CREATES

Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2008-59

Testing for long memory in potentially nonstationary perturbed fractional processes

Per Frederiksen and Frank S. Nielsen

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

Aarhus School of Business
University of Aarhus

Handelshøjskolen
Aarhus Universitet

UNIVERSITY OF COPENHAGEN

Testing for long memory in potentially nonstationary perturbed fractional processes*

Per Frederiksen
Equity Trading & Derivatives
Nordea Markets

Frank S. Nielsen†
University of Aarhus and CREATES

November 17, 2008

Abstract

In this paper, we propose new tests for long memory in stationary and nonstationary time series possibly perturbed by short-run noise which may be serially correlated. The tests are all based on semiparametric estimators and exploit the self-similarity property of long memory processes. We offer simulation results that show good size properties of the tests, with power against spurious long memory. An empirical study of daily log-squared returns series of exchange rates and DJIA30 stocks shows that indeed there is long memory in exchange rate volatility and stock return volatility.

JEL Classifications: C14, C22, C43.

Keywords: Temporal aggregation, semiparametric estimation, fractional integration, self-similarity, perturbed fractional processes.

*We are grateful to Niels Haldrup, Esben Høg, and Morten Ø. Nielsen for valuable suggestions and comments. We would also like to thank Birgitte Højklint Nielsen for proofreading the manuscript. This work was partly done while P. Frederiksen was visiting Northwestern University; their hospitality is gratefully acknowledged. F.S. Nielsen is grateful for financial support from the Danish Social Sciences Research Council (grant no. FSE275-05-0199) and Center for Research in Econometric Analysis of Time Series (CREATES), funded by the Danish National Research Foundation.

†Please address correspondence to: Frank S. Nielsen, School of Economics and Management, University of Aarhus, Universitetsparken building 1322, 8000 Aarhus, Denmark; phone: +45 8942 5419; e-mail: fnielsen@creates.au.dk

1 Introduction

Recent interest in the properties of financial markets has supported the notion that volatility has long memory, is fractionally integrated[1] or self-similar, see e.g. Calvet, Fisher & Mandelbrot (1997), Fisher, Calvet & Mandelbrot (1997) and Mandelbrot, Fisher & Calvet (1997). The three definitions are not exactly equivalent. Nonetheless, if a process has long memory, is fractionally integrated or self-similar, its autocovariance function decays hyperbolically as $\gamma(h) \sim c_h |h|^{2d-1}$ as $h \to \infty$, or equivalently its spectral density satisfies $f(\lambda) \sim c_\lambda |\lambda|^{-2d}$ as $\lambda \to 0$, where c_h and c_λ are constant proportionality factors, and d is the long memory parameter, parameter of fractional integration or parameter of self-similarity, see Beran (1994) and Taqqu (2003).

Neglecting a normalizing constant, a definition of self-similarity is that nonoverlapping temporal aggregates of the original series, see eqn. (14) below, are distributed as the original series, and hence, their autocorrelation functions or spectral densities are asymptotically equal. Two of the first self-similar and continuous-time fractionally integrated processes introduced into statistics were the fractional Brownian motion (fBm) and its increments, the fractional Gaussian noise (fGn), see Mandelbrot & van Ness (1968). Later, Granger & Joyeux (1980) and Hosking (1981) introduced an extension of the fBm in discrete-time and named it the autoregressive fractionally integrated moving average (ARFIMA) process, which is often referred to in the literature as a long memory process. More recently, temporal aggregation for long memory processes has been considered. Beran & Ocker (2000) and Man & Tiao (2001, 2006) consider Gaussian ARFIMA models, and Souza (2003*a*, 2003*b*, 2004) present more general results all showing that temporal aggregation only acts to modify the short memory properties, and therefore, a true long memory processes is defined as a process that has the same memory at all levels of sampling frequency.

This phenomenon inspired Ohanissian, Russell & Tsay (2008) to formalize a test for self-similarity and naming it more popularly a test for long memory. In the following, we will use the terms self-similar, fractionally integrated or long memory process liberally, covering the same property that the spectral density of the underlying process is only assumed to decay according to a power law near frequency zero, see eqn. (1) below.

Even though the past two decades have witnessed an increasing interest in fractionally integrated processes as a convenient way of describing the long memory properties of many time series, a debate has begun concerning the possibility of confusing long memory processes and processes which are capable of exhibiting a hyperbolic decaying autocovariance function without being self-similar. We refer to these processes as being spurious long memory processes. It has been known for some years that stationary short memory processes contaminated by level shifts, Markov switching models with independent, identically distributed (iid) regimes or generalized autoregressive conditional heteroskedasticity (GARCH) regimes, white noise models with deterministic trends and higher order factor generalized autoregressive moving average (GARMA) models are capable of exhibiting spurious long memory, see e.g. Bhattacharya, Gupta & Waymire (1983), Woodward, Cheng & Gray (1998), Diebold & Inoue (2001), Granger

[1] See among others Ding, Granger & Engle (1993), Baillie, Bollerslev & Mikkelsen (1996), Comte & Renault (1998), Ray & Tsay (2000), Andersen, Bollerslev, Diebold & Ebens (2001), Andersen, Bollerslev, Diebold & Labys (2001, 2003), Wright (2002), Hurvich & Ray (2003) and Arteche (2004*b*).

& Hyung (2004), and Perron & Qu (2004). Since inference under a stationary true long memory model is different from the spurious case, it is very important to be able to distinguish between them. Ohanissian, Russell & Tsay (2003) have shown in a simulation study that misspecifying the long memory property of asset return volatility results in a serious mispricing of call options. This result is an example of the well-known issue that long-lived impacts of shocks found in long memory models have very different impacts on asset pricing compared to short memory models.

Exploiting the long memory parameter's invariance to temporal aggregation, Ohanissian et al. (2008) derived the joint distributional properties of the log-periodogram regression (LPR) estimator of Geweke & Porter-Hudak (1983) and Robinson (1995*b*) applied to different aggregation levels of the original series, and introduced a formal test of true long memory. A simulation study showed that the size of the test and the power against the spurious long memory models mentioned above are very good.[2]

However, as mentioned by Ohanissian et al. (2008), the methodology is mostly directed towards high frequency data because of the long time series needed to facilitate aggregation, but many financial time series with much shorter length also exhibit long memory, see e.g. the surveys by Robinson (1994*b*), Baillie (1996) and Henry & Zaffaroni (2003) and the references therein for some examples. Hence, to be general, the methodology should be applicable to smaller samples. Furthermore, Ohanissian et al. (2008) extend their test to long memory signal plus added noise models (perturbed fractional processes). However, especially in small samples, it should be noted that the LPR estimator may be substantially downward biased when the noise-to-signal ratio is high, and that the bias increases as the bandwidth m, used in the estimator, increases, see Deo & Hurvich (2001, 2002). Therefore, especially in finite samples, the LPR estimates of the long memory in the temporally aggregated series and their variances might be significantly different depending on the properties of the noise. The issue about the variance is to some extent dealt with in Ohanissian et al. (2008) since they use the variance approximation recommended by Geweke & Porter-Hudak (1983) in their test.

In this paper, we extend the idea of Ohanissian et al. (2008) in several important directions. We first extend the test based on the LPR estimator to nonstationary Gaussian time series and then, by applying tapering, we avoid the assumption of Gaussianity. This does, however, increase the asymptotic variance in the distributional result. Secondly, we derive the local Whittle (LW) of Künsch (1987) and Robinson (1995*a*) analogue of the test in Ohanissian et al. (2008). Basing the test on the local Whittle estimator suggested by Künsch (1987) has the advantage that Gaussianity is not assumed and that theoretically it should have better power since the variance is smaller. Thirdly, we further derive a test that is robust to cases where the signal process (potentially serially correlated) is a long memory process with memory parameter d which is perturbed by an additive noise term which may be serially correlated. More specifically, we use the local polynomial Whittle with noise[3] (LPWN) framework set forward

[2] The power against stationary short memory processes contaminated by level shifts can also indirectly be induced from the paper by Perron & Qu (2004) who showed that applying the log-periodogram estimator on such processes results in very different (spurious) long memory estimates depending on the bandwidths.

[3] Includes both the LW estimator and the local Whittle with noise (LWN) estimator of Hurvich & Ray (2003) and Hurvich, Moulines & Soulier (2005, parameterization (P1)) as special cases.

by Frederiksen, Nielsen & Nielsen (2008). Frederiksen et al. (2008) proposed to approximate both the spectrum of the short-memory component of the signal and the spectrum of the perturbation by polynomials of even and finite order in a shrinking neighborhood of the zero frequency, instead of constants, thereby obtaining a reduction depending on the smoothness of the polynomials. By approximating the short-memory component by a polynomial, Andrews & Sun (2004) showed for non-perturbed processes that it is possible to reduce the theoretical bias of the classical LW estimator if the spectral density is sufficiently smooth. Frederiksen et al. (2008) extended this to include potentially perturbed processes by modeling the spectrum of the perturbation in the same manor, see Frederiksen et al. (2008) for further details.

Although the popular LPR estimator and the LW estimator both preserve consistency and asymptotic normality when applied to perturbed fractional processes, as shown by Deo & Hurvich (2001) and Arteche (2004*b*), these estimators can be severely biased since they do not take the perturbation into account. Indeed, for non-perturbed processes the bias of the standard semiparametric frequency domain estimators is of order $O(\lambda_m^2)$, where $\lambda_m = 2\pi m/n$ and n is the sample size, and m is a user chosen bandwidth number, which tends to infinity slower than n such that $\lambda_m \to 0$, whereas the leading bias term when there is perturbation is of order $O(\lambda_m^{2d})$. As shown in Deo & Hurvich (2001) and Arteche (2004*b*), this bias is typically negative and can be very large (note that $d < 1$). Therefore, estimating long memory in perturbed time series can be a challenging task and calls for an estimator which explicitly accounts for the perturbation.

To alleviate the problem of the sometimes quite large difference between the asymptotic variance and the sample variance, see e.g. Hurvich & Ray (2003) and Ohanissian et al. (2008), we use finite sample approximations of the asymptotic variances of the estimators in the implementation of the tests. Finally, we show that all the applied tests inherit the structure of the covariance matrix derived in Ohanissian et al. (2008).

Simulations show that all the tests have good size properties and power against spurious long memory alternatives. In conclusion, an empirical study of the 30 Dow Jones Industrial Average (DJIA) stocks and the DEM/USD, YEN/USD, and USD/GBP exchange rates are implemented. The empirical analysis shows that there is indeed long memory in volatility.

The remainder of the paper is organized as follows. Section 2 gives an introduction to semiparametric estimation in fractional processes potentially perturbed by some noise term that may be serially correlated. Section 3 introduces the setup of temporal aggregation and some auxiliary results. The limiting distribution results for the semiparametric estimators of long memory under temporal aggregation are shown in section 4. In section 5, we analyze the tests with respect to finite sample size and power. Finally, section 6 provides an empirical study of daily log-squared returns series of DJIA stocks and the DEM/USD, YEN/USD, and USD/GBP exchange rates. Finally, section 7 offers some concluding remarks. All proofs are in Appendix A, and Appendix B contains the specification of the power simulation study.

2 Semiparametric estimation of perturbed fractional processes

For semiparametric frequency domain estimators, it is common to use the approximation

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

where G is a constant, and the symbol "$\sim$" means that the ratio of the left and right hand sides tends to one in the limit. Thus, the semiparametric estimators enjoy robustness to short-run dynamics since they use only information from the periodogram ordinates in the vicinity of the origin. Note that the approximation in (1) is valid for a broad range of processes, where e.g. the ARFIMA process is a special case, since it only assumes that the spectral density decays according to a power law near frequency zero.

Probably the most commonly applied semiparametric estimator is the LPR estimator introduced by Geweke & Porter-Hudak (1983) and analyzed in detail by Robinson (1995*b*). Taking logs in (1) and inserting sample quantities, we get the approximate regression relationship

$$\log I_z(\lambda_j) = \log G - 2d \log \lambda_j + \text{error}, \tag{2}$$

where $\lambda_j = 2\pi j/n$ are the Fourier frequencies, and $I_z(\lambda) = |w_z(\lambda)|^2$ is the periodogram of z_t, where $w_z(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} z_t e^{it\lambda}$ is the discrete Fourier transform of z_t. The LPR estimator is defined as the OLS estimator in the regression (2) using $j = 1, ..., m$, where m is a user chosen bandwidth number which tends to infinity as $n \to \infty$ but at a slower rate. Note that the estimator is invariant to a non-zero mean since $j = 0$ is left out of the regression.

Under suitable regularity conditions, including z_t being Gaussian (later relaxed by Velasco (2000)) and a restriction on the bandwidth, Robinson (1995*b*) derived the asymptotically normal limit distribution for the LPR estimator

$$\sqrt{m}\left(\hat{d}_{LPR} - d\right) \xrightarrow{d} N\left(0, \pi^2/24\right), \tag{3}$$

when d is in the stationary and invertible range $(-1/2, 1/2)$. Recently, Kim & Phillips (2006) and Velasco (1999*b*) demonstrated that the range of consistency is $d \in (-1/2, 1]$ and the range of asymptotic normality is $d \in (-1/2, 3/4)$, where the distributional result in (3) still holds if sufficient trimming of the very first periodogram ordinates is applied, see Velasco (1999*b*, Theorem 3) for details.

By assuming that z_t is a fourth order stationary linear process and applying tapering, Velasco (2000) avoided the Gaussianity assumption. The tapering asymptotically removes the bias of the periodogram ordinates introduced by not assuming Gaussianity and thus paves the way for asymptotic normality results. Defining the tapered periodogram,

$$I_z^T(\lambda_j) = \left|w_z^T(\lambda_j)\right|^2, \tag{4}$$

where $w_z^T(\lambda_j) = (2\pi \sum_{t=1}^{n} h_t^2)^{-1} \sum_{t=1}^{n} h_t z_t \exp(i\lambda_j t)$ and then interchanging $I_z(\lambda_j)$ with $I_z^T(\lambda_j)$ in (2), the tapered LPR estimator (TLPR) is defined as the OLS estimator using $j = 3, 6, \ldots, m$, where m is divisible by 3. If no tapering is used ($h_t = 1$) and the summation is $j = 1, 2, \ldots, m$, then the TLPR reduces to the classical (non-tapered) LPR estimator.

