nonstationarity-extended local whittle estimation', *Journal of Econometrics* **141**, 1353–1384.

Abadir, K. M. & Taylor, R. (1999), 'On the definitions of (co-)integration', *Journal of Time Series Analysis* **20**, 129–13.

Andrews, D. W. K. & Guggenberger, P. (2003), 'A bias-reduced log-periodogram regression estimator for the long memory parameter', *Econometrica* **71**, 675–712.

Andrews, D. W. K. & Sun, Y. (2004), 'Adaptive local polynomial Whittle estimation of long-range dependence', *Econometrica* **72**, 569–614.

Arteche, J. (2004), 'Gaussian semiparametric estimation in long memory in stochastic volatility and signal plus noise models', *Journal of Econometrics* **119**, 131–154.

Beran, J. (1994), *Statistics for Long-Memory Processes*, Chapman-Hall, New York.

Black, F. & Cox, J. C. (1976), 'Valuing corporate securities: Some effects of bond indenture provisions', *Journal of Finance* **31**, 351–67.

Dalla, V., Giraitis, L. & Hidalgo, J. (2006), 'Consistent estimation of the memory parameter for nonlinear time series', *Journal of Time Series Analysis* **27**, 211–251.

Das, S. (1995), 'Credit risk derivatives', *The Journal of Derivatives* **2**, 7–23.

Das, S. & Tufano, P. (1995), 'Pricing credit-sensitive debt when interest rates, credit ratings and credit spreads are stochastic', *Journal of Financial Engineering* **5**, 161–198.

Davies, R. B. & Harte, D. S. (1987), 'Tests for hurst effects', *Biometrika* **74**, 95–102.

Diebold, F. X. & Inoue, A. (2001), 'Long memory and regime switching', *Journal of Econometrics* **105**, 131–159.

Duffie, D. & Huang, M. (1996), 'Swap rates and credit quality', *Journal of Finance* **51**, 921–49.

Duffie, D. & Singleton, K. J. (1999), 'Modeling term structures of defaultable bonds', *Review of Financial Studies* **12**, 687–720.

Frederiksen, P. H. & Nielsen, F. S. (2008), 'Testing for spurious long memory in potentially nonstationary perturbed fractional processes', *CREATES RP 2008-59, University of Aarhus, and Working Paper, Nordea Markets* .

Frederiksen, P. H., Nielsen, F. S. & Nielsen, M. Ø. (2008), 'Local polynomial Whittle estimation of perturbed fractional processes', *CREATES RP 2008-29, University of Aarhus, and Working Paper, Nordea Markets and Cornell University* .

Geweke, J. & Porter-Hudak, S. (1983), 'The estimation and application of long-memory time series models', *Journal of Time Series Analysis* **4**, 221–238.

Granger, C. W. J. & Hyung, N. (2004), 'Occasional structural breaks and long memory with an application to the s&p 500 absolute stock returns', *Journal of Empirical Finance* **11**, 399–421.

Granger, C. W. J. & Terasvirta, T. (1999), 'A simple nonlinear time series model with misleading linear properties', *Economics Letters* **62**, 161–165.

Haldrup, N. & Nielsen, M. Ø. (2007), 'Estimation of fractional integration in the presence of data noise', *Computational Statistics and Data Analysis* **51**, 3100–3114.

Hull, J. & White, A. (1995), 'The impact of default risk on the prices of options and other derivative securities', *Journal of Banking & Finance* **19**, 299–322.

Hurvich, C. M. & Chen, W. W. (2000), 'An efficient taper for potentially overdifferenced long-memory time series', *Journal of Time Series Analysis* **21**, 155–180.

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

Hurvich, C. M. & Ray, B. K. (1995), 'Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes', *Journal of Time Series Analysis* **16**, 17–41.

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

Jarrow, R. A., Lando, D. & Turnbull, S. M. (1997), 'A markov model for the term structure of credit risk spreads', *Review of Financial Studies* **10**, 481–523.

Jarrow, R. A. & Turnbull, S. M. (1995), 'Pricing derivatives on financial securities subject to credit risk', *Journal of Finance* **50**, 53–85.

Kim, C. S. & Phillips, P. C. B. (2006), Log periodogram regression: The nonstationary case, Cowles Foundation Discussion Papers 1587, Cowles Foundation, Yale University.

Künsch, H. R. (1987), Statistical aspects of self-similar processes, *in* Y. Prokhorov & V. V. Sazanov, eds, 'Proceedings of the First World Congress of the Bernoulli Society', VNU Science Press, Utrecht, pp. 67–74.

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

Leland, H. E. & Toft, K. B. (1996), 'Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads', *Journal of Finance* **51**, 987–1019.

Longstaff, F. A. & Schwartz, E. (1995), 'Valuing credit derivatives', *Journal of Fixed Income* **5**, 6–12.

Madan, D. & Unal, H. (1996), Pricing the risks of default, Center for Financial Institutions Working Papers 94-16, Wharton School Center for Financial Institutions, University of Pennsylvania.

Manzoni, K. (2002), 'Modeling credit spreads: An application to the sterling eurobond market', *International Review of Financial Analysis* **11**, 183–218.

Marinucci, D. & Robinson, P. M. (1999), 'Alternative forms of fractional brownian motion', *Journal of Statistical Planning and Inference* **80**, 111–122.

Merton, R. C. (1974), 'On the pricing of corporate debt: The risk structure of interest rates', *Journal of Finance* **29**, 449–70.

Newey, W. K. & McFadden, D. (1986), Large sample estimation and hypothesis testing, *in* R. F. Engle & D. McFadden, eds, 'Handbook of Econometrics'.

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

Ohanissian, A., Russell, J. R. & Tsay, R. S. (2008), 'True or spurious long memory? a new test', *Journal of Business & Economic Statistics* **26**(2), 161–175.

Phillips, P. C. B. (1999), 'Discrete fourier transformation of fractional processes', *Discussion Paper* **1243**(Yale University (Cowles Foundation)).

Phillips, P. C. B. & Shimotsu, K. (2004), 'Local whittle estimation in nonstationary and unit root cases', *The Annals of Statistics* **32**, 656–692.

Ramaswamy, K. & Sundaresan, S. M. (1986), 'The valuation of floating-rate instruments : Theory and evidence', *Journal of Financial Economics* **17**, 251–272.

Ratta, L. C. & Urga, G. (2005), Modeling credit spreads: A fractional integration approach, Working Paper CEA-07-2005, Cass Business School, City University London.

Robinson, P. M. (1994), 'Semiparametric analysis of long-memory time series', *The Annals of Statistics* **22**, 515–539.

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

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

Robinson, P. M. (2005), 'The distance between nonstationary fractional processes', *Journal of Econometrics* **128**, 195–236.

Shimotsu, K. (2006), 'Exact local whittle estimation of fractional integration with unknown mean and time trend', *Working Paper, Department of Economics, Queen's University, Canada* (1061).