Compared to the LPR estimator, the advantage of the taper is that Gaussianity of the underlying series is not assumed. However, the asymptotic variance is increased by a factor 3 caused by basing the estimator on only a third of the Fourier frequencies, and from Velasco (2000) we know that

$$m^{1/2}(\hat{d}_{TLPR} - d) \xrightarrow{d} N(0, 3\pi^2/24). \tag{5}$$

Another popular semiparametric estimator which does not assume Gaussianity of z_t is the local Whittle (LW) approach suggested by Künsch (1987) and later analyzed in detail by Robinson (1995*a*). The LW estimator is defined as the minimizer of the (negative local Whittle likelihood) function

$$Q(G,d) = \frac{1}{m}\sum_{j=1}^{m}\left(\log\left(G\lambda_j^{-2d}\right) + \frac{I_z(\lambda_j)}{G\lambda_j^{-2d}}\right), \tag{6}$$

i.e.

$$\hat{d}_{LW} = \arg\min_{d\in D}\left(\log\hat{G}(d) - 2d\frac{1}{m}\sum_{j=1}^{m}\log\lambda_j\right), \quad \hat{G}(d) = \frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2d}I_z(\lambda_j). \tag{7}$$

For $D \subset (-1/2, 1/2)$, Robinson (1995*a*) showed that

$$m^{1/2}(\hat{d}_{LW} - d) \xrightarrow{d} N(0, 1/4). \tag{8}$$

Like the LPR estimator, the range of consistency and the range of the above asymptotic normality result have later been shown by Velasco (1999*a*) and Phillips & Shimotsu (2004) to be $d \in (-1/2, 1]$ and $d \in (-1/2, 3/4)$, respectively.

To reduce the asymptotic bias of the standard LW estimator, Andrews & Sun (2004) have suggested the local polynomial Whittle (LPW) estimator, where the constant, $\log G$, in (6) is replaced by the polynomial $\log G - \sum_{r=1}^{R}\theta_r\lambda_j^{2r}$. That is, by modeling the logarithm of the spectral density of the short-run dynamics in the vicinity of the origin by a polynomial instead of a constant, we reduce potential bias introduced by short-run contamination of the signal. As shown by Andrews & Sun (2004), this method does, however, increase the asymptotic variance of $\hat{d}$ in (8) by a multiplicative constant.

For non-perturbed fractional processes, the asymptotic bias of $\hat{d}_{LW}$, $\hat{d}_{LPR}$, and $\hat{d}_{TLPR}$ is of order $O(\lambda_m^2)$, and for $\hat{d}_{LPW}$ it is of order $O(\lambda_m^{\min\{s,2+2R\}})$, where s is a measure of the smoothness of the spectral density near frequency zero.

In this paper, we will focus on perturbed fractional processes

$$z_t = y_t + w_t, \tag{9}$$

where the signal process y_t is a long memory process with memory parameter d, which is perturbed by the additive noise term w_t. These processes have found extensive use in modeling long memory characteristics of observed time series. In particular, they are a version of the random walk plus noise model, except that the signal here is a long memory process rather than a random walk, see e.g. Harvey (1989). Since the analyzed estimators are functions of the periodogram at nonzero frequencies only, we assume without loss of generality[4] that the signal process y_t has zero mean.

Assuming that the processes $\{y_t\}$ and $\{w_t\}$ are independent, the spectral density of z_t can be written as

$$f_z(\lambda) = \lambda^{-2d}\phi_y(\lambda) + \phi_w(\lambda) = \lambda^{-2d}G\left(\frac{\phi_y(\lambda)}{\phi_y(0)} + \lambda^{2d}\frac{\phi_w(\lambda)}{\phi_y(0)}\right), \tag{10}$$

where $\lambda^{-2d}\phi_y(\lambda)$ is the spectrum of the signal y_t, $\phi_w(\lambda)$ is the spectrum of the noise term w_t, and d is the degree of long memory in y_t or equivalently in z_t. Contrary to the case of non-perturbed processes, applying the above-mentioned estimators for perturbed fractional processes, the bias is of order $O(\lambda_m^{2d})$, and, as shown by e.g. Deo & Hurvich (2001), Hurvich & Ray (2003) and Arteche (2004*b*), this bias is typically negative and can be very severe.

More specifically, for perturbed fractional processes we have the spectral representation (10) rather than (1). There are two main consequences: first, the extra additive term in (10) needs to be taken into account to avoid serious asymptotic bias as mentioned above, and second, the rate of convergence of the estimators is reduced if the extra term is not modeled. The latter follows because the choice of bandwidth parameter is severely constrained for perturbed fractional processes when the perturbation term in (10) is not modeled. Thus, for non-perturbed processes the bandwidth requirement is typically $m = o(n^{4/5})$, whereas (apart from logarithmic terms) for perturbed processes it is $m = o(n^{2d/(1+2d)})$. Since $d \leq 1$ and the estimator is $\sqrt{m}-$consistent, this is a serious constraint.

Frederiksen et al. (2008) therefore propose to approximate (10) locally near the zero frequency by[5]

$$g(\lambda) = \lambda^{-2d}G\left(1 + h_y(\theta_y, \lambda) + \lambda^{2d}h_w(\theta_w, \lambda)\right), \tag{11}$$

where $h_y(\theta_y, \lambda) = \sum_{r=1}^{R_y}\theta_{y,r}\lambda^{2r}$, $h_w(\theta_w, \lambda) = \sum_{r=0}^{R_w}\theta_{w,r}\lambda^{2r}$. If $R_y = 0$, we set $h_y(\theta_y, \lambda) = 0$. Defining also the polynomial $h(d, \theta, \lambda) = h_y(\theta_y, \lambda) + \lambda^{2d}h_w(\theta_w, \lambda)$ with $\theta = (\theta_y', \theta_w')'$, this yields the (concentrated) likelihood

$$Q(d, \theta) = \log\hat{G}(d, \theta) + \frac{1}{m}\sum_{j=1}^{m}\log\left(\lambda_j^{-2d}\left(1 + h(d, \theta, \lambda_j)\right)\right), \tag{12}$$

$$\hat{G}(d, \theta) = \frac{1}{m}\sum_{j=1}^{m}\frac{\lambda_j^{2d}I_z(\lambda_j)}{1 + h(d, \theta, \lambda_j)}, \tag{13}$$

with estimates,

$$(\hat{d}, \hat{\theta}) = \underset{(d,\theta)\in D\times\Theta}{\arg\min}\, Q(d, \theta),$$

where Θ is a compact and convex set in $\mathbb{R}^{R+1}$, $R = R_y + R_w$, and $D = [d_1, d_2]$ with $0 < d_1 < d_2 < 1$. Frederiksen et al. (2008) call this estimator the local polynomial Whittle with noise (LPWN) estimator.

[4] In the nonstationary case, the zero mean assumption implies that z_t is free of linear trends, which does entail a loss of generality in that case.

[5] Note that $\phi_y(\lambda)$ and $\phi_w(\lambda)$ are symmetric around $\lambda = 0$ and are therefore approximated by even polynomials.

Note that $h(\theta, \lambda) = 0$ is the standard local Whittle specification in (6), which does not explicitly account for the perturbation. For $R_y = R_w = 0$, we get $h(\theta, \lambda) = \theta$, where $\phi_y(\lambda)$ and $\phi_w(\lambda)$ in (10) are both modeled locally by constants. This is the local Whittle with noise (LWN) estimator of Hurvich & Ray (2003) and Hurvich et al. (2005, parameterization (P1)). Thus, the above model parameterization includes the standard LW estimator and the LWN estimator as special cases.

3 Temporal aggregation and the asymptotic properties

The framework for the test for spurious long memory builds on the self-similarity property of fractionally integrated processes, see Mandelbrot & van Ness (1968). If the time series z_t has long memory, then the memory will be the same at all levels of sampling frequency, and thus the long memory parameter should be invariant to temporal aggregation of the original series, see e.g. Souza (2003*a*, 2003*b*, 2004).

Defining the k period nonoverlapping temporal aggregates of z_t as

$$\tau_{z,t}^{(k)} = \sum_{i=1}^{k} z_{k(t-1)+i}, \; t = 1, \ldots, \frac{n}{k}, \tag{14}$$

$\tau_{z,t}^{(k)}$ and z_t should have the same memory. Note that the number of periodogram ordinates, $m^{(k)} = \left[(n/k)^{\delta} \right]$, used in the estimation of the memory parameter d depends on the length of the series, and since temporal aggregation decreases the length, we have that $m^{(k_u)} > m^{(k_v)}$ for $k_u < k_v$. Furthermore, changing the length of the series also implies different frequencies, λ_j, used in the estimation of d; i.e. $\lambda_{j,k_u} = \frac{2\pi k_u}{n} j$ such that $\lambda_{j,k_u} = \frac{k_u}{k_v} \lambda_{j,k_v}$.

The self-similarity property implies that there is a straightforward relationship between the periodogram of the original series, $I_z(\lambda_j)$, and the periodogram of the k period nonoverlapping temporal aggregates, $I_{\tau_z}^{(k)}(\lambda_j)$. Defining

$$\begin{aligned} I_{\tau_z}^{(k)}(\lambda_j) &= \frac{k}{2\pi n} \left| \sum_{t=1}^{n/k} \left(\sum_{i=1}^{k} z_{k(t-1)+i} \right) \exp(i\lambda_j t) \right|^2 \\ &= \frac{k}{2\pi n} \left| \sum_{t=1}^{n} z_t \exp(im\lambda_j \lfloor t/k \rfloor) \right|^2 \end{aligned}$$

for $1 \leq j \leq n/k$, where $\lfloor x \rfloor$ denotes the smallest integer greater than or equal to x, we state the following result.

Lemma 1 *Let z_t be a long memory process with spectral density satisfying the assumptions outlined in the subsections regarding the individual semiparametric estimators, then for the k'th aggregation level*

$$I_{\tau_z}^{(k)}(\lambda_j) - kI_z(\lambda_j) = O_p\left(\frac{j}{n^{\kappa}}\right),$$

where $\kappa = 1 - 2d$ for $d \in (-1/2, 1/2)$ and $\kappa = 2 - 2d$ for $d \in [1/2, 1)$.

Notice that Lemma 1 extends the result of Ohanissian et al. (2008) to the nonstationary case, and is used in establishing the limiting distributional results for arbitrary linear combinations of LPR and TLPR estimates of long memory obtained using temporally aggregated series, see Theorem 1 and Theorem 2 below. In order to prove similar results for the LPWN estimators, we need the following lemma.

Lemma 2 *Under the assumptions of Lemma 1, we have that*

$$f_{\tau_z}^{(k)}(\lambda_j) - k f_z(\lambda_j) = O_p\left(\frac{j}{n^{\kappa}}\right) \quad \textit{and} \quad \frac{I_{\tau_z}^{(k)}(\lambda_j)}{f_{\tau_z}^{(k)}(\lambda_j)} - \frac{I_z(\lambda_j)}{f_z(\lambda_j)} = O_p\left(\frac{j}{n^{\kappa}}\right).$$

This furthermore implies that for any two arbitrarily chosen aggregation levels, $k_u < k_v$ we have

$$I_{\tau_z}^{(k_v)}(\lambda_j) - \frac{k_v}{k_u} I_{\tau_z}^{(k_u)}(\lambda_j) = O_p\left(\frac{j}{n^{\kappa}}\right) \quad \textit{and} \quad \frac{I_{\tau_z}^{(k_v)}(\lambda_j)}{f_{\tau_z}^{(k_v)}(\lambda_j)} - \frac{I_{\tau_z}^{(k_u)}(\lambda_j)}{f_{\tau_z}^{(k_u)}(\lambda_j)} = O_p\left(\frac{j}{n^{\kappa}}\right),$$

where $\kappa = 1 - 2d$ for $d \in (-1/2, 1/2)$ and $\kappa = 2 - 2d$ for $d \in [1/2, 1)$.

Now to use the fact that the memory of the original series and the k period nonoverlapping temporal aggregates are equal, and thus to derive a formal test, it is necessary to know the joint distribution of the long memory estimates for different aggregation levels. Depending on the estimator, applying results for empirical spectral processes (see Soulier (2001), Moulines & Soulier (2003) or Hurvich et al. (2005)), it can be shown that arbitrary linear combinations of the memory estimates of the aggregated series are asymptotically normal. Ohanissian et al. (2008) apply the central limit theorem for functions of Gaussian vectors in Soulier (2001). An analogue for non-Gaussian vectors is found in Moulines & Soulier (2003, Theorem 21), which can be used when basing the test on the tapered LPR estimator. Alternatively, Hurvich et al. (2005, Proposition 4.1, Proposition 4.2, and Proposition A.2) can be used in the log-periodogram case and has the advantage that neither Gaussianity nor stationarity of z_t is required. It is straightforward to show that if the assumptions of Soulier (2001, Theorem 4.1) or Moulines & Soulier (2003, Theorem 21) are satisfied, then applying the LPR estimators (see e.g. the proof of Theorem 1 and Theorem 2 in the appendix) implies that the assumptions of Hurvich et al. (2005, Proposition A.2) are satisfied. This has the advantage that our results are valid for $d \in (0, 3/4)$. For the LPWN estimator, the limiting results can be based on Hurvich et al. (2005, Proposition 4.1, Proposition 4.2, and Proposition A.2), or one can straightforwardly extend Lemma 3 of Frederiksen et al. (2008) to this setup utilizing our Lemma 2 and the assumptions underlying the specific estimator.