Shimotsu, K. & Phillips, P. (2005), 'Exact local whittle estimation of fractional integration', *The Annals of Statistics* **33**, 1890–1933.

Tanaka, K. (1999), 'The nonstationary fractional unit root', *Econometric Theory* **15**, 549–582.

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

Velasco, C. (1999*b*), 'Non-stationary log-periodogram regression', *Journal of Econometrics* **91**, 325–371.

Table 1: Simulation results for ARFIMA(0,d,0) with n = 512.

	LW		LPW (r=1)		LPW (r=2)		ExtLW		ExtLPW (r=1)		ExtLPW (r=2)	
d	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Panel A: $m = \lfloor n^{0.5} \rfloor$												
-0.3	-0.0087	0.1450	-0.0490	0.2708	-0.0546	0.3984	-0.0059	0.1385	-0.0309	0.2459	-0.0219	0.3769
0	-0.0123	0.1385	-0.0548	0.2659	-0.0616	0.3958	-0.0119	0.1387	-0.0500	0.2549	-0.0287	0.3724
0.3	-0.0162	0.1413	-0.0476	0.2631	-0.0670	0.3972	-0.0161	0.1423	-0.0510	0.2595	-0.0634	0.3893
0.7	0.0117	0.1480	-0.0432	0.2729	-0.0359	0.3975	-0.0020	0.1462	-0.0539	0.2667	-0.0464	0.4003
1	-0.0208	0.1273	-0.0542	0.2370	-0.0570	0.3352	-0.0221	0.1427	-0.0642	0.2609	-0.0663	0.3750
1.3	-0.1937	0.2314	-0.1935	0.2769	-0.1902	0.3459	-0.0138	0.1395	-0.0415	0.2568	-0.0423	0.3483
1.7	-0.5869	0.6103	-0.5540	0.5927	-0.5408	0.5979	-0.0053	0.1393	-0.0184	0.2256	-0.0153	0.3272
2	-0.9094	0.9256	-0.8703	0.8969	-0.8478	0.8851	-0.0146	0.1383	-0.0492	0.2547	-0.0352	0.3298
Panel B: $m = \lfloor n^{0.65} \rfloor$												
-0.3	0.0008	0.0771	-0.0151	0.1283	-0.0176	0.1807	0.0007	0.0777	-0.0125	0.1223	-0.0091	0.1706
0	-0.0017	0.0763	-0.0168	0.1279	-0.0203	0.1740	-0.0017	0.0763	-0.0167	0.1275	-0.0198	0.1723
0.3	-0.0075	0.0783	-0.0230	0.1309	-0.0232	0.1783	-0.0071	0.0793	-0.0226	0.1319	-0.0192	0.1862
0.7	0.0100	0.0806	0.0031	0.1359	0.0019	0.1849	-0.0037	0.0781	-0.0097	0.1329	-0.0133	0.1821
1	-0.0140	0.0691	-0.0205	0.1165	-0.0287	0.1598	-0.0160	0.0772	-0.0278	0.1309	-0.0315	0.1759
1.3	-0.2141	0.2344	-0.1961	0.2293	-0.1904	0.2405	-0.0128	0.0797	-0.0169	0.1266	-0.0136	0.1810
1.7	-0.6229	0.6382	-0.5882	0.6103	-0.5709	0.5986	-0.0158	0.0770	-0.0100	0.1224	-0.0043	0.1625
2	-0.9506	0.9581	-0.9177	0.9311	-0.8990	0.9165	-0.0186	0.0785	-0.0210	0.1282	-0.0206	0.1789
Panel C: $m = \lfloor n^{0.8} \rfloor$												
-0.3	0.0115	0.0448	-0.0021	0.0713	0.0004	0.0949	0.0115	0.0448	-0.0018	0.0705	0.0017	0.0907
0	-0.0040	0.0435	-0.0093	0.0709	-0.0107	0.0941	-0.0040	0.0435	-0.0093	0.0709	-0.0107	0.0941
0.3	-0.0093	0.0446	-0.0067	0.0711	-0.0067	0.0952	-0.0093	0.0446	-0.0067	0.0711	-0.0068	0.0950
0.7	-0.0147	0.0510	0.0079	0.0738	0.0083	0.0977	-0.0280	0.0537	-0.0075	0.0708	-0.0077	0.0917
1	-0.0363	0.0531	-0.0065	0.0603	-0.0082	0.0817	-0.0378	0.0574	-0.0090	0.0699	-0.0115	0.0933
1.3	-0.2528	0.2662	-0.2052	0.2261	-0.2001	0.2257	-0.0472	0.0644	-0.0085	0.0703	-0.0111	0.0944
1.7	-0.6847	0.6918	-0.6315	0.6431	-0.6187	0.6327	-0.0595	0.0734	-0.0007	0.0690	-0.0014	0.0912
2	-0.9950	1.0001	-0.9439	0.9523	-0.9308	0.9417	-0.0746	0.0876	-0.0129	0.0703	-0.0146	0.0954

Notes: LW, LPW, ExtLW, and ExtLPW denotes the local Whittle estimator of Robinson (1995a), local polynomial Whittle estimator of Andrews & Sun (2004), extended local Whittle estimator of Abadir et al. (2007), and our proposed estimator the extended local polynomial Whittle estimator, respectively. r denotes the degree of parameterization of the polynomial, i.e. $P_r = \sum_{\tau=1}^{r} \theta_\tau \lambda_j^{2\tau}$.

Table 2: Simulation results for ARFIMA(0,d,1) with $\beta = -0.8$ and n = 512.

	LW		LPW (r=1)		LPW (r=2)		ExtLW		ExtLPW (r=1)		ExtLPW (r=2)	
d	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Panel A: $m = \lfloor n^{0.5} \rfloor$												
-0.3	-0.1159	0.1940	-0.0132	0.2611	0.0077	0.3638	-0.1184	0.1950	-0.0054	0.2493	0.0192	0.3509
0	-0.1669	0.2168	-0.0859	0.2771	-0.0816	0.3852	-0.1666	0.2167	-0.0801	0.2630	-0.0694	0.3417
0.3	-0.1606	0.2151	-0.0842	0.2659	-0.0609	0.3838	-0.1604	0.2156	-0.0944	0.2531	-0.0842	0.3558
0.7	-0.1444	0.2101	-0.0565	0.2659	-0.0412	0.3699	-0.0826	0.1932	0.0003	0.2822	0.0053	0.3946
1	-0.1275	0.1866	-0.0739	0.2553	-0.0611	0.3594	-0.1654	0.2164	-0.0879	0.2674	-0.0636	0.3742
1.3	-0.2494	0.2746	-0.2099	0.2973	-0.2036	0.3608	-0.1727	0.2248	-0.1027	0.2585	-0.0998	0.3398
1.7	-0.6037	0.6207	-0.5666	0.6007	-0.5428	0.6014	-0.0851	0.1889	-0.0074	0.2669	0.0142	0.3400
2	-0.9138	0.9276	-0.8721	0.8978	-0.8456	0.8858	-0.1677	0.2188	-0.0775	0.2573	-0.0470	0.3369
Panel B: $m = \lfloor n^{0.65} \rfloor$												
-0.3	-0.3216	0.3365	-0.1284	0.1915	-0.0467	0.1859	-0.3378	0.3505	-0.1342	0.1972	-0.0463	0.1846
0	-0.3603	0.3699	-0.1773	0.2215	-0.1025	0.2091	-0.3580	0.3682	-0.1769	0.2210	-0.0999	0.2039
0.3	-0.3697	0.3794	-0.1875	0.2274	-0.1039	0.2030	-0.3697	0.3794	-0.1875	0.2274	-0.1041	0.2033
0.7	-0.3467	0.3586	-0.1609	0.2115	-0.0774	0.1976	-0.3388	0.3541	-0.1262	0.2098	-0.0325	0.2156
1	-0.3056	0.3225	-0.1326	0.1854	-0.0645	0.1754	-0.3653	0.3741	-0.1816	0.2215	-0.0984	0.1991
1.3	-0.3311	0.3372	-0.2439	0.2644	-0.2084	0.2545	-0.3748	0.3844	-0.1850	0.2271	-0.1040	0.2032
1.7	-0.6457	0.6523	-0.6045	0.6209	-0.5822	0.6075	-0.3493	0.3623	-0.1951	0.2299	-0.1501	0.2253
2	-0.9516	0.9581	-0.9168	0.9301	-0.8965	0.9153	-0.3674	0.3756	-0.1834	0.2245	-0.1026	0.2009
Panel C: $m = \lfloor n^{0.8} \rfloor$												
-0.3	-0.4946	0.4996	-0.3523	0.3650	-0.2386	0.2636	-0.5094	0.5133	-0.3693	0.3797	-0.2290	0.2549
0	-0.5333	0.5362	-0.3930	0.4006	-0.2866	0.3028	-0.4455	0.4739	-0.3889	0.3958	-0.2859	0.3015
0.3	-0.5469	0.5500	-0.3992	0.4073	-0.2960	0.3128	-0.5469	0.5500	-0.3992	0.4073	-0.2960	0.3128
0.7	-0.5359	0.5397	-0.3712	0.3805	-0.2614	0.2808	-0.5359	0.5397	-0.3696	0.3798	-0.2183	0.2556
1	-0.4980	0.5046	-0.3223	0.3371	-0.2216	0.2462	-0.5145	0.5167	-0.3855	0.3920	-0.2836	0.2974
1.3	-0.4657	0.4730	-0.3425	0.3493	-0.2900	0.2988	-0.5739	0.5768	-0.3986	0.4065	-0.2928	0.3081
1.7	-0.7071	0.7086	-0.6516	0.6563	-0.6341	0.6427	-0.5625	0.5665	-0.3740	0.3834	-0.2601	0.2825
2	-1.0010	1.0035	-0.9502	0.9555	-0.9364	0.9444	-0.5301	0.5336	-0.3890	0.3962	-0.3080	0.3309

Notes: LW, LPW, ExtLW, and ExtLPW denotes the local Whittle estimator of Robinson (1995a), local polynomial Whittle estimator of Andrews & Sun (2004), extended local Whittle estimator of Abadir et al. (2007), and our proposed estimator the extended local polynomial Whittle estimator, respectively. r denotes the degree of parameterization of the polynomial, i.e. $P_r = \sum_{\tau=1}^{r} \theta_\tau \lambda_j^{2\tau}$.

Table 3: Simulation results for ARFIMA(0,d,1) with $\beta = -0.5$ and n = 512.

	LW		LPW (r=1)		LPW (r=2)		ExtLW		ExtLPW (r=1)		ExtLPW (r=2)	
d	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Panel A: $m = \lfloor n^{0.5} \rfloor$												
-0.3	-0.0293	0.1431	-0.0478	0.2630	-0.0522	0.3787	-0.0291	0.1429	-0.0364	0.2503	-0.0240	0.3537
0	-0.0365	0.1509	-0.0553	0.2649	-0.0646	0.3864	-0.0364	0.1505	-0.0496	0.2591	-0.0437	0.3666
0.3	-0.0384	0.1527	-0.0485	0.2676	-0.0506	0.3782	-0.0402	0.1505	-0.0578	0.2564	-0.0580	0.3645
0.7	-0.0154	0.1448	-0.0278	0.2758	-0.0470	0.3862	-0.0191	0.1450	-0.0355	0.2747	-0.0500	0.3926
1	-0.0329	0.1306	-0.0441	0.2481	-0.0453	0.3532	-0.0406	0.1476	-0.0509	0.2647	-0.0428	0.3865
1.3	-0.2058	0.2402	-0.2073	0.2842	-0.2119	0.3359	-0.0422	0.1496	-0.0615	0.2542	-0.0711	0.3341
1.7	-0.5915	0.6136	-0.5554	0.5931	-0.5365	0.5968	-0.0186	0.1394	-0.0027	0.2247	0.0105	0.3361
2	-0.9134	0.9269	-0.8732	0.8968	-0.8480	0.8828	-0.0383	0.1435	-0.0383	0.2535	-0.0236	0.3273
Panel B: $m = \lfloor n^{0.65} \rfloor$												
-0.3	-0.0956	0.1225	-0.0233	0.1318	-0.0114	0.1763	-0.0962	0.1237	-0.0223	0.1291	-0.0074	0.1695
0	-0.1028	0.1288	-0.0365	0.1352	-0.0320	0.1733	-0.1028	0.1288	-0.0365	0.1349	-0.0313	0.1710
0.3	-0.1014	0.1269	-0.0334	0.1326	-0.0249	0.1802	-0.1014	0.1269	-0.0335	0.1329	-0.0248	0.1818
0.7	-0.0901	0.1241	-0.0170	0.1363	-0.0087	0.1782	-0.0956	0.1294	-0.0215	0.1383	-0.0135	0.1802
1	-0.0827	0.1112	-0.0247	0.1236	-0.0207	0.1654	-0.1082	0.1317	-0.0294	0.1280	-0.0216	0.1737
1.3	-0.2358	0.2479	-0.2047	0.2371	-0.1989	0.2462	-0.1135	0.1372	-0.0371	0.1390	-0.0351	0.1815
1.7	-0.6342	0.6452	-0.5990	0.6177	-0.5806	0.6057	-0.0969	0.1270	-0.0147	0.1290	-0.0031	0.1680
2	-0.9528	0.9590	-0.9192	0.9315	-0.8992	0.9165	-0.1168	0.1402	-0.0360	0.1377	-0.0272	0.1751
Panel C: $m = \lfloor n^{0.8} \rfloor$												
-0.3	-0.2392	0.2441	-0.1047	0.1262	-0.0386	0.0975	-0.2278	0.2328	-0.1029	0.1232	-0.0374	0.0947
0	-0.2582	0.2629	-0.1142	0.1348	-0.0565	0.1099	-0.2582	0.2629	-0.1142	0.1348	-0.0565	0.1099
0.3	-0.2655	0.2696	-0.1117	0.1318	-0.0493	0.1037	-0.2655	0.2696	-0.1117	0.1318	-0.0493	0.1037
0.7	-0.2621	0.2679	-0.0966	0.1223	-0.0314	0.1042	-0.2583	0.2644	-0.1023	0.1254	-0.0356	0.1029
1	-0.2287	0.2390	-0.0781	0.1084	-0.0312	0.0916	-0.2852	0.2895	-0.1118	0.1324	-0.0486	0.1037
1.3	-0.3183	0.3200	-0.2354	0.2468	-0.2146	0.2359	-0.2966	0.3006	-0.1146	0.1369	-0.0530	0.1101
1.7	-0.6865	0.6913	-0.6289	0.6394	-0.6128	0.6278	-0.2818	0.2884	-0.1125	0.1388	-0.0490	0.1148
2	-0.9998	1.0036	-0.9493	0.9567	-0.9370	0.9468	-0.3191	0.3229	-0.1169	0.1359	-0.0543	0.1058