Next we will distinguish between log-periodogram estimation and local Whittle estimation. We will only discuss the modeling of approximating the spectrum of the signal and perturbation by polynomials in the local Whittle setting, but it should be noted that the ideas of Frederiksen et al. (2008) could be transferred to the log-periodogram setting by utilizing the results of Andrews & Guggenberger (2003), Sun & Phillips (2003), and Arteche (2006), where Andrews & Guggenberger (2003) model the LPR equivalent of the LPW estimator and Sun & Phillips (2003) and Arteche (2006) discuss the LPR equivalent of the LWN estimator.

For the local Whittle framework, we will focus on the LPWN methodology as it encompasses the LW and LWN estimators as special cases. We do this to simplify the presentation, and we only list one set of assumptions even though these could be relaxed when looking at the special cases (e.g. LW or LWN).

In the following, true values of the parameters are denoted by subscript zero, and $\lfloor x \rfloor$ denotes the integer part of a real number x. We also define a function $\phi(\lambda)$ to be smooth of order s at $\lambda = 0$ if, in a neighborhood of $\lambda = 0$, $\phi(\lambda)$ is $\lfloor s \rfloor$ times continuously differentiable with $\lfloor s \rfloor$ −derivative, $\phi^{(\lfloor s \rfloor)}$, satisfying $|\phi^{(\lfloor s \rfloor)}(\lambda) - \phi^{(\lfloor s \rfloor)}(0)| \le C |\lambda|^{s - \lfloor s \rfloor}$ for some constant $C < \infty$.

3.1 Log-periodogram regression asymptotics

To extend the setup of Ohanissian et al. (2008) to the nonstationary case, we need the following assumptions of Velasco (1999*b*).

A1 z_t has memory parameter $d_0 \in (1/2, 3/2)$ if the mean zero stationary process $\varepsilon_t = \Delta z_t$ has spectral density

$$f_\varepsilon(\lambda) = \lambda^{-2(d_0-1)} f^*(\lambda),$$

where $f^*(\lambda)$ is a positive, integrable, even function on $[-\pi, \pi]$ which is bounded above and away from zero and is continuous at $\lambda = 0$.

Assumption A1 allows us to define the pseudo spectral density

$$f_z(\lambda) = \lambda^{-2d_0} f^*(\lambda).$$

Since $f_z(\lambda)$ does not have a clear statistical interpretation, Velasco (1999*b*) stated his assumptions in terms of $f_\varepsilon(\lambda)$.

A2 The spectral density $f_\varepsilon(\lambda)$ satisfies for $\rho \in (0, 2]$, $G \in (0, \infty)$, and $d_0 \in [1/2, 3/2)$

$$f_\varepsilon(\lambda) = G\lambda^{-2(d_0-1)} + O(\lambda^{-2(d_0-1)+\rho}) \quad \text{as } \lambda \to 0^+.$$

Notice that $\rho = 2$ if ε_t is a stationary and invertible ARFIMA process.

A3 In a neighborhood $(0, \epsilon)$ of the origin, $f_\varepsilon(\lambda)$ is differentiable and

$$\left| \frac{d}{d\lambda} f_\varepsilon(\lambda) \right| = O(\lambda^{-1-2(d-1)}) \quad \text{as } \lambda \to 0^+.$$

For the stationary case, these assumptions are equivalent to Assumptions 1 and 2 in Robinson (1995*b*) and thus imply that $f^*(\lambda)$ is bounded from above and away from zero and is continuous in an interval $(0, \epsilon)$, $\epsilon > 0$, and that $f_z(\lambda)$ has bounded first derivative. Lastly, we need to restrict the bandwidth number.

A4 For any fixed aggregation level k, the bandwidth $m^{(k)} = m^{(k)}(n)$ is a non-decreasing sequence such that

$$\frac{\left(m^{(k)}\right)^{3/2}\log m^{(k)}}{(n/k)^{\kappa}} + \frac{\left(m^{(k)}\right)^{1/2}\log m^{(k)}}{l^{2(1-d)}} + \frac{l\log^2(n/k)}{m^{(k)}} + \frac{\left(m^{(k)}\right)^{1+1/2\rho}}{(n/k)} \to 0, \tag{15}$$

where κ is defined in Lemma 1 and l is a user-chosen trimming number to avoid the very first periodogram ordinates.

The first term of the (15) is an extra term compared to e.g. Velasco (1999*b*). It is needed to ensure that we can derive a scaling factor α, which is needed in deriving the limiting distributional results for arbitrary linear combinations of the estimator. We can now state the following limiting distributional result for arbitrary linear combinations of LPR estimates of long memory obtained using the K temporal aggregated series.

Theorem 1 *Let the assumptions of Lemma 1 and Assumptions A1 - A4 be satisfied and assume that ε_t is Gaussian. Then for d_0 in the interior of $D = [d_1, d_2]$, $0 < d_1 < d_2 < 3/4$,*

$$\alpha \sum_{i=1}^{K} \gamma_i(\hat{d}_i - d_0) \xrightarrow{d} N\left(0, \pi^2/24\right),$$

where $\alpha = 2\left(\sum_{i=1}^{K}\frac{\gamma_i^2}{m^{(k_i)}} + 2\sum_{i=2}^{K}\sum_{s=1}^{i-1}\frac{\gamma_s\gamma_i}{m^{(k_s)}}\right)^{-1/2}$, and the γ_i's are scalar weights that can be chosen arbitrarily.

The result of Theorem 1 is very convenient since confidence bands of the long memory estimates of e.g. volatility do generally suggest that nonstationarity is a possibility.

To avoid the assumption of Gaussianity, we need the following assumptions of Velasco (1999*b*, 2000) besides A1 and A3.

B1 ε_t is covariance stationary with $d_0 \in [0, 1/2)$ and satisfies

$$\varepsilon_t = \sum_{j=0}^{\infty}\alpha_j\eta_{t-j}, \quad \sum_{j=0}^{\infty}\alpha_j < \infty, \quad \left|\frac{d}{d\lambda}\alpha(\lambda)\right| = O(\lambda^{-1}\left|\alpha(\lambda)\right|), \quad \alpha(\lambda) = \sum_{j=0}^{\infty}\alpha_j e^{ij\lambda},$$

as $\lambda \to 0^+$, where η_t is *iid* with $E(\eta_t) = 0$, $E\left(\eta_t^2\right) = 1$, $E\left(\eta_t^3\middle|\mathcal{F}_{t-1}\right) = \mu_3 < \infty$ and $E\left(\eta_t^4\middle|\mathcal{F}_{t-1}\right) = \mu_4 < \infty$, where $\mathcal{F}_{t-1}$ is the σ-field generated by $\{\eta_s : s \le t-1\}$.

B2 η_t has characteristic function $Q(\varpi) = E\left(e^{i\varpi\eta_t}\right)$ satisfying

$$\sup_{|\varpi|\ge\varpi_0}|\mathcal{Q}(\varpi)| = \delta(\varpi_0), \quad \text{for all } \varpi_0 > 0, \quad \text{and} \int_{-\infty}^{\infty}|\mathcal{Q}(\varpi)|^p\,d\varpi < \infty,$$

for some integer $p > 1$.

As explained by Velasco (2000), Assumption B2 implies that the probability distribution of η_t has a bounded and continuous density.

B3 The spectral density $f_\varepsilon(\lambda)$ satisfies for $\rho \in (0, 2]$, $G \in (0, \infty)$, $E_\rho < \infty$, and $d_0 \in [1/2, 3/2)$

$$f_\varepsilon(\lambda) = G\lambda^{-2(d_0-1)} + GE_\rho\lambda^{-2(d_0-1)+\rho} + o(\lambda^{-2(d_0-1)+\rho}) \quad \text{as } \lambda \to 0^+.$$

Notice that this assumption is satisfied if ε_t e.g. is a stationary and invertible ARFIMA process ($\rho = 2$). However, for the stationary case these assumptions are equivalent to Assumptions 1 and 2 in Robinson (1995*b*), and thus imply that $f^*(\lambda)$ is bounded from above and away from zero and is continuous in an interval $(0, \epsilon)$, $\epsilon > 0$, and that $f_z(\lambda)$ has bounded first derivative. Lastly, we need to restrict the bandwidth number.

B4 For any fixed aggregation level k, the bandwidth $m^{(k)} = m^{(k)}(n)$ is a non-decreasing sequence such that for $d_0 \in [0, 1/2)$

$$\frac{\left(m^{(k)}\right)^{3/2} \log m^{(k)}}{(n/k)^\kappa} + \frac{1}{l} + \frac{l \log^2(n/k)}{m^{(k)}} + \frac{\left(m^{(k)}\right)^{1+1/2\rho}}{(n/k)} \to 0,$$

and for $d_0 \in [1/2, 3/2)$ we also need $\left(m^{(k)}\right)^{1/2}/l \to 0$ for $n \to \infty$, where κ is defined in Lemma 1, and l is a user-chosen trimming number to avoid the very first periodogram ordinates.

Note that for $d \in [0, 1/2)$, interchanging $d-1$ in Assumption B3 with just d, Velasco (2000) proved the normality result of the tapered LPR estimator based on the covariance stationary series ε_t. However, the tapered estimator of Velasco (1999*b*) is still based on Gaussianity of ε_t. Thus, the following result is based on the conjecture of Velasco (1999*b*, p. 351) that instead of Gaussianity, Assumption B1 and B2 can be applied in the nonstationary case as well.

We can now state the following limiting distributional result for arbitrary linear combinations of TLPR estimates of long memory obtained using the K temporal aggregated series.

Theorem 2 *Let the assumptions of Lemma 1 and Assumptions A1, A3, B1 - B4 be satisfied. Then for d_0 in the interior of $D = [d_1, d_2]$, $0 < d_1 < d_2 < 3/4$,*

$$\alpha \sum_{i=1}^{K} \gamma_i(\hat{d}_i - d_0) \xrightarrow{d} N\left(0, 3\pi^2/24\right),$$

where $\alpha = 2\left(\sum_{i=1}^{K} \frac{\gamma_i^2}{m^{(k_i)}} + 2\sum_{i=2}^{K}\sum_{s=1}^{i-1} \frac{\gamma_s\gamma_i}{m^{(k_s)}}\right)^{-1/2}$, and the γ_i's are scalar weights that can be chosen arbitrarily.

If the signal process $\{y_t\}$ is contaminated by a noise term $\{w_t\}$, we need to restrict the bandwidth choice of the LPR and TLPR for the limiting distributional results to hold. That is, under the assumptions of Theorem 1 or Theorem 2 and $m^{(k)} = o\left((n/k)^{\frac{4d}{4d+1}}\right)$, Theorem 1 and Theorem 2 still hold for perturbed fractional processes, see e.g. Deo & Hurvich (2001).

Note that the asymptotic variance in the limiting distributional result for both the non-tapered and tapered log-periodogram regression equals the (scaled) asymptotic variance of each individual estimate. This implies that the asymptotic covariance matrix is known given the user-chosen bandwidth. Theorem 1 and 2 give the following structure of the covariance matrix.

Proposition 2.1 *Under the assumptions of Theorem 1 or Theorem 2 and with d_0 in the interior of $D = [d_1, d_2]$, $0 < d_1 < d_2 < 3/4$,*

$$\lim_{n\to\infty} \left(m^{(k_i)} \left(Cov\left(\hat{d}_i, \hat{d}_j\right) - Var\left(\hat{d}_i\right)\right)\right) = 0 \tag{16}$$

for $1 \leq i < j \leq K$.

Notice that the structure of the covariance matrix parallels the result of Hausman (1978) that an efficient estimator must have zero asymptotic covariance with the difference between the efficient estimator and any other consistent, asymptotically normal estimator. In this setup the estimate based on the least aggregated series is efficient in the class of LPR estimates obtained using temporal aggregates.

Since it was shown by Ohanissian et al. (2008) that the test for long memory based on the LPR estimator is better behaved if one uses a finite sample analogue of (16), we present this in Proposition 2.2. This was originally put forth by Geweke & Porter-Hudak (1983).

Proposition 2.2 *Treating $I_z(\lambda_j)$ as independently distributed as $\frac{1}{2}\lambda_j^{-2d_0} G\chi_2^2$, omitting a multiplying constant that converges to one in probability, we obtain the finite approximate variance expression*

$$Var(\hat{d}_i) \approx \pi^2 \left[24 \sum_{j}^{m^{(k_i)}} \left(\log\lambda_j - \frac{1}{m^{(k_i)}} \sum_{q}^{m^{(k_i)}} \log\lambda_q \right)^2 \right]^{-1},$$

where we note that the summation is $j, q = 1, 2, ..., m^{(k_i)}$ and $j, q = 3, 6, ..., m^{(k_i)}$ for the LPR and TLPR, respectively.

Note that basing the variance expression on $j, q = 3, 6, \ldots, m^{(k_i)}$ makes it approximately three times larger compared to the LPR estimator.

3.2 Local Whittle asymptotics

These assumptions follow from Frederiksen et al. (2008) adapted to our setting.

C1 The noise process $\{w_t\}$ is independent of the signal process $\{y_t\}$.

C2 The spectral density of z_t is $f_z(\lambda) = \lambda^{-2d_0} G_0 \frac{\phi_y(\lambda)}{\phi_y(0)} + \phi_w(\lambda)$, where $\phi_y(\lambda)$ and $\phi_w(\lambda)$ are real, even, positive, continuous functions on $[-\pi, \pi)$ and $d_0 \in D = [d_1, d_2]$ with $0 < d_1 < d_2 < 1$.

C3 The functions $\phi_y(\lambda)$ and $\phi_w(\lambda)$ are smooth of orders s_y and s_w at $\lambda = 0$, where $s_y > 2R_y$, $s_w > 2R_w$, and $s_y, s_w \geq 1$.

Assumption C1 is the independence assumption used above to write the spectral density of z_t as the sum of the (pseudo) spectral densities of y_t and w_t. Assumption C3 is a smoothness condition on the functions $\phi_y(\lambda)$ and $\phi_w(\lambda)$ similar to that applied by Andrews & Sun (2004). Note that Assumption C3 holds for all $s_y < \infty$ when, e.g., y_t is a finite order ARFIMA process, and for all $s_w < \infty$ when, e.g., w_t is a finite order ARMA process.

C4 (a) The signal y_t has zero mean and admits an infinite order moving average representation $y_t = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}$ (stationary case) or $\Delta y_t = x_t = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}$ (nonstationary case), where $\sum_{j=0}^{\infty} \alpha_j^2 < \infty$ and ε_t satisfies, for all t, $E\left(\varepsilon_t | \mathcal{F}_{t-1}\right) = 0$, $E\left(\varepsilon_t^2 \middle| \mathcal{F}_{t-1}\right) = 1$, $E\left(\varepsilon_t^3 \middle| \mathcal{F}_{t-1}\right) = \mu_3 < \infty$, and $E\left(\varepsilon_t^4 \middle| \mathcal{F}_{t-1}\right) = \mu_4 < \infty$ almost surely, where F_{t-1} is the σ-field generated by $\{\varepsilon_s, s < t\}$.

(b) There exists a random variable ε with $E(\varepsilon^2) < \infty$ such that for all $\tau > 0$ and some $K > 0$, $P(|\varepsilon_t| > \tau) < KP(|\varepsilon| > \tau)$.

(c) In a neighborhood of the origin, $\frac{\partial}{\partial \lambda} \alpha(\lambda) = O(|\alpha(\lambda)|/\lambda)$ as $\lambda \to 0$, where $\alpha(\lambda) = \sum_{k=0}^{\infty} \alpha_k e^{ik\lambda}$.

C5 (a) The noise w_t has zero mean and admits an infinite order moving average representation $w_t = \sum_{j=0}^{\infty} \beta_j \eta_{t-j}$, where $\sum_{j=0}^{\infty} \beta_j^2 < \infty$ and η_t satisfies, for all t, $E\left(\eta_t | \mathcal{F}_{t-1}\right) = 0$, $E\left(\eta_t^2 \middle| \mathcal{F}_{t-1}\right) = 1$, $E\left(\eta_t^3 \middle| \mathcal{F}_{t-1}\right) = \mu_3 < \infty$, and $E\left(\eta_t^4 \middle| \mathcal{F}_{t-1}\right) = \mu_4 < \infty$ almost surely, where F_{t-1} is the σ-field generated by $\{\eta_s, s < t\}$.

(b) There exists a random variable ε with $E(\varepsilon^2) < \infty$ such that for all $\tau > 0$ and some $K > 0$, $P(|\eta_t| > \tau) < KP(|\varepsilon| > \tau)$.

(c) In a neighborhood of the origin, $\frac{\partial}{\partial \lambda} \beta(\lambda) = O(|\beta(\lambda)|/\lambda)$ as $\lambda \to 0$, where $\beta(\lambda) = \sum_{k=0}^{\infty} \beta_k e^{ik\lambda}$.

Importantly, Assumptions C4 and C5 allow for non-Gaussian processes. Note that Assumptions C1-C4 plus the assumption that w_t is white noise with finite fourth moment imply the assumptions needed on y_t and w_t to prove consistency and asymptotic normality (if, in addition, $d_2 < 3/4$) of the LWN estimator of Hurvich & Ray (2003). It follows from Theorem 3 below that their results for the LWN estimator are also valid for our more general assumptions on w_t in Assumption C5.

C6 For any fixed aggregation level k, the bandwidth $m^{(k)} = m^{(k)}(n)$ is a non-decreasing sequence such that

$$\frac{\left(m^{(k)}\right)^{1+4R_y}}{(n/k)^{4R_y}} + \frac{\left(m^{(k)}\right)^{1+4(d_0+R_w)}}{(n/k)^{4(d_0+R_w)}} \to \infty, \tag{17}$$

$$\frac{\left(m^{(k)}\right)^{3/2} \log m^{(k)}}{(n/k)^{\kappa}} + \frac{\left(m^{(k)}\right)^{2\varphi_y+1}}{(n/k)^{2\varphi_y}} + \frac{\left(m^{(k)}\right)^{2\varphi_w+4d_0+1}}{(n/k)^{2\varphi_w+4d_0}} \to 0 \tag{18}$$

where κ is defined as in Lemma 2 and $\varphi_a = \min\{s_a, 2 + 2R_a\}, a = y, w$.

Condition (17) guarantees that all the elements of the scaling matrix $B_{m^{(k)}}$ (that is used to normalize the gradient and Hessian of the scaled log-likelihood) diverge as $n/k \to \infty$, which is a minimal condition for consistency. (18) restricts the expansion rate of the bandwidth to control bias and ensures that the estimator uses only relevant information from periodogram ordinates sufficiently near the zero frequency.

C7 Θ is a compact and convex subset of $\mathbb{R}^{R+1}$, and θ_0 lies in the interior of Θ.

Frederiksen et al. (2008) show that the LPWN estimator is consistent for $d \in (0,1)$ for $k=1$ when the bandwidth is chosen as $\frac{1}{m^{(1)}} + \frac{m^{(1)}}{n} \to 0$.

Theorem 3 *Let Assumptions C1-C7 hold with d_0 in the interior of $D = [d_1, d_2]$, $0 < d_1 < d_2 < 3/4$, then $\hat{d}$ and $\hat{\theta}$ are both consistent and*

$$\sum_{i=1}^{K} \gamma_i \widetilde{B}_{m^{(k_i)}} \begin{pmatrix} \hat{d}_i - d_0 \\ \hat{\theta}_i - \theta_0 \end{pmatrix} \xrightarrow{d} N(0, \left(\alpha \Omega_{R_y,R_w}\right)^{-1}), \quad \Omega_{R_y,R_w} = \begin{pmatrix} 4 & \mu'_{R_y} & \nu'_{R_w} \\ \mu_{R_y} & \Gamma_{R_y} & \psi'_{R_y,R_w} \\ \nu_{R_w} & \psi_{R_w,R_y} & \Psi_{R_w} \end{pmatrix},$$

and

$$\alpha = \left(\sum_{i=1}^{K} \frac{\gamma_i^2}{m^{(k_i)}} + 2 \sum_{i=2}^{K} \sum_{s=1}^{i-1} \frac{\gamma_s \gamma_i}{m^{(k_s)}} \right)^{-1/2},$$

where the γ_i's are scalar weights that can be chosen arbitrarily, $\widetilde{B}_{m^{(k_i)}} = 1/\sqrt{m^{(k_i)}} B_{m^{(k_i)}}$ and $B_{m^{(k_i)}} = B_{m^{(k_i)}}(d_0)$ is the $(R+2) \times (R+2)$ deterministic diagonal matrix with diagonal elements

$$\left(B_{m^{(k_i)}}\right)_{11} = \sqrt{m^{(k_i)}}, \; \left(B_{m^{(k_i)}}\right)_{s+1,s+1} = \sqrt{m^{(k_i)}} \lambda_{m^{(k_i)}}^{2s} \text{ for } s = 1, \ldots, R_y,$$
$$\text{and } \left(B_{m^{(k_i)}}\right)_{s+R_y+2, s+R_y+2} = \sqrt{m} \lambda_{m^{(k_i)}}^{2d_0+2s} \text{ for } s = 0, \ldots, R_w,$$

μ_{R_y} and $\nu_{R_w} = \nu_{R_w}(d_0)$ are the vectors

$$(\mu_{R_y})_s = \frac{-4s}{(1+2s)^2} \text{ for } s = 1, \ldots, R_y \text{ and } (\nu_{R_w})_{s+1} = \frac{-4(d_0+s)}{(1+2d_0+2s)^2} \text{ for } s = 0, \ldots, R_w,$$

Γ_{R_y} and $\Psi_{R_w} = \Psi_{R_w}(d_0)$ are the $R_y \times R_y$ and $(R_w+1) \times (R_w+1)$ matrices

$$\left(\Gamma_{R_y}\right)_{ls} = \frac{4ls}{(1+2l+2s)(1+2l)(1+2s)} \text{ for } l, s = 1, \ldots, R_y,$$
$$(\Psi_{R_w})_{l+1,s+1} = \frac{4(d_0+l)(d_0+s)}{(1+2l+2s+4d_0)(1+2l+2d_0)(1+2s+2d_0)} \text{ for } l, s = 0, \ldots, R_w,$$

and $\psi_{R_w,R_y} = \psi_{R_w,R_y}(d_0)$ is the $(R_w+1) \times R_y$ matrix

$$(\psi_{R_w,R_y})_{l+1,s} = \frac{4s(d_0+l)}{(1+2d_0+2s+2l)(1+2d_0+2l)(1+2s)} \text{ for } l = 0, \ldots R_w, s = 1, \ldots, R_y.$$

If $R_y = R_w = 0$ define $\Omega_{0,0} = \begin{pmatrix} 4 & \nu'_0 \\ \nu_0 & \Psi_0 \end{pmatrix}$.

First of all, we note that by setting $h(d,\theta,\lambda)=0$ or $R_y=R_w=0$, we obtain as special cases the LW and LWN estimators, respectively

$$\alpha\sum_{i=1}^{K}\gamma_i\left(\hat{d}_i^{(LW)}-d_0\right)\xrightarrow{d}N(0,1/4),$$
$$\alpha\sum_{i=1}^{K}\gamma_i\left(\hat{d}_i^{(LWN)}-d_0\right)\xrightarrow{d}N\left(0,\frac{(1+2d_0)^2}{16d_0^2}\right).$$

Secondly, we note that the asymptotic variance of $\alpha\sum_{i=1}^{K}\gamma_i(\hat{d}-d_0)$ is free of the polynomial parameters θ_0, but it depends on d_0. Moreover, the use of the polynomials $h_y(\theta_y,\lambda)$ and $h_w(\theta_w,\lambda)$ increases the asymptotic variance of $\hat{d}$ by a multiplicative constant compared to LWN, seen by use of the formula for the inverse of a partitioned matrix.

Note that the condition (18) implies that if $\phi_y(\lambda)$ and $\phi_w(\lambda)$ are infinitely smooth near frequency zero, then any (R_y,R_w) can be chosen and the estimator is $(n/k)^{1/2-\tau}$ consistent for all $\tau>0$. Hence, in that case, the rate of convergence is arbitrarily close to the parametric rate. Thus, the conditions (17) and (18) allow the bandwidth $m^{(k)}$ to be much larger than for the LWN estimator and the standard LW estimator, which require that (assuming $s_y\geq 2, s_w\geq 2$) $\left(m^{(k)}\right)^5(n/k)^{-4}\to 0$ and $\left(m^{(k)}\right)^{4d_0+1}(n/k)^{-4d_0}\to 0$, respectively, see Hurvich & Ray (2003) and Arteche (2004*b*) in a non-temporal aggregation context. Therefore, Theorem 3 provides an improvement in the rate of convergence relative to existing estimators of the memory parameter for perturbed fractional processes. This comes at the cost of an increase in the asymptotic variance by a multiplicative constant, but this is clearly more than off-set by the faster rate of convergence, at least asymptotically.

The LPWN setup inherits the same approximate variance structure as for the LPR setup.

Proposition 3.1 *Under the assumptions of Theorem 3 and with d_0 in the interior of $D=[d_1,d_2]$, $0<d_1<d_2<3/4$,*

$$\lim_{n\to\infty}\left(m^{(k_i)}\left(Cov\left(\hat{d}_i,\hat{d}_j\right)-Var\left(\hat{d}_i\right)\right)\right)=0 \tag{19}$$

for $1\leq i<j\leq K$.

In Proposition 3.2, we present the analogue for the LPWN estimator to Proposition 2.2.

Proposition 3.2 *Treating $I_z(\lambda_j)$ as independently distributed as $\frac{1}{2}\lambda_j^{-2d_0}G\chi_2^2$, omitting a multiplying constant that converges to one in probability, we obtain that*

$$cov\left[m^{(k)}\nabla Q(d_0,\theta_0)\right]\approx\Gamma,$$

where

$$\begin{aligned}
\Gamma_{11} &= 4\sum_{j=1}^{m^{(k_i)}} v_{1,j}^2, \\
\Gamma_{1,(1+s)} &= 2\sum_{j=1}^{m^{(k_i)}} v_{1,j} v_{(1+s),j}, \\
\Gamma_{(1+s),1} &= \Gamma_{1,(1+s)}, \\
\Gamma_{1,(2+R_y+l)} &= 2\sum_{j=1}^{m^{(k_i)}} v_{1,j} v_{(2+R_y+l),j}, \\
\Gamma_{(2+R_y+l),1} &= \Gamma_{1,(2+R_y+l)}, \\
\Gamma_{(2+R_y+l),(1+s)} &= \sum_{j=1}^{m^{(k_i)}} v_{(2+R_y+l),j} v_{(1+s),j}, \\
\Gamma_{(1+s),(2+R_y+l)} &= \Gamma_{(2+R_y+l),(1+s)} \\
\Gamma_{(1+s),(1+s)} &= \sum_{j=1}^{m^{(k_i)}} v_{(1+s),j}^2, \\
\Gamma_{(2+R_y+l),(2+R_y+l)} &= \sum_{j=1}^{m^{(k_i)}} v_{(2+R_y+l),j}^2,
\end{aligned}$$

where

$$\begin{aligned}
v_{1,j} &= \log\lambda_j - \frac{\lambda_j^{2d_0} h_w(\theta_{0,w},\lambda_j)\log\lambda_j}{1+h(d,\theta_0,\lambda_j)} - \frac{1}{m^{(k_i)}}\sum_{q=1}^{m^{(k_i)}}\left(\log\lambda_q - \frac{\lambda_q^{2d_0} h_w(\theta_{0,w},\lambda)\log\lambda_q}{1+h(d,\theta_0,\lambda_q)}\right), \\
v_{(1+s),j} &= \frac{-\lambda_j^{2s}}{1+h(d,\theta_0,\lambda_j)} - \frac{1}{m^{(k_i)}}\sum_{q=1}^{m^{(k_i)}} \frac{-\lambda_q^{2s}}{1+h(d,\theta_0,\lambda_q)}, \\
v_{(2+R_y+l),j} &= \frac{-\lambda_j^{2d_0+2l}}{1+h(d,\theta_0,\lambda_j)} - \frac{1}{m^{(k_i)}}\sum_{q=1}^{m^{(k_i)}} \frac{-\lambda_q^{2d_0+2l}}{1+h(d,\theta_0,\lambda_q)},
\end{aligned}$$

for $s = 1, ..., R_y$, $l = 0, 1, ..., R_w$, *and* $h(d,\theta,\lambda) = h_y(\theta_y,\lambda) + \lambda^{2d} h_w(\theta_w,\lambda)$. *Then the finite sample approximate variance of* $\hat{d}$ *is given as the element* $\left[\Gamma^{-1}\right]_{11}$, *where we replace the unknown true parameters* (d_0, θ_0) *with the consistent estimate* $\left(\hat{d}_i, \hat{\theta}_i\right)$.