Notes: LW, LPW, ExtLW, and ExtLPW denotes the local Whittle estimator of Robinson (1995a), local polynomial Whittle estimator of Andrews & Sun (2004), extended local Whittle estimator of Abadir et al. (2007), and our proposed estimator the extended local polynomial Whittle estimator, respectively. r denotes the degree of parameterization of the polynomial, i.e. $P_r = \sum_{\tau=1}^{r} \theta_\tau \lambda_j^{2\tau}$.

Table 4: Simulation results for ARFIMA(1,d,0) with $\alpha = 0.8$ and n = 512.

	LW		LPW (r=1)		LPW (r=2)		ExtLW		ExtLPW (r=1)		ExtLPW (r=2)	
d	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Panel A: $m = \lfloor n^{0.5} \rfloor$												
-0.3	0.1530	0.2114	-0.0133	0.2535	-0.0323	0.3805	0.1530	0.2114	-0.0012	0.2475	0.0080	0.3891
0	0.1404	0.2015	-0.0290	0.2553	-0.0572	0.3771	0.1413	0.2034	-0.0248	0.2552	-0.0260	0.3647
0.3	0.1436	0.2030	-0.0266	0.2505	-0.0578	0.3806	0.1477	0.2088	-0.0237	0.2511	-0.0479	0.3733
0.7	0.1518	0.2117	0.0021	0.2495	-0.0041	0.3575	0.1421	0.2039	-0.0129	0.2413	-0.0126	0.3671
1	0.0751	0.1567	-0.0265	0.2323	-0.0582	0.3499	0.1385	0.2036	-0.0201	0.2608	-0.0431	0.3671
1.3	-0.1604	0.2210	-0.1813	0.2621	-0.1960	0.3381	0.1511	0.2108	0.0029	0.2683	-0.0195	0.3720
1.7	-0.5879	0.6133	-0.5554	0.5946	-0.5425	0.6058	0.1383	0.1978	0.0115	0.2225	0.0040	0.3320
2	-0.9022	0.9199	-0.8565	0.8878	-0.8322	0.8761	0.1407	0.2023	0.0088	0.2670	0.0192	0.3579
Panel B: $m = \lfloor n^{0.65} \rfloor$												
-0.3	0.4058	0.4144	0.1595	0.2037	0.0584	0.1855	0.4074	0.4171	0.1597	0.2036	0.0624	0.1791
0	0.4053	0.4138	0.1634	0.2103	0.0602	0.1916	0.4163	0.4276	0.1637	0.2112	0.0609	0.1937
0.3	0.4026	0.4106	0.1597	0.2078	0.0658	0.1873	0.4114	0.4195	0.1692	0.2216	0.0722	0.1983
0.7	0.3729	0.3816	0.1671	0.2121	0.0766	0.1898	0.3962	0.4049	0.1565	0.2064	0.0609	0.1853
1	0.1873	0.2329	0.0985	0.1590	0.0338	0.1653	0.4045	0.4154	0.1732	0.2295	0.0707	0.2103
1.3	-0.1653	0.2412	-0.1598	0.2219	-0.1681	0.2314	0.4058	0.4139	0.1935	0.2335	0.1281	0.2201
1.7	-0.6165	0.6385	-0.5792	0.6080	-0.5606	0.5938	0.3895	0.3988	0.1600	0.2082	0.0842	0.2217
2	-0.9361	0.9489	-0.8980	0.9193	-0.8765	0.9034	0.3864	0.3940	0.2500	0.3172	0.2115	0.3369
Panel C: $m = \lfloor n^{0.8} \rfloor$												
-0.3	0.6635	0.6655	0.4595	0.4665	0.3122	0.3279	0.6687	0.6713	0.4595	0.4665	0.3122	0.3279
0	0.6490	0.6510	0.4583	0.4648	0.3058	0.3202	0.6603	0.6623	0.4641	0.4721	0.3061	0.3209
0.3	0.6367	0.6386	0.4640	0.4701	0.3132	0.3278	0.6406	0.6425	0.4731	0.4791	0.3226	0.3384
0.7	0.5339	0.5415	0.4318	0.4393	0.3026	0.3175	0.6209	0.6232	0.4649	0.4713	0.3088	0.3244
1	0.1922	0.2685	0.2135	0.2625	0.1693	0.2101	0.6120	0.6143	0.4876	0.4958	0.3438	0.3641
1.3	-0.2298	0.2829	-0.1696	0.2415	-0.1630	0.2294	0.5840	0.5863	0.4694	0.4757	0.3290	0.3434
1.7	-0.6783	0.6883	-0.6187	0.6384	-0.6017	0.6267	0.5514	0.5547	0.4572	0.4637	0.3158	0.3378
2	-1.0016	1.0052	-0.9513	0.9584	-0.9386	0.9486	0.4845	0.4854	0.4534	0.4568	0.4315	0.4466

Notes: LW, LPW, ExtLW, and ExtLPW denotes the local Whittle estimator of Robinson (1995*a*), local polynomial Whittle estimator of Andrews & Sun (2004), extended local Whittle estimator of Abadir et al. (2007), and our proposed estimator the extended local polynomial Whittle estimator, respectively. r denotes the degree of parameterization of the polynomial, i.e. $P_r = \sum_{\tau=1}^{r} \theta_\tau \lambda_j^{2\tau}$.

Table 5: Simulation results for ARFIMA(1,d,0) with $\alpha = 0.5$ and n = 512.