We remark that for the LW and LWN cases, i.e. $h(d,\theta,\lambda) = 0$ and $h(d,\theta,\lambda) = \theta\lambda^{2d}$, respectively, the corresponding entries in the above finite sample covariance matrix are deleted.

The validity of the finite sample approximate variance expression is influenced by the approximation of the spectrum of the signal and noise by polynomials. That is, e.g. in the case of the LWN estimator the expression is more accurate when $\theta\lambda^{2d}$ is small; i.e. when either d is large or θ is small. Therefore, the variance expression can be inaccurate when d is close to zero. However, unreported simulations show that the expression has the advantage that it does not blow up when d is close to zero given realistic noise-to-signal ratios (see Deo & Hurvich (2001)).

In the next section, we setup the test statistic.

4 Testing for long memory

Our null hypothesis is that the original series is a long memory process in which case the memory parameter should be the same across aggregation levels. If we let $\hat{d} \equiv \left(\hat{d}_1, \hat{d}_2, \ldots \hat{d}_K\right)'$ be the vector of memory parameters for the K aggregated series, where $\hat{d}_1$ is the memory estimate of the original series, $\hat{d}_2$ is the memory estimate of $\tau_{z,t}^{(2)} = z_{2t-1} + z_{2t}$ etc. (see eqn. (14)), and let the finite sample covariance matrix of the estimates be Λ, then applying the limiting normal distributional results derived in sections 3.1 and 3.2, the test statistic follows directly from Ohanissian et al. (2008).

Corollary 1 *Let $\hat{d}$ be the K dimensional vector of long memory estimates of the K aggregated series, i.e. $\hat{d} \equiv \left(\hat{d}_1, \hat{d}_2, \ldots \hat{d}_K\right)'$, and let $Var\left(\hat{d}\right) = \Lambda$. Then (a)*

$$W = \left(\hat{d} - d_0\right)' \Lambda^{-1} \left(\hat{d} - d_0\right) \xrightarrow{d} \chi_K^2,$$

(b) since d is unknown, we propose a feasible test statistic given as

$$\hat{W} = \left(T\hat{d}\right)' \left(T\Lambda T'\right)^{-1} \left(T\hat{d}\right) \xrightarrow{d} \chi_{K-1}^2,$$

where we use the mean value of the estimates as an approximation of d, and the $(K-1) \times K$ matrix T is defined with elements $[T]_{kk} = 1 - 1/K$ for $k = 1, \ldots, K-1$ and $-1/K$ elsewhere.

Note that we use the feasible test statistic in Corollary 1 (b) even though it is straightforward to specify the null hypothesis that each long memory estimate should be identical across aggregation levels using Corollary 1 (a); i.e. the Wald test on its standard form. However, since d_0 in practice is unknown, it seems more reasonable to test whether the individual estimates are different from the mean value.

5 Finite sample properties of the tests

The size and power of the different tests are evaluated in a finite sample setup where we use the corresponding 95% χ_{K-1}^2 critical value. In investigating the size, we will limit the analysis to where the signal is an $ARFIMA(p,d,0)$ for $p \in \{0,1\}$ and potentially contaminated by a perturbation. We will for the sake of simplicity not consider the case where there is serial correlation in the perturbation, see e.g. the Monte Carlo (non-temporal aggregation) setup in Frederiksen et al. (2008). That is, we will therefore only implement the LPWN estimator where we set $R_y = 1$ and $R_w = 0$.

In a non-temporal aggregation setup, we know from Hurvich & Ray (2003) that the LWN estimator is superior to the LW estimator in terms of bias and RMSE in the context of the standard LMSV model. Furthermore, Hurvich et al. (2005) show that the polynomial log-periodogram regression estimator of Andrews & Guggenberger (2003) suffers from severe bias in the case of perturbed fractional processes, and the LPW estimator is expected to perform similarly. Therefore, we would expect the LWN and LPWN(1,0) to outperform the LPR, TLPR, LW, and LPW in both size and power, where we contaminate the signal by a noise term. Additionally, Frederiksen et al. (2008) show that the LPWN estimator is superior to the LWN when we have short-run contamination in the signal besides contamination of the signal by an additive noise term.

5.1 Simulation setup

We consider the model

$$z_t = y_t + \mathbb{I} w_t, \tag{20}$$

where $\mathbb{I}$ is an indicator function taking the value 0 if no noise is added. $\{y_t\}$ is the signal process, and $\{w_t\}$ is the perturbation process. We model $\{w_t\}$ as

$$w_t = \log u_t^2, \quad u_t \sim NID(0,1). \tag{21}$$

Note that the variance of w_t is $\sigma_w^2 = \pi^2/2$ regardless of the variance of u_t. The signal process $\{y_t\}$ follows different DGPs. For brevity, we consider four different DGPs for the signal process. The general setup for $\{y_t\}$ is

$$(1 - \alpha_y L)(1 - L)^d y_t = \left(1 + \beta_y\right) \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2), \tag{22}$$

with parameter configurations

$$\begin{aligned} \text{Model I} &: \mathbb{I} = 0, \alpha_y = \beta_y = 0, \\ \text{Model II} &: \mathbb{I} = 1, \alpha_y = \beta_y = 0, \\ \text{Model III} &: \mathbb{I} = 1, \alpha_y = 0.8, \beta_y = 0, \\ \text{Model IV} &: \mathbb{I} = 1, \alpha_y = 0, \beta_y = -0.8, \end{aligned}$$

We remark that in all the models the noise-to-signal ratio is given as[6]

$$nsr = \frac{f_w(0)}{f_{(1-L)^d y_t}(0)} = \frac{\frac{\pi^2}{2}}{\sigma_\eta^2 \frac{\left(1+\beta_y\right)^2}{\left(1-\alpha_y\right)^2}}. \tag{23}$$

[6]This allows us to control the true value of $h(d, \theta)$ (in theory, as $\lambda \to 0$), see eqn (12).

For each Monte Carlo DGP, we generated $10,000$ artificial time series with a sample size of 4096 and 8192 with ordered aggregation levels as consecutive powers of two; i.e. $k_i = 2^{i-1}$ for $i = 1, 2, \ldots, K = 5$.[7] This is in sharp contrast to Ohanissian et al. (2008) who employ $n = 610,304$. However, our choice is motivated by the fact that we want to apply the tests to shorter time series. For both size and power studies we set the bandwidth $m^{(k_i)} = \lfloor (n/k_i)^{\delta} \rfloor$ for all aggregate series, where $\delta \in \{0.5, 0.6, 0.7\}$. For the LPR, TLPR and LW estimators, we conjecture that the "optimal" empirical[8] choice of δ is 0.5, whereas for the LWN and LPWN estimators, we apply the same bandwidth for simple comparison. However, regarding the estimators where we model a polynomial term also, we can have issues of pinning down the θ's estimates. Therefore, a larger sample of periodogram ordinates is needed, and thus the inclusion of $\delta \in \{0.6, 0.7\}$. Furthermore, a higher bandwidth value may lead to lower bias for the estimators that implement a polynomial due to the fact that the polynomial parameters are only consistently estimated if the bandwidth grows sufficiently fast relative to the sample size, see Assumption C6. So in that sense, high bandwidth values generate better estimates of the polynomial parameters leading to lower bias.

The parameter of interest, d, is set equal to either 0.4 or 0.6. For the noise-to-signal ratio, we choose $nsr \in \{1, 5, 10\}$, and the variance σ_η^2 is set as a function of α_y and β_y such that the nsr has the desired value. The values of d, nsr, α_y, β_y, and the sample sizes are chosen to reflect empirical findings on long memory in volatility, see the references in the introduction for some examples.

The signal $\{y_t\}$ is generated by the circulant embedding method as described in Davies & Harte (1987), i.e. the stationary type I fractionally integrated process in the terminology of Marinucci & Robinson (1999), see also Beran (1994, pp. 215-217). To generate nonstationary series with $d \geq 1/2$, we simulate the ARFIMA process with integration order $d - 1$ and cumulate the resulting series. Numerical optimization was carried out in Matlab v7.2 using the BFGS optimization routine. The initial values were set as follows. For the local Whittle type estimators, we used the LPR estimate, $\hat{d}_{LPR}$, if it was in the interior of the admissible space of d, i.e. $[0.01, 0.99]$.[9] Otherwise, d was set equal to 0.1. As initial values for the polynomial parameters, we used 1 for all estimators.

Like Ohanissian et al. (2008), we consider power properties by applying the tests to simulated realizations of models which are capable of exhibiting spurious long memory. More specifically, we consider stationary and nonstationary random level shift models, Markov switching models with *iid* or *GARCH* regimes, and a white noise model with a slow deterministic trend. The exact specifications of the models are in Appendix B.

To conserve space, we present only a subset of the results. The left-out results (Size study: Model IV, $d = 0.6$ (for Model I-III), $n = 8192$, and $m^{(k_i)} = \lfloor (n/k_i)^{0.6} \rfloor$, and for the power study: $m^{(k)} = \lfloor (n/k)^{\delta} \rfloor$ for $\delta \in \{0.6, 0.7\}$) are qualitatively similar when comparing the different

[7] The number of observations is chosen as a power of two in order to use the fast Fourier transform in calculating the periodogram. This speeds up the estimations considerably compared to using the discrete Fourier transform.

[8] It is well-known that the bias of the LPR and the LW estimators is increasing in the bandwidth when the long memory series is perturbed, and that choosing $\delta = 0.5$ renders fairly unbiased results, see e.g. Sun & Phillips (2003) and Arteche (2004*b*, 2004b).

[9] We set the admissible parameter space of the memory parameter equal to the consistency region. Although, we note that the limiting distribution results only hold for $d_0 \in (0, 3/4)$.

estimators to the ones presented, and are available from the authors upon request.

5.2 Simulation results

Before turning to the results, note that a general problem stems from the fact that using temporally aggregated series is asymptotically equivalent to using a smaller bandwidth in estimating the memory of the original series. Since it is well-known that the estimators that do not model the perturbation directly, i.e. LPR, TLPR, LW and LPW, become less biased when applying smaller bandwidths in perturbed fractional processes, see e.g. Deo & Hurvich (2002), Sun & Phillips (2003), Arteche (2004*a*, 2004*b*), Frederiksen & Nielsen (2008), and Frederiksen et al. (2008), we might see a large variability in the estimates across aggregates, and thus, possible overrejection. Tables 1 to 3 show that this is generally the case since the size increases as we include more aggregated series (K increases) in the tests. From a practical point of view, this is expected since the sample of periodogram ordinates decreases as K increases, and it therefore becomes harder for the estimators to pin down the memory parameter precisely, and for the LWP, LWN, and LPWN to pin down the polynomial estimates.

[Tables 1-3 about here]

In the case where there is no contamination by short-run dynamics in the signal and with no perturbation, i.e. Model I with results displayed in Table 1, we see that the size generally increases as K increases. But overall, the size is generally fairly close to the theoretical type I error, $\alpha = 0.05$.

In Table 2, we consider Model II, i.e. a perturbed fractional process without any contamination of the signal. Here we would presume that the tests based on the LWN and LPWN estimators would outperform the tests that do not take the perturbation into account, i.e. the LPR, TLPR, LW, and LPW estimators. We clearly see that the LWN and LPWN estimators introduce less size distortion in the empirically relevant scenario of $d = 0.4$, when $K < 4$. However, it is noteworthy that when $K = 5$, and sometimes when $K = 4$, these tests have larger size compared to the LPR, TLPR, LW and LPW tests when the nsr is low. This is caused by the fact that the impact from the perturbation diminishes when K increases, which benefits the LPR, TLPR, LW, and LPW estimators and because using fewer periodogram ordinates makes $\phi_w(\lambda)$ in eqn. (10) approach zero inducing more volatile estimates of d for the LWN and LPWN estimators.

Results for Model III where $\left(\alpha_y, \beta_y\right) = (0.8, 0)$ are shown in Table 3. We would expect the LPWN estimator to outperform the LWN estimator as it explicitly models the contamination in the signal. This is also seen in Table 3, although when K increases, the size of the two tests are comparable as the bias in estimating d introduced by not modeling the short-run contamination diminishes as K increases. Furthermore, looking at the tests based on the LPR, TLPR, and LW estimators, there is clear overrejection from not properly dealing with the short-run contamination and perturbation.

For the nonstationary case (results omitted), we see that the theoretical extension carries over into practice. The sizes of the tests are equivalent to the stationary case, but it seems that the tests perform slightly better under this scenario. This is due to the stronger memory in the signal, which makes it easier for the estimator to pin down even in smaller samples.

To complete the size analysis, we also tried modeling the signal innovations as t_4; i.e. Student t distributed with four degrees of freedom, and thus introducing fat tails. This did not change the results. This is expected for the test based on the TLPR, LW, LPW, LWN, and LPWN estimators since this does not violate the assumptions behind the estimators. However, it is noteworthy that the test based on the LPR estimator still performs well since this is a violation of one of its assumptions.