	LW		LPW (r=1)		LPW (r=2)		ExtLW		ExtLPW (r=1)		ExtLPW (r=2)	
d	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Panel A: $m = \lfloor n^{0.5} \rfloor$												
-0.3	0.0108	0.1404	-0.0426	0.2469	-0.0382	0.3678	0.0123	0.1369	-0.0305	0.2400	-0.0107	0.3774
0	0.0088	0.1441	-0.0598	0.2682	-0.0818	0.3961	0.0086	0.1448	-0.0534	0.2642	-0.0529	0.3668
0.3	0.0167	0.1470	-0.0621	0.2660	-0.0754	0.4006	0.0174	0.1501	-0.0616	0.2690	-0.0690	0.3854
.7	0.0295	0.1451	-0.0210	0.2485	-0.0165	0.3702	0.0131	0.1389	-0.0350	0.2460	-0.0219	0.3864
1	0.0104	0.1251	-0.0399	0.2309	-0.0464	0.3604	0.0159	0.1391	-0.0450	0.2528	-0.0415	0.4000
1.3	-0.1915	0.2349	-0.2109	0.2887	-0.2113	0.3579	0.0043	0.1407	-0.0547	0.2594	-0.0395	0.3600
1.7	-0.5861	0.6113	-0.5580	0.5953	-0.5428	0.6002	0.0121	0.1422	-0.0217	0.2276	-0.0136	0.3212
2	-0.9062	0.9227	-0.8612	0.8891	-0.8321	0.8728	0.0094	0.1363	-0.0286	0.2457	-0.0047	0.3275
Panel B: $m = \lfloor n^{0.65} \rfloor$												
-0.3	0.0950	0.1221	-0.0070	0.1296	-0.0261	0.1750	0.0950	0.1221	-0.0052	0.1263	-0.0185	0.1624
0	0.0916	0.1194	-0.0064	0.1260	-0.0234	0.1734	0.0916	0.1194	-0.0064	0.1259	-0.0226	0.1707
0.3	0.0921	0.1191	-0.0060	0.1256	-0.0191	0.1741	0.0946	0.1240	-0.0049	0.1285	-0.0174	0.1776
0.7	0.1027	0.1300	0.0168	0.1320	0.0027	0.1756	0.0913	0.1208	-0.0014	0.1262	-0.0146	0.1712
1	0.0476	0.0872	-0.0110	0.1205	-0.0237	0.1710	0.0827	0.1130	-0.0104	0.1317	-0.0216	0.1833
1.3	-0.1954	0.2285	-0.1931	0.2309	-0.1906	0.2428	0.0859	0.1181	0.0065	0.1407	-0.0026	0.1892
1.7	-0.6226	0.6379	-0.5876	0.6102	-0.5697	0.5974	0.0808	0.1137	-0.0012	0.1286	-0.0093	0.1665
2	-0.9521	0.9602	-0.9200	0.9340	-0.9012	0.9197	0.0823	0.1135	-0.0068	0.1350	-0.0134	0.1877
Panel C: $m = \lfloor n^{0.8} \rfloor$												
-0.3	0.3043	0.3081	0.1112	0.1331	0.0402	0.1017	0.3043	0.3081	0.1112	0.1331	0.0404	0.1012
0	0.2897	0.2937	0.1054	0.1274	0.0330	0.1019	0.2897	0.2937	0.1054	0.1274	0.0330	0.1019
0.3	0.2804	0.2844	0.1064	0.1273	0.0364	0.0988	0.2892	0.2935	0.1072	0.1293	0.0364	0.0989
0.7	0.2588	0.2632	0.1173	0.1385	0.0484	0.1105	0.2631	0.2673	0.1081	0.1292	0.0360	0.1002
1	0.1107	0.1465	0.0676	0.0966	0.0255	0.0864	0.2497	0.2542	0.1088	0.1323	0.0366	0.1014
1.3	-0.2351	0.2676	-0.1952	0.2263	-0.1958	0.2240	0.2460	0.2512	0.1322	0.1574	0.0644	0.1258
1.7	-0.6751	0.6867	-0.6195	0.6366	-0.6049	0.6256	0.2178	0.2232	0.1103	0.1317	0.0378	0.1010
2	-0.9966	1.0010	-0.9449	0.9528	-0.9317	0.9419	0.1978	0.2041	0.1068	0.1293	0.0332	0.1030

Notes: LW, LPW, ExtLW, and ExtLPW denotes the local Whittle estimator of Robinson (1995*a*), local polynomial Whittle estimator of Andrews & Sun (2004), extended local Whittle estimator of Abadir et al. (2007), and our proposed estimator the extended local polynomial Whittle estimator, respectively. r denotes the degree of parameterization of the polynomial, i.e. $P_r = \sum_{\tau=1}^{r} \theta_\tau \lambda_j^{2\tau}$.

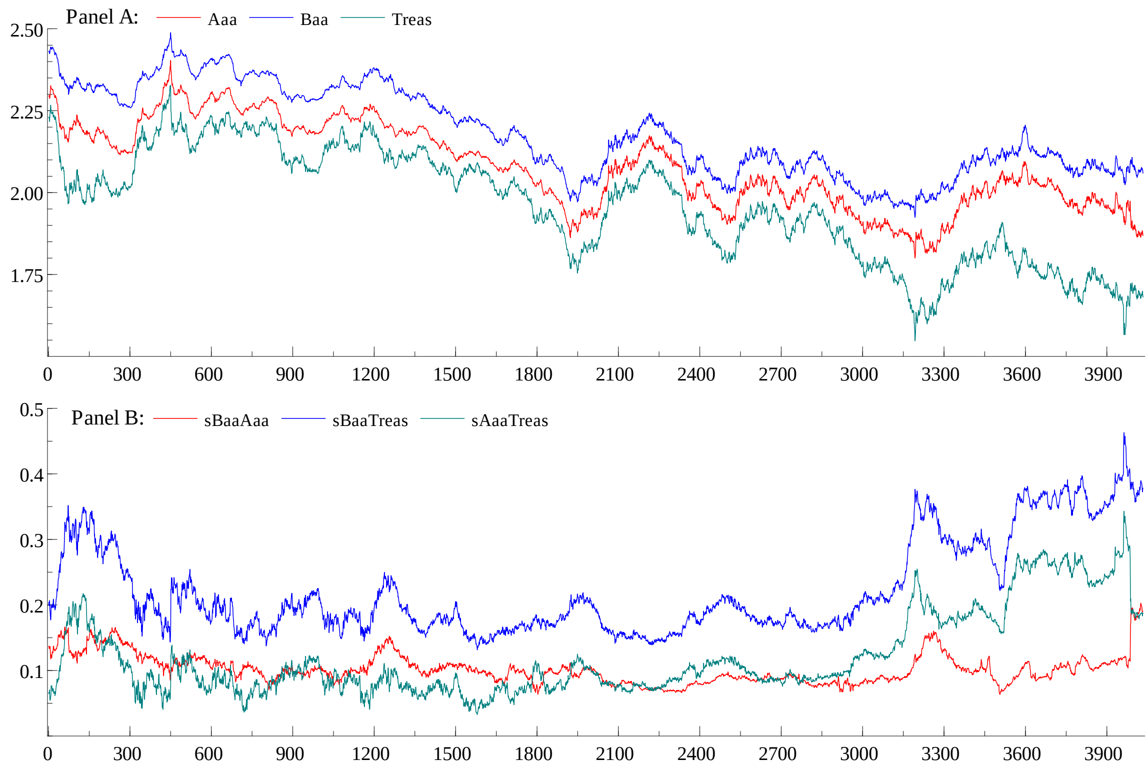


Figure 1: Time series plot of log yields (Panel A) and their respective spreads (Panel B).