[**Tables 4-5 about here**]

Focusing on the power properties of the tests, Tables 4 to 5 show that all the tests generally have power above 90% against the mentioned models capable of exhibiting spurious long memory. However, the test based on the LWN and LPWN estimators always obtains (very close to) 100% power. This result holds for both the generic and the empirical parameter values. Taking a closer look, we have marked in bold font the test which obtains the lowest power for the different aggregation levels (K's) and spurious long memory models. Now it becomes more noticeable that the tests based on the LPR, TLPR, LW, and LPW estimators have the lowest power. This was expected in comparison to the estimator that explicitly models the perturbation in the returns, but as seen the tests based on the LPR and TLPR estimators are comparable to the tests based on the LW. This was not expected since the LW obtains the lowest variance. Furthermore, we see that the overall lowest power results are found for the nonstationary random level shift (NSRLS) model and the white noise with trend (WNT) model when based on the empirical parameter values. It is also seen that using a higher K, i.e. increasing the number of aggregated series used in the test, increases the power. This result is not surprising since the NSRLS and WNT models give the most stable long memory estimates across aggregation levels (not reported here) although the stability decreases as K increases. For the LPR estimator, this is also obvious from Table III in Ohanissian et al. (2008).

Increasing the bandwidth choice (results omitted) generally results in increased power and in almost all cases with close to 100% power for all estimators.

For completeness, we also investigated the power of the tests against the $AR(1)$ alternative with autoregressive parameter $\phi = 0.684$, which was shown by Nielsen (2004) to be the most troublesome for parametric tests of fractional integration. However, unreported results show that the power is above 96% for all estimators.

Overall, we find that the proposed tests all perform well in terms of both size and power. However, in the empirically relevant scenario of perturbed long memory series ($d \in \{0.4, 0.6\}$, $\mathbb{I} = 1$, and $nsr \in \{1, 5, 10\}$), the test based on the LWN and LPWN estimators should be considered as good alternatives to the test based on the LPR estimator derived in Ohanissian et al. (2008), especially for higher bandwidth. Furthermore, when we contaminate the signal by short-run dynamics, i.e. AR or MA noise, the LPWN estimator outperforms the LWN estimator in terms of size. This is of course not surprising as the LPWN estimator models this contamination by a polynomial whereas the LWN does not.

6 Empirical applications

This section investigates whether it is reasonable to assume that there is long memory in daily log-squared returns series of exchange rates (DEM/USD, YEN/USD, and USD/GBP) and DJIA30 Stocks. Based on the above finite sample analysis, we apply the series to the tests based on the LPR, TLPR, LW, LPW, LWN, and LPWN estimators. We implement the LPWN estimator with (R_y, R_w) equal to $(1,0)$ and with starting values etc. as in the Monte Carlo study above.[10] Including the test based on the LWN and LPWN estimators is motivated by the study of daily log-squared returns series of DM/USD and Yen/USD presented in Ohanissian et al. (2008), where rejecting the null hypothesis of long memory based on the LPR estimates comes fairly close for the DM/USD (p-value is 0.08 based on the approximate variance expression).

In both empirical studies, to avoid the problem of taking logarithm of zero, we based the analysis on adjusted log-squared returns using the method of Fuller (1996); i.e. we analyze

$$\log \tilde{r}_t^2 = \log\left(r_t^2 + \nu\right) - \frac{\nu}{r_t^2 + \nu},$$

where $\nu = \frac{0.02}{n}\sum_{t=1}^{n} r_t^2$. Furthermore, we set the bandwidth equal to $m^{(k)} = \left\lfloor (n/k)^{\delta} \right\rfloor$, $\delta \in \{0.5, 0.7\}$.

6.1 Long memory in exchange rate volatility

This subsection analyses empirically the long memory in volatility of daily returns series of DEM/USD, YEN/USD, and USD/GBP exchange rates obtained from the U.S. Federal Reserve Board of Governors H.10 release. The sample covers the period 12/1/1986 –11/30/2006 for a total of $n = 5000$ observations.[11,12] We apply aggregation levels $k_i = 2^{i-1}$ for $i = 1, 2, \ldots, K = 4$ such that the aggregated series have 5000, 2500, 1250 and 625 observations, respectively. The number of aggregation levels implies that the critical value of the test statistic is $\chi_3^2(0.95) = 7.82$.

[Tables 6 and 7 about here]

Tables 6 and 7 report the long memory estimates for the 30 stocks, the test statistics, and the respective p-values.

When we set $m = \left\lfloor (n/k)^{0.5} \right\rfloor$, i.e. Table 6, we reject the null for the DEM/USD for the tests baed on the LPR, TLPR, and LW estimators whereas, we cannot reject for the LPW, LWN, and LPWN estimators. Furthermore, we reject the null for the YEN/USD when using the LPR estimator. We note that for the LWN estimator we hit the lower bound for the DEM/USD and

[10] We also implemented the local polynomial Whittle with noise estimator with (R_y, R_w) equal to $(0,1)$ and $(1,1)$. The results where qualitatively the same as for the LPWN(1,0) parameterization, and therefore omitted for space reasons.

[11] After the adoption of the Euro on January 1, 1999, the DEM/USD exchange rate has been calculated using the USD/EUR exchange rate and the fixed 1.95583 DEM/EUR exchange rate.

[12] The sample is originally $n = 5186$, but to obtain reasonable aggregated series we deleted the first 186 observations.

YEN/USD and for the LPWN estimator for the DEM/USD. Hitting the lower bound inflates the variance and therefore results in a non-rejection of the null. Increasing the bandwidth choice mitigates this problem because we get more periodogram ordinates to pin down the θ's.

We also note that for the estimators that model the potential perturbation in returns, i.e. the LWN and LPWN estimators, the estimate of the memory parameter is somewhat higher than for the estimators that do not model the potential perturbation, i.e. the LPR, TLPR, LW, and LPW estimators. Especially when we increase the bandwidth. This is expected given the theoretical, simulation, and empirical evidence of this in the literature, see e.g. Deo & Hurvich (2002), Sun & Phillips (2003), Arteche (2004*a*, 2004*b*), Frederiksen & Nielsen (2008), Frederiksen et al. (2008), among others. Furthermore, the apparent perturbation in returns causes the test based on the LPR, TLPR, LW, and LPW estimators to reject the null hypothesis of long memory when in fact the test based on the LWN and LPWN estimators does not. Overall, we conclude that there indeed is long memory in exchange rate volatility.

Ohanissian et al. (2008) analyze the DEM/USD and YEN/USD exchange rates using the LPR estimator and they find that there indeed is true long memory in exchange rates, although on another sample period than ours. However, their estimates of the long memory parameter d for $k = 1, 2, 3, 4$ and $m = \lfloor (n/k)^{0.5} \rfloor$ are reasonably close to our LPR estimates.

6.2 Long memory in DJIA stock volatility

This subsection analyzes the long memory in daily log-squared returns series of the 30 DJIA stocks corrected for the effects of stock splits and dividends from January 1 1990 to March 31 2008, for a sample of $n = 4400$,[13] and we apply aggregation levels $k_i = 2^{i-1}$ for $i = 1, 2, \ldots, K = 4$ such that the aggregated series have 4400, 2200, 1100 and 550 observations, respectively. The number of aggregation levels implies that the critical value of the test statistic is $\chi^2_3(0.95) = 7.82$.

[Table 8 about here]

Table 8 reports the long memory estimates for the 30 stocks, the test statistics and the respective p-values. For space reasons, we only display the results for the LPR estimator and LWN estimator. The results for the TLPR, LW, and LPW tests are comparable to the test based on the LPR estimator, and results for the LPWN test are comparable to the test based on the LWN estimator. To make a fair comparison of the estimators, we focus on setting the bandwidth equal to $m^{(k)} = \lfloor (n/k)^{0.5} \rfloor$ as we know from the simulation study that increasing the number of periodogram ordinates potentially introduces severe size distortion for the tests based on the TLPR, LPR, LW, and LPW estimators. Furthermore, the reason for focusing on the LWN estimator instead of the LPWN when $m^{(k)} = \lfloor (n/k)^{0.5} \rfloor$ is that we need more periodogram ordinates to pin down the polynomial coefficient in the case of the LPWN estimator.

Notice that for some of the DJIA stocks, the perturbation in the returns causes the test based on the LPR estimator to reject the null hypothesis of long memory when in fact the test

[13] The sample is originally $n = 4753$, but to obtain reasonable aggregated series we deleted the first 353 observations.

based on the LWN estimator does not. It is only for General Electric (GE) that both tests reject, although removing $k = 4$ from the analysis makes the test based on the LWN estimator not reject. In some cases, we hit the lower bound of 0.01 for the LWN estimator (C, JPM, and MSFT). This results in a variance inflation and hence we cannot reject the null. This is mitigated by increasing the bandwidth and in this case the memory estimate when $k = 4$ is comparable to the other levels of aggregation, and hence we still cannot reject the null.

It is noticeable that the LWN estimates of the memory parameter in the log-squared returns series are somewhat larger than the LPR estimates, especially when increasing the bandwidth as the LPR estimates get downward biased due to the perturbation in returns. From the LWN memory estimates, we see that volatility might be nonstationary. This is in line with the results obtained by Frederiksen et al. (2008) where they in an empirical investigation of the DJIA30 Stocks show that the LWN and LPWN estimators indicate stronger persistence in volatility than standard estimators, and for most stocks produce estimates of d in the nonstationary region. Nonetheless, the results of the tests show that there is indeed long memory in stock return volatility for the majority of the stocks.

7 Concluding remarks

In this paper, we have proposed new tests for long memory in stationary and nonstationary time series possibly perturbed by short-run noise which may be serially correlated. The different tests are all based on semiparametric estimators and exploit the invariance of the long memory parameter to temporal aggregation of the original series. We have shown that the analyzed estimators are asymptotically normal under temporal aggregation and we derived Wald-type tests based on these semiparametric estimators. Simulations showed that the tests overall have good size properties and power against series that exhibit spurious long memory. More specifically, in the case where the signal is perturbed by a noise term, the tests based on the local Whittle with noise and local polynomial Whittle with noise estimators outperform the standard semiparametric estimators (e.g. log-periodogram regression and standard local Whittle estimators). Furthermore, the local polynomial Whittle with noise estimator outperforms the local Whittle with noise when we, besides the perturbation, introduce short-run contamination in the signal. This is due to the fact that the local polynomial Whittle with noise estimator allows the spectrum of the signal (and potentially also the perturbation) to be modeled as even finite polynomials, instead of constants near the zero frequency, and thereby yielding a bias reduction depending on the smoothness of the spectra. In conclusion, the tests were applied to the daily log-squared returns series of exchange rates (DEM/USD, YEN/USD, and USD/GBP) and DJIA30 Stocks, which showed that there is long memory in exchange rate and stock return volatility.

Future research will focus on deriving methods for selecting the optimal level of temporal aggregation, and how to select the bandwidth given temporal aggregation. Additionally, it would also be of interest to look at bootstrap methods to improve the size properties. Furthermore, we want to apply the theory to high frequency data where especially the test based on the LPWN estimator will be appropriate.

8 Appendix A: Proofs

The following proofs are heavily based on the methodology of Hurvich et al. (2005), Frederiksen et al. (2008), and the ideas of Ohanissian et al. (2008). In the following theorems, we set $k_1 = 1$ without loss of generality. However, if $k_1 \neq 1$, the results of Lemma 2 can be used to prove the theorems.

8.1 Proof of Lemma 1

Here we focus on the nonstationary case since the proof for $0 \leq d < \frac{1}{2}$ is found in Ohanissian et al. (2008).

Define $R_{j,k} = \left(I_{\tau_z}^{(k)}(\lambda_j) - k I_{\tau_z}^{(1)}(\lambda_j) \right)$, where $I_{\tau_z}^{(k)}(\lambda_j)$ is the periodogram of the kth aggregated series. Now, we show that this remainder term can be bounded. Since $R_{j,k}$ can be written as a quadratic form of a mean zero Gaussian vector, we only need to show that its first two moments can be bounded, for details see Ohanissian et al. (2008). Using the results in Velasco (1999*b*, Proof of Theorem 1) combined with the approach in Ohanissian et al. (2008), we can write, for $\frac{1}{2} \leq d < 1$

$$E[R_{j,k}] = E\left(I_{\tau_z}^{(k)}(\lambda_j) - k I_{\tau_z}^{(1)}(\lambda_j) \right)$$

$$= \frac{k}{2\pi n} \sum_{t_1=1}^{n} \sum_{h_1=1}^{t_1} \sum_{t_2=1}^{n} \sum_{h_2=1}^{t_2} E(\varepsilon_{h_1}\varepsilon_{h_2}) \left(\exp\left(i\lambda_j k \left(\left\lfloor \frac{t_1}{k} \right\rfloor - \left\lfloor \frac{t_2}{k} \right\rfloor \right) \right) - \exp\left(i\lambda_j (t_1 - t_2) \right) \right),$$

where $\lfloor x \rfloor$ denotes the smallest integer greater than or equal to x. This expression can be written as

$$\begin{aligned}
&\frac{k}{2\pi n} \sum_{t_1=1}^{n} \sum_{h_1=1}^{t_1} \sum_{t_2=1}^{n} \sum_{h_2=1}^{t_2} E(\varepsilon_{h_1}\varepsilon_{h_2}) \left(\cos\left(\lambda_j k \left(\left\lfloor \frac{t_1}{k} \right\rfloor - \left\lfloor \frac{t_2}{k} \right\rfloor \right) \right) - \cos\left(\lambda_j (t_1 - t_2) \right) \right) \\
= &\sum_{t_1=1}^{n} \sum_{h_1=1}^{t_1} \sum_{t_2=1}^{n} \sum_{h_2=1}^{t_2} E(\varepsilon_{h_1}\varepsilon_{h_2}) \frac{k}{2\pi n} \left(\cos\left(\lambda_j k \left(\left\lfloor \frac{t_1}{k} \right\rfloor - \left\lfloor \frac{t_2}{k} \right\rfloor \right) \right) - \cos\left(\lambda_j (t_1 - t_2) \right) \right) \\
= &a(\varepsilon_{h_1}\varepsilon_{h_2})\, b(t_1, t_2) = a(\varepsilon_{h_1}\varepsilon_{h_2})\, O\left(\frac{j}{n^2} \right),
\end{aligned}$$

where

$$\begin{aligned}
a(\varepsilon_{h_1}\varepsilon_{h_2}) &= \sum_{t_1=1}^{n} \sum_{h_1=1}^{t_1} \sum_{t_2=1}^{n} \sum_{h_2=1}^{t_2} E(\varepsilon_{h_1}\varepsilon_{h_2}) \\
b(t_1, t_2) &= \frac{k}{2\pi n} \left(\cos\left(\lambda_j k \left(\left\lfloor \frac{t_1}{k} \right\rfloor - \left\lfloor \frac{t_2}{k} \right\rfloor \right) \right) - \cos\left(\lambda_j (t_1 - t_2) \right) \right),
\end{aligned}$$

see Ohanissian et al. (2008) for details. Since $b(t_1, t_2)$ is uniformly zero for $t_1 = t_2$, straightforward calculations show that the expectation can be bounded by

$$E[R_{j,k}] = O\left(j/n^2 \right) O\left(n^2 \right) O\left(n^{2(d-1)} \right) = O\left(\frac{j}{n^{2-2d}} \right).$$

Applying the same approach, it can be shown that $Var\left(R_{j,k}\right) = O(\left(j/n^{2-2d}\right)^2)$, which implies that $I_{\tau_z}^{(k)}\left(\lambda_j\right) - kI_{\tau_z}^{(1)}\left(\lambda_j\right) = O_p\left(j/n^{2-2d}\right)$. This concludes the proof.

8.2 Proof of Lemma 2

From Ohanissian et al. (2008), we know that

$$I_{\tau_z}^{(k)}\left(\lambda_j\right) - kI_{\tau_z}^{(1)}\left(\lambda_j\right) = O_p\left(\frac{j}{n^{\kappa}}\right),$$

and for arbitrary aggregation levels, $k_u < k_v$, that

$$I_{\tau_z}^{(k_v)}\left(\lambda_j\right) - \frac{k_v}{k_u} I_{\tau_z}^{(k_u)}\left(\lambda_j\right) = O_p\left(\frac{j}{n^{\kappa}}\right),$$

where $\kappa = 1 - 2d$ when $d \in (0, 1/2)$ and from Lemma 1, we have $\kappa = 2 - 2d$ when $d \in [1/2, 1)$.

Since $I_{\tau_z}^{(k)}\left(\lambda_j\right) = f_{\tau_z}^{(k)}\left(\lambda_j\right) + O_p\left(j/n\right)$, see e.g. Priestley (1981), we can deduce that

$$f_{\tau_z}^{(k)}\left(\lambda_j\right) - kf_{\tau_z}^{(1)}\left(\lambda_j\right) = O_p\left(\frac{j}{n^{\kappa}}\right).$$

Combining the above results, we get

$$\frac{I_{\tau_z}^{(k)}\left(\lambda_j\right)}{f_{\tau_z}^{(k)}\left(\lambda_j\right)} - \frac{I_{\tau_z}^{(1)}\left(\lambda_j\right)}{f_{\tau_z}^{(1)}\left(\lambda_j\right)} = O_p\left(\frac{j}{n^{\kappa}}\right) \quad \text{and} \quad \frac{I_{\tau_z}^{(k_v)}\left(\lambda_j\right)}{f_{\tau_z}^{(k_v)}\left(\lambda_j\right)} - \frac{I_{\tau_z}^{(k_u)}\left(\lambda_j\right)}{f_{\tau_z}^{(k_u)}\left(\lambda_j\right)} = O_p\left(\frac{j}{n^{\kappa}}\right).$$

This concludes the proof.

8.3 Proof of Theorem 1

Since Soulier (2001) assumes stationarity of z_t, we cannot directly apply the results of Ohanissian et al. (2008) to prove the theorem. However, since

$$\sum_{j=1}^{m^{(k_1)}} \alpha c_j = 0 \quad \text{and} \quad \sum_{j=1}^{m^{(k_1)}} \left(\alpha c_j\right)^2 = 1,$$

by the construction of c_j and by defining $\alpha = 2\left(\sum_{i=1}^{K} \frac{\gamma_i^2}{m^{(k_i)}} + 2\sum_{i=2}^{K}\sum_{s=1}^{i-1} \frac{\gamma_s\gamma_i}{m^{(k_s)}}\right)^{-1/2}$, (see Ohanissian et al. (2008) for the definitions c_j), and since

$$\lim_{n\to\infty} \alpha^2 \left(\sum_{j=1}^{m^{(k_1)}} \left|c_j - c_{j+1}\right| + \left|c_{m^{(k_1)}}\right|\right)^2 \log(n) = 0,$$

because $c_j = O\left(\log m^{(k_1)}/m^{(k_1)}\right)$ and $\alpha = O(\sqrt{m^{(k_1)}})$, we can apply Hurvich et al. (2005, Proposition A.2) together with the results of Lemma 1 and 2 to prove the theorem. This concludes the proof.

8.4 Proof of Theorem 2

Using the conjecture of Velasco (2000)[14], it can be shown that the joint distribution of arbitrary linear combinations of log-periodogram regression (LPR) estimates obtained using temporally aggregated possibly non-Gaussian series is asymptotically normal.

For every aggregation level $(k_1, k_2, \dots k_K)$, the LPR estimator based on the tapered discrete Fourier transform is

$$\hat{d}_i^T = -\frac{1}{2S_{k_i}} \sum_{j}^{m^{(k_i)}} a_{j,k_i} \log I^{T,(k_i)}(\lambda_j), \quad j = 3, 6, 9, \dots, m^{(k_i)}, \tag{24}$$

where

$$\begin{aligned} a_{j,k_i} &= \log \lambda_j - \frac{1}{m^{(k_i)}} \sum_{j}^{m^{(k_i)}} \log \lambda_j = \log j - \frac{1}{m^{(k_i)}} \sum_{j}^{m^{(k_i)}} \log j, \\ S_{k_i} &= \sum_{j}^{m^{(k_i)}} a_{j,k_i}^2. \end{aligned}$$

Note that here and in the remainder of the proof, $\sum_j^{m^{(k_i)}}$ denotes the sum over $j = 3, 6, 9, \dots, m^{(k_i)}$, where $m^{(k_i)}$ is divisible by 3. If no tapering is used ($h_t = 1$) and the summation is $j = 1, 2, \dots, m^{(k_i)}$, then (24) reduces to the classical (non-tapered) LPR estimator.

Since we know from Lemma 2 that given the bandwidth restrictions

$$I_{\tau_z}^{(k_v)}(\lambda_j) - \frac{k_v}{k_u} I_{\tau_z}^{(k_u)}(\lambda_j) = o_p(1),$$

we can deduce that

$$\mathrm{Re}\left(w_{\tau_z}^{(k_v)}(\lambda_j) - \sqrt{\frac{k_v}{k_u}} w_{\tau_z}^{(k_u)}(\lambda_j) \right) = o_p(1),$$

where $\mathrm{Re}(x)$ is the real part of x. Therefore, by simple insertion, it follows that

$$I_{\tau_z}^{T,(k_v)}(\lambda_j) - \frac{k_v}{k_u} I_{\tau_z}^{T,(k_u)}(\lambda_j) = o_p(1).$$

Using this notation we can therefore write the weighted sum of K tapered LPR estimates as

$$\sum_{i=1}^{K} \gamma_i \hat{d}_i^T = -\sum_{i=1}^{K} \frac{\gamma_i}{2S_{k_i}} \sum_{j}^{m^{(k_i)}} a_{j,k_i} \log I^{T,(k_i)}(\lambda_j),$$

[14] Velasco (2000) proves the asymptotic normality of the tapered log-periodogram regression estimate for non-Gaussian time-series using pooled periodogram ordinates. However, he conjectures that extending the arguments, the limiting distribution theory holds for estimates without pooling.

where the K estimates are based on the K temporally aggregated series. Following Ohanissian et al. (2008), this can be written as

$$\begin{aligned}
&= -\frac{\gamma_1}{2S_{k_1}} \sum_{j}^{m^{(k_1)}} a_{j,k_1} \log I^{T,(k_1)}(\lambda_j) - \sum_{i=2}^{K} \frac{\gamma_i}{2S_{k_i}} \sum_{j}^{m^{(k_i)}} a_{j,k_i} \log I^{T,(k_1)}(\lambda_j) \\
&\quad - \sum_{i=2}^{K} \frac{\gamma_i}{2S_{k_i}} \sum_{j}^{m^{(k_i)}} a_{j,k_i} O_p\left(\frac{j}{n^{\kappa}}\right) \\
&= \sum_{j}^{m^{(k_1)}} c_j \log I^{T,(k_1)}(\lambda_j) + O_p\left(\frac{m^{(k_1)} \log m^{(k_1)}}{n^{\kappa}}\right), \qquad (25)
\end{aligned}$$

since $a_{j,k_i} = O(\log m^{(k_i)})$ and

$$c_j = \begin{cases} -\sum_{i=1}^{K} \frac{\gamma_i a_{j,k_i}}{2S_{k_i}} & \text{for } 1 \leq j \leq m^{(k_K)} \\ -\sum_{i=1}^{K-1} \frac{\gamma_i a_{j,k_i}}{2S_{k_i}} & \text{for } m^{(k_K)} + 1 \leq j \leq m^{(k_{K-1})} \\ \vdots & \vdots \\ -\frac{\gamma_1 a_{j,k_1}}{2S_{k_1}} & \text{for } m^{(k_2)} + 1 \leq j \leq m^{(k_1)} \end{cases}.$$

To finish the proof, we can use Moulines & Soulier (2003, Theorem 21) to state that (25) is asymptotically normal if

$$\sum_{j}^{m^{(k_1)}} (\alpha c_j)^2 = 1, \qquad (26)$$

$$\lim_{n\to\infty} \max_{1\leq j\leq m^{(k_1)}} |\alpha c_j| = 0, \qquad (27)$$

$$\lim_{n\to\infty} n^{-1} \sum_{h\neq j=1}^{m^{(k_1)}} (\alpha c_h)(\alpha c_j) = \tau, \qquad (28)$$

and

$$\max_{1\leq j\leq m^{(k_1)}} |\alpha c_j| = O(1), \qquad (29)$$

see Moulines & Soulier (2003) for details. The only difference to assuming Gaussianity of the underlying series is (29). So from Ohanissian et al. (2008), we know that (26) - (28) are fulfilled if α is defined accordingly.

Focusing on (26), we can write

$$\begin{aligned}\sum_{j}^{m^{(k_1)}} c_j^2 &= \frac{1}{4}\left(\sum_{j}^{m^{(k_1)}}\sum_{i=1}^{K}\left(\frac{\gamma_i a_{j,k_i}}{S_{k_i}}\right)^2 + 2\sum_{j}^{m^{(k_1)}}\sum_{i=2}^{K}\sum_{s=1}^{i-1}\frac{\gamma_s\gamma_i a_{j,k_s}a_{j,k_i}}{S_{k_s}S_{k_i}}\right)\\ &= \frac{1}{4}\left(\sum_{j}^{m^{(k_1)}}\sum_{i=1}^{K}\frac{\gamma_i^2 a_{j,k_i}^2}{S_{k_i}^2} + 2\sum_{j}^{m^{(k_1)}}\sum_{i=2}^{K}\sum_{s=1}^{i-1}\frac{\gamma_s\gamma_i a_{j,k_i}^2}{S_{k_s}S_{k_i}}\right.\\ &\left. +2\sum_{j}^{m^{(k_1)}}\sum_{i=2}^{K}\sum_{s=1}^{i-1}\frac{\gamma_s\gamma_i(\frac{1}{m^{(k_i)}}\sum_j^{m^{(k_i)}}\log j - \frac{1}{m^{(k_s)}}\sum_j^{m^{(k_s)}}\log j)a_{j,k_i}}{S_{k_s}S_{k_i}}\right),\end{aligned}$$

where

$$\begin{aligned}&2\sum_{j}^{m^{(k_1)}}\sum_{i=2}^{K}\sum_{s=1}^{i-1}\frac{\gamma_s\gamma_i(\frac{1}{m^{(k_i)}}\sum_j^{m^{(k_i)}}\log j - \frac{1}{m^{(k_s)}}\sum_j^{m^{(k_s)}}\log j)a_{j,k_i}}{S_{k_s}S_{k_i}}\\ =\;& 2O(\log j!)\sum_{i=2}^{K}\sum_{s=1}^{i-1}\frac{\gamma_s\gamma_i\sum_j^{m^{(k_1)}} a_{j,k_i}}{S_{k_s}S_{k_i}} = o(1).\end{aligned}$$

By definition we have that $\sum_j^{m^{(k_1)}} a_{j,k_i}^2 = S_{k_i}$, and by approximating sums by integrals we get $S_{k_i} = m^{(k_i)}$. Hence,

$$\lim_{n\to\infty}\sum_{j}^{m^{(k_1)}} c_j^2 = \frac{1}{4}\left(\sum_{i=1}^{K}\frac{\gamma_i^2}{m^{(k_i)}} + 2\sum_{i=2}^{K}\sum_{s=1}^{i-1}\frac{\gamma_s\gamma_i}{m^{(k_s)}}\right).$$

Defining

$$\alpha = 2\left(\sum_{i=1}^{K}\frac{\gamma_i^2}{m^{(k_i)}} + 2\sum_{i=2}^{K}\sum_{s=1}^{i-1}\frac{\gamma_s\gamma_i}{m^{(k_s)}}\right)^{-1/2}$$

then $\lim_{n\to\infty}\sum_j^{m^{(k_1)}}(\alpha c_j)^2 = 1$.

Notice that (26) and (28) imply that

$$\tau = \lim_{n\to\infty} n^{-1}(\sum_{j}^{m^{(k_1)}}\alpha c_j)^2.$$

Hence, $\tau = 0$ when $\sum_j^{m^{(k_1)}}\alpha c_j = 0$, which we have in our case since a_{j,k_i} is by definition deviations from mean so (28) is fulfilled. From Moulines & Soulier (2003, Section 9.3), we know that (28) is generally fulfilled by the LPR estimator and therefore also for the TLPR estimator and the sum of estimates. Lastly, note that since we know from Moulines & Soulier (2003) that for every aggregation level, k_i,

$$\lim_{n\to\infty}\max_{1\le j\le m^{(k_i)}}|c_{j,k_i}| = 0,$$

we have that the weighted sum c_j satisfies (27).