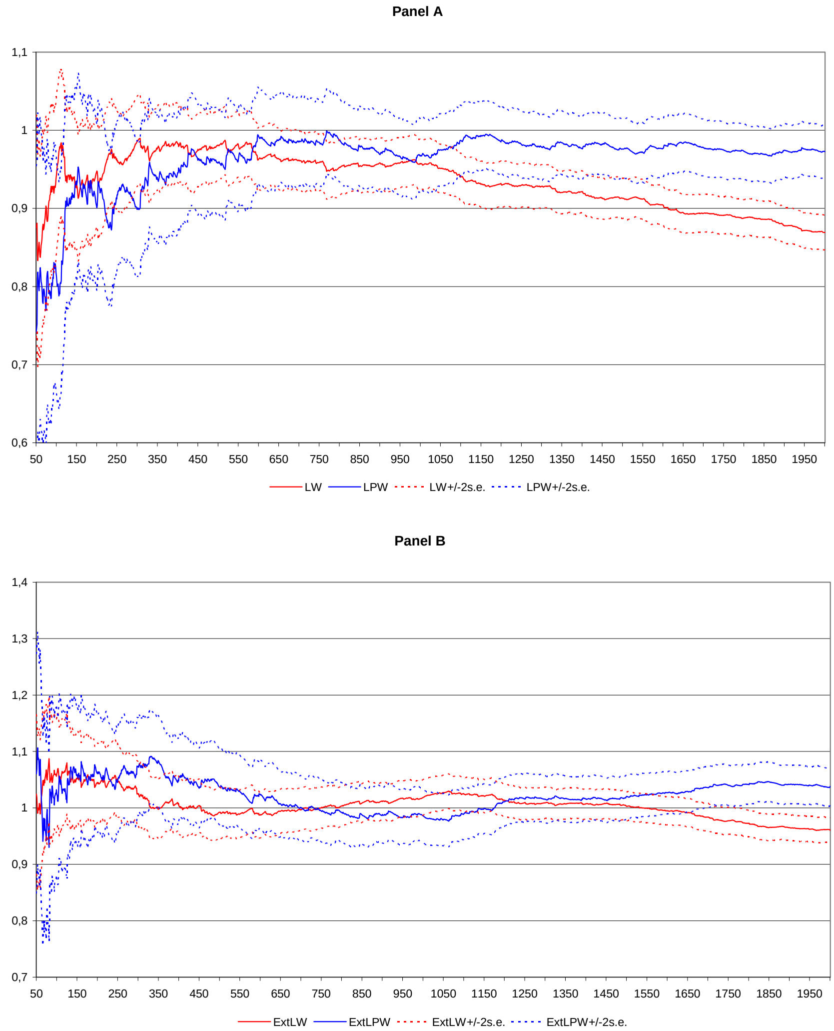


Figure 2: Estimated long memory of log Aaa yield for bandwidth equal to 50 through 2000.

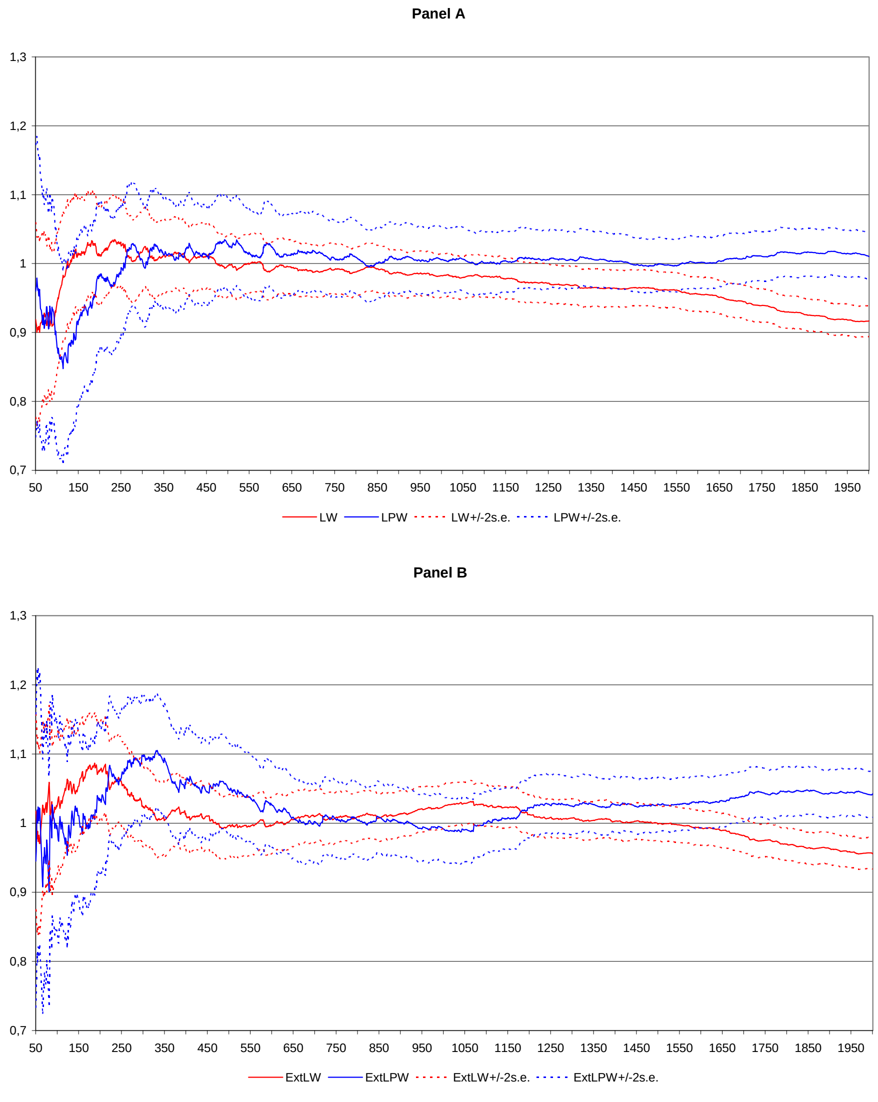


Figure 3: Estimated long memory of log Baa yield for bandwidth equal to 50 through 2000.

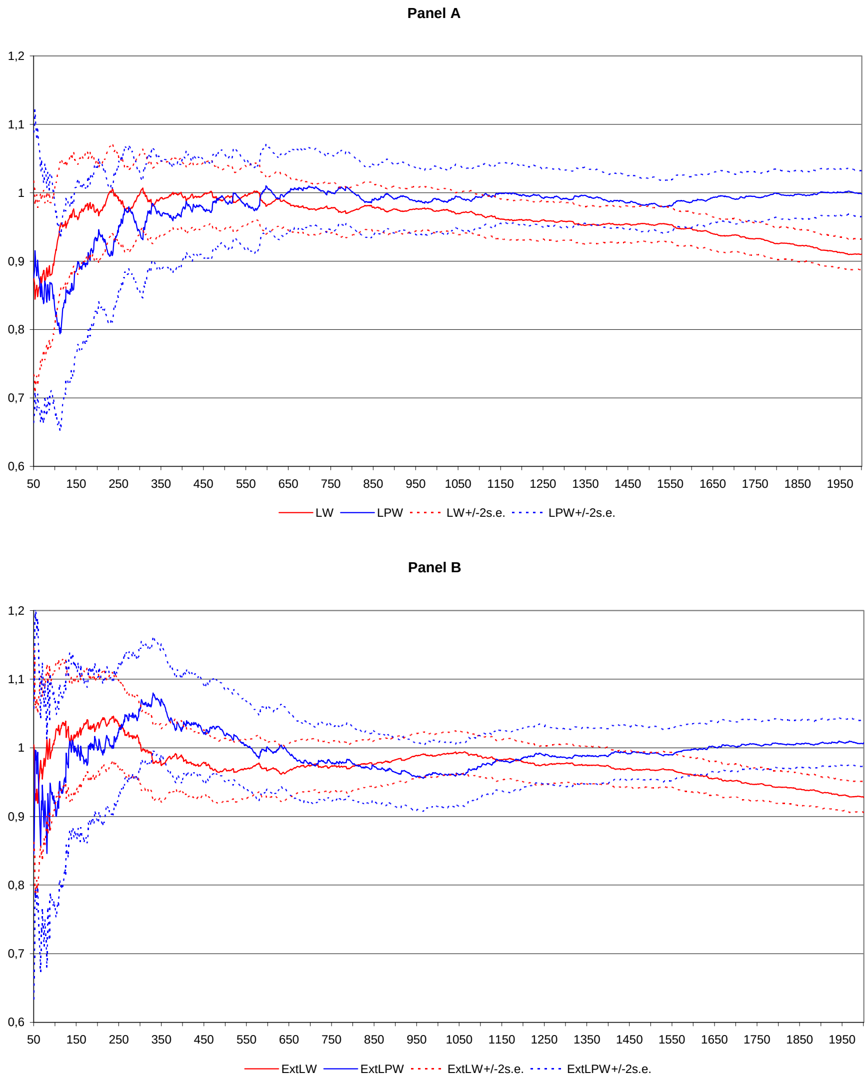


Figure 4: Estimated long memory of log Treasury yield for bandwidth equal to 50 through 2000.

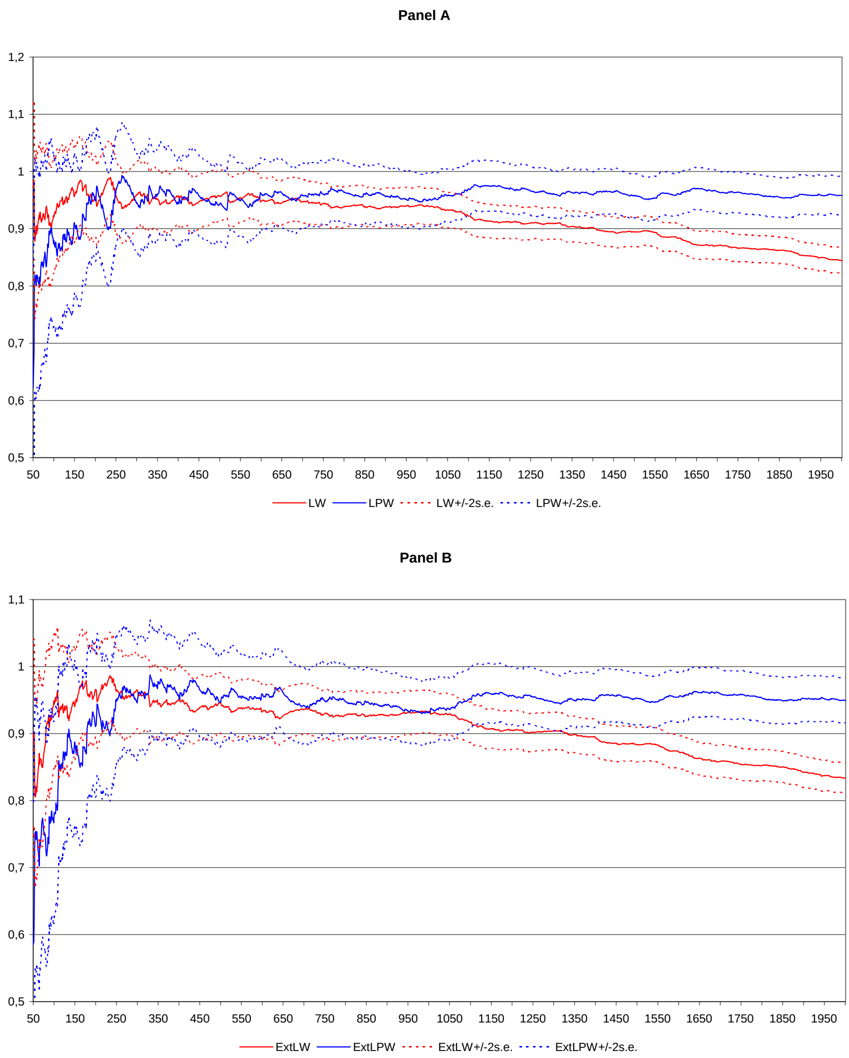


Figure 5: Estimated long memory of Aaa spread over Treasury yield for bandwidth equal to 50 through 2000.

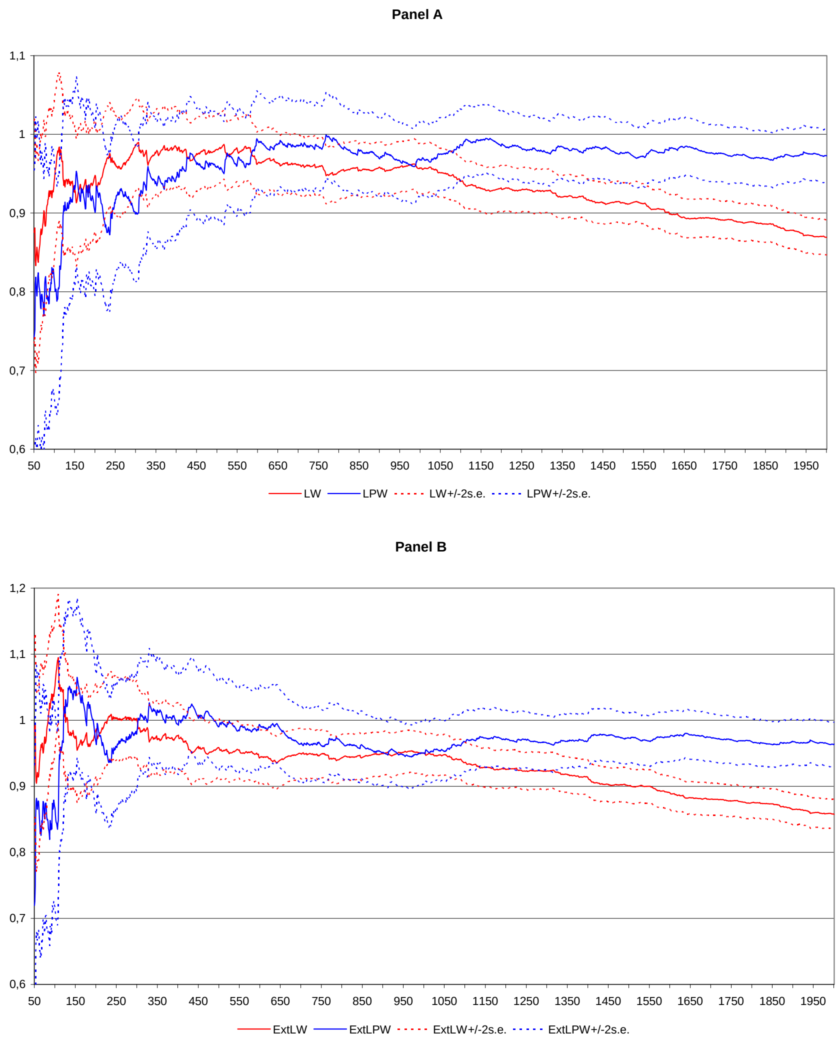
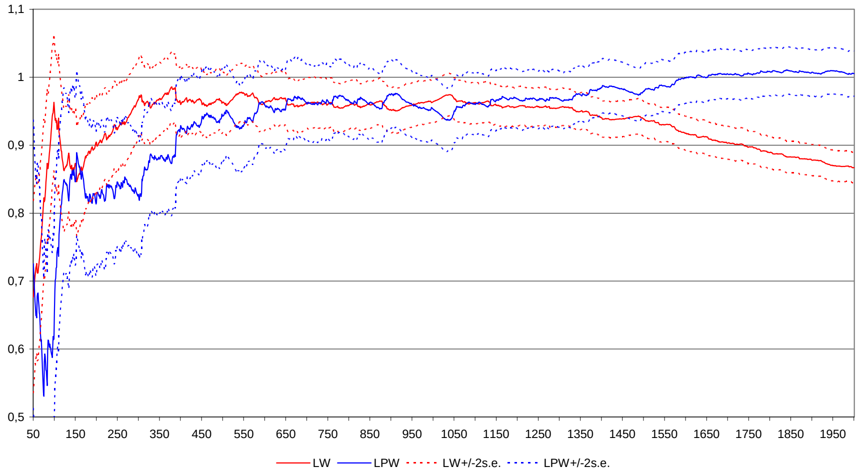


Figure 6: Estimated long memory of Baa spread over Treasury yield for bandwidth equal to 50 through 2000.

Panel A



Panel B

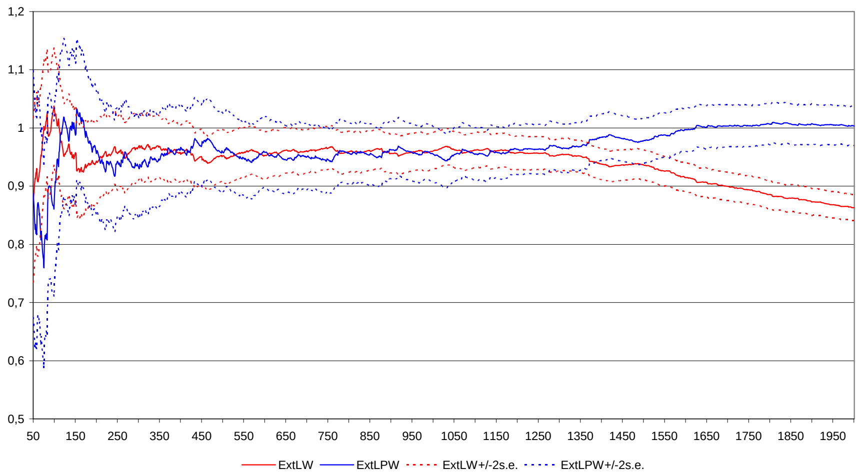


Figure 7: Estimated long memory of Baa spread over Aaa yield for bandwidth equal to 50 through 2000.

Research Papers 2008

Center for Research in Econometric Analysis of Time Series

2008-15	Jie Zhu: Testing for Expected Return and Market Price of Risk in Chinese A-B Share Market: A Geometric Brownian Motion and Multivariate GARCH Model Approach
2008-16:	Jie Zhu: FIEGARCH-M and and International Crises: A Cross-Country Analysis
2008-17:	Almut E. D. Veraart: Inference for the jump part of quadratic variation of Itô semimartingales
2008-18:	Michael Sørensen: Parametric inference for discretely sampled stochastic differential equations
2008-19:	Anne Péguin-Feissolle, Birgit Strikholm and Timo Teräsvirta: Testing the Granger noncausality hypothesis in stationary nonlinear models of unknown functional form
2008-20:	Stefan Holst Bache, Christian M. Dahl and Johannes Tang Kristensen: Determinants of Birthweight Outcomes: Quantile Regressions Based on Panel Data
2008-21:	Ole E. Barndorff-Nielsen, José Manuel Corcuera, Mark Podolskij and Jeannette H.C. Woerner: Bipower variation for Gaussian processes with stationary increments
2008-22:	Mark Podolskij and Daniel Ziggel: A Range-Based Test for the Parametric Form of the Volatility in Diffusion Models
2008-23:	Silja Kinnebrock and Mark Podolskij: An Econometric Analysis of Modulated Realised Covariance, Regression and Correlation in Noisy Diffusion Models
2008-24:	Matias D. Cattaneo, Richard K. Crump and Michael Jansson: Small Bandwidth Asymptotics for Density-Weighted Average Derivatives
2008-25:	Mark Podolskij and Mathias Vetter: Bipower-type estimation in a noisy diffusion setting
2008-26:	Martin Møller Andreasen: Ensuring the Validity of the Micro Foundation in DSGE Models
2008-27:	Tom Engsted and Thomas Q. Pedersen: Return predictability and intertemporal asset allocation: Evidence from a bias-adjusted VAR model
2008-28:	Frank S. Nielsen: Local polynomial Whittle estimation covering non-stationary fractional processes