Notice that since $\alpha = O(\sqrt{m^{(k_1)}})$ this implies that $\alpha \sum_{i=1}^{K} \gamma_i \hat{d}_i^T$ can be written as

$$\sum_{j}^{m^{(k_1)}} c_j \log I^{T,(k_1)}(\lambda_j) + O_p\left(\frac{\left(m^{(k_1)}\right)^{3/2} \log m^{(k_1)}}{n^{\kappa}}\right),$$

and thus, given the restrictions on the bandwidths, we can apply Moulines & Soulier (2003, Theorem 21) to state that

$$2\left(\sum_{i=1}^{K} \frac{\gamma_i^2}{m^{(k_i)}} + 2\sum_{i=2}^{K}\sum_{s=1}^{i-1} \frac{\gamma_s \gamma_i}{m^{(k_s)}}\right)^{-1/2} \sum_{i=1}^{K} \gamma_i (\hat{d}_i^T - d)$$

is asymptotically normal with zero mean and variance $3\pi^2/24$.

Alternatively, we could have applied Hurvich et al. (2005, Proposition A.2) to state the same result since

$$\sum_{j}^{m^{(k_1)}} \alpha c_j = 0 \quad \text{and} \quad \sum_{j}^{m^{(k_1)}} (\alpha c_j)^2 = 1$$

by the construction of c_j, and since

$$\lim_{n\to\infty} \alpha^2 \left(\sum_{j}^{m^{(k_1)}} |c_j - c_{j+1}| + \left|c_{m^{(k_1)}}\right|\right)^2 \log(n) = 0,$$

because $c_j = O\left(\log m^{(k_1)}/m^{(k_1)}\right)$ and $\alpha = O(\sqrt{m^{(k_1)}})$. This concludes the proof.

8.5 Proof of Proposition 2.1

Proof follows by Ohanissian et al. (2008, Lemma 2). This concludes the proof.

8.6 Proof of Proposition 2.2

The proof follows by Geweke & Porter-Hudak (1983). This concludes the proof.

8.7 Proof of Theorem 3

We know from the non-temporal aggregated case that Lemma 3(e) of Frederiksen et al. (2008) implies Proposition 4.1 of Hurvich et al. (2005) and Lemma 3(a,b,c,d) of Frederiksen et al. (2008) implies Proposition 4.2 of Hurvich et al. (2005). Therefore, in order to prove Theorem 3 it is sufficient to adapt either of these results to our setting, see Hurvich et al. (2005) and Frederiksen et al. (2008).

Define

$$\begin{aligned} S_{m^{(k_i)}}(d_i,\theta_i) &= \frac{1}{m^{(k_i)}}\sum_{j=1}^{m^{(k_i)}} \varphi_j(d_i,\theta_i) j^{2d_i-2d_0}\varepsilon_j^{(k_i)}, \\ U_{m^{(k_i)}}(d_i,\theta_i) &= m^{(k_i)} S_{m^{(k_i)}}(d_i,\theta_i)\nabla Q_{m^{(k_i)}}(d_i,\theta_i), \end{aligned}$$

where

$$\varphi_j(d_i,\theta_i) = \frac{1+h\left(d_0,\theta_0,\lambda_j\right)}{1+h\left(d_i,\theta_i,\lambda_j\right)},$$

$$\varepsilon_j^{(k_i)} = I_{\tau_z}^{(k_i)}\left(\lambda_j\right)\left(f_{\tau_z}^{(k_i)}\left(\lambda_j\right)\right)^{-1},$$

and $\nabla Q_{m^{(k_i)}}(d_i,\theta_i)$ is the first-order derivative of the objective function. Here $I_{\tau_z}^{(k_i)}\left(\lambda_j\right)$ and $f_{\tau_z}^{(k_i)}\left(\lambda_j\right)$ are the periodogram and spectral density of the k_ith aggregated series, respectively. Furthermore, define

$$\begin{aligned} v_{1,j,m^{(k_i)}}\left(d_i,\theta_i\right) &= 2\log\left(\lambda_j\right) - \frac{\partial_d h\left(d_i,\theta_i,\lambda_j\right)}{1+h\left(d_i,\theta_i,\lambda_j\right)} - 2\frac{1}{m^{(k_i)}}\sum_{q=1}^{m^{(k_i)}}\log\left(\lambda_q\right) \\ &\quad +\frac{1}{m}\sum_{q=1}^{m^{(k_i)}}\frac{\partial_d h\left(d_i,\theta_i,\lambda_q\right)}{1+h\left(d_i,\theta_i,\lambda_q\right)} \\ v_{r,j,m^{(k_i)}}\left(d_i,\theta_i\right) &= \frac{\partial_{\theta_r} h\left(d_i,\theta_i,\lambda_j\right)}{1+h\left(d_i,\theta_i,\lambda_j\right)} - \frac{1}{m}\sum_{q=1}^{m^{(k_i)}}\frac{\partial_{\theta_r} h\left(d_i,\theta_i,\lambda_q\right)}{1+h\left(d_i,\theta_i,\lambda_q\right)},\ r=2,...,R_y+R_w+2, \\ N_{j,m^{(k_i)}}\left(d_i,\theta_i\right) &= \left(v_{1,j,m^{(k_i)}}\left(d_i,\theta_i\right),...,v_{R_y+R_w+2,j,m^{(k_i)}}\left(d_i,\theta_i\right)\right), \\ S_{m^{(k_i)}}^{(0)} &= S_{m^{(k_i)}}^{(0)}\left(d_0,\theta_0\right),\ U_{m^{(k_i)}}^{(0)} = U_{m^{(k_i)}}\left(d_0,\theta_0\right),\ N_{j,m^{(k_i)}}^{(0)} = N_{j,m^{(k_i)}}\left(d_0,\theta_0\right). \end{aligned}$$

This notation allows us to write the scaled derivative vector evaluated at the true parameters as

$$\begin{aligned} m^{(k_i)}B_{m^{(k_i)}}^{-1}\nabla Q_{m^{(k_i)}}\left(d_0,\theta_0\right) &= \left(S_{m^{(k_i)}}^{(0)}\right)^{-1} B_{m^{(k_i)}}^{-1} U_{m^{(k_i)}}^{(0)} \\ &= \left(S_{m^{(k_i)}}^{(0)}\right)^{-1} B_{m^{(k_i)}}^{-1}\sum_{j=1}^{m^{(k_i)}} N_{j,m^{(k_i)}}^{(0)}\varepsilon_j^{(k_i)}, \end{aligned}$$

where

$$\left(B_{m^{(k_i)}}\right)_{11} = \sqrt{m^{(k_i)}},\ \left(B_{m^{(k_i)}}\right)_{s+1,s+1} = \sqrt{m^{(k_i)}}\lambda_{m^{(k_i)}}^{2s} \text{ for } s=1,\dots,R_y,$$
$$\text{and } \left(B_{m^{(k_i)}}\right)_{l+R_y+2,l+R_y+2} = \sqrt{m^{(k_i)}}\lambda_{m^{(k_i)}}^{2d_0+2l} \text{ for } l=0,\dots,R_w.$$

From Frederiksen et al. (2008) we know that $\left(S^{(0)}_{m^{(k_i)}}\right)^{-1} B^{-1}_{m^{(k_i)}} U^{(0)}_{m^{(k_i)}}$ is asymptotically Gaussian with zero mean and covariance matrix Ω_{R_y,R_w}, such that $B_{m^{(k_i)}}\left(\hat{d}_i - d_0, \hat{\theta}_i - \theta_0\right)'$ is asymptotically Gaussian with mean zero and covariance matrix $\Omega^{-1}_{R_y,R_w}$. We need to show that the properly scaled, $\sum_{i=1}^{K} \gamma_i \widetilde{B}_{m^{(k_i)}}\left(\hat{d}_i - d_0, \hat{\theta}_i - \theta_0\right)'$ is also asymptotically Gaussian with mean zero and covariance matrix $(\alpha\Omega)^{-1}$, where $\widetilde{B}_{m^{(k_i)}} = \frac{1}{\sqrt{m^{(k_i)}}} B_{m^{(k_i)}}$. Two steps are needed as $n \to \infty$: (1) $\sum_{i=1}^{K} \gamma_i m^{(k_i)} \widetilde{B}^{-1}_{m^{(k_i)}} \nabla Q_{m^{(k_i)}}(d_i, \theta_i) \xrightarrow{d} N\left(0, \alpha\Omega_{R_y,R_w}\right)$ and (2) $\sum_{i=1}^{K} \gamma_i m^{(k_i)} \widetilde{B}^{-1}_{m^{(k_i)}} \nabla^2 Q_{m^{(k_i)}}(d_i, \theta_i) \widetilde{B}^{-1}_{m^{(k_i)}} \xrightarrow{p} \alpha\Omega_{R_y,R_w}$. (1) Follows by the adaption of Lemma 3(e) of Frederiksen et al. (2008) and/or Proposition 4.1 of Hurvich et al. (2005) to our setting. (2) Follows by the adaption of Lemma 3(a,b,c,d) of Frederiksen et al. (2008) and/or Proposition 4.2 of Hurvich et al. (2005) to our setting.

More specifically, following the idea of Ohanissian et al. (2008), write

$$\begin{aligned}
\sum_{i=1}^{K} \gamma_i m^{(k_i)} \widetilde{B}^{-1}_{m^{(k_i)}} \nabla Q_{m^{(k_i)}}(d_i, \theta_i) &= \sum_{i=1}^{K} \gamma_i \left(S_{m^{(k_i)}}\right)^{-1} \widehat{B}^{-1}_{m^{(k_i)}} \sum_{j=1}^{m^{(k_i)}} N_{j,m^{(k_i)}} \varepsilon_j^{(k_i)} \\
&= \sum_{i=1}^{K} \gamma_i \widehat{B}^{-1}_{m^{(k_i)}} \sum_{j=1}^{m^{(k_i)}} N_{j,m^{(k_i)}} \varepsilon_j^{(k_i)} + o_p(1). \\
&= \sum_{i=1}^{K} \gamma_i \widehat{B}^{-1}_{m^{(k_i)}} \sum_{j=1}^{m^{(k_i)}} N_{j,m^{(k_i)}} \varepsilon_j^{(k_1)} + O_p\left(\frac{m^{(k_1)} \log m^{(k_1)}}{n^{\kappa}}\right) \\
&= \sum_{j=1}^{m^{(k_1)}} c_j \varepsilon_j^{(k_1)} + O_p\left(\frac{m^{(k_1)} \log m^{(k_1)}}{n^{\kappa}}\right),
\end{aligned}$$

where $\widehat{B} = \sqrt{m^{(k_i)}} B_{m^{(k_i)}}$ and since

$$c_j = \begin{cases} \sum_{i=1}^{K} \gamma_i \widehat{B}^{-1}_{m^{(k_i)}} N_{j,m^{(k_i)}} & \text{for } 1 \le j \le m^{(k_K)} \\ \sum_{i=1}^{K-1} \gamma_i \widehat{B}^{-1}_{m^{(k_i)}} N_{j,m^{(k_i)}} & \text{for } m^{(k_K)} + 1 \le j \le m^{(k_{K-1})} \\ \vdots & \vdots \\ \gamma_1 \widehat{B}^{-1}_{m^{(k_1)}} N_{j,m^{(k_1)}} & \text{for } m^{(k_2)} + 1 \le j \le m^{(k_1)} \end{cases}.$$

The proof of asymptotic normality of $\sum_{j=1}^{m^{(k_1)}} c_j \varepsilon_j^{(k_1)}$ is based on the Wold device. Define, for any $x \in R^{R_y+R_w+2}$

$$t_n^2(x) = \sum_{j=1}^{m^{(k_1)}} \left(x' c_j\right)^2, \quad k_{n,j}(x) = \tau_n^{-1}(x)\, x' c_j, \quad \text{and} \quad T_n = \sum_{j=1}^{m^{(k_1)}} k_{n,j}(x)\, \varepsilon_j^{(k_1)}.$$

Using this notation, we have $x' \sum_{j=1}^{m^{(k_1)}} c_j \varepsilon_j^{(k_1)} = \tau_n(x) T_n$ and it suffices to prove that T_n is asymptotically standard Gaussian and that $\lim_{n\to\infty} t_n^2(x) = x'\alpha\Omega x$. Using the idea of

Ohanissian et al. (2008) and the proof to Proposition 4.1 in Hurvich et al. (2005) the last property can be obtained by writing

$$t_n^2(x) = \sum_{j=1}^{m^{(k_1)}} \left(x'c_j\right)^2 = t_{n,1,1}^2(x) + \sum_{l=1}^{R_y+R_w+2} \sum_{r=2}^{R_y+R_w+2} t_{n,l,s}^2(x), \tag{30}$$

and approximating sums by integrals. More specifically, without loss of generality focus on $t_{n,1,1}^2(x)$ (the other being routine applications of the same technique), where

$$t_{n,1,1}^2(x) = x_{11}^2 \sum_{j=1}^{m^{(k_1)}} \left(\sum_{i=1}^{K} \frac{\gamma_i}{m^{(k_i)}} v_{1,j,m^{(k_i)}} \right)^2.$$

Rewriting $t_{n,1,1}^2(x)$, we get

$$\begin{aligned}
\lim_{n\to\infty} t_{n,1,1}^2(x) &= \lim_{n\to\infty} x_{11}^2 \left(\sum_{i=1}^{K} \sum_{j=1}^{m^{(k_i)}} \left(\frac{\gamma_i}{m^{(k_i)}} v_{1,j,m^{(k_i)}} \right)^2 \right. \\
&\quad \left. + 2 \sum_{i=2}^{K} \sum_{s=1}^{i-1} \sum_{j=1}^{m^{(k_i)}} \frac{\gamma_s \gamma_i}{m^{(k_s)} m^{(k_i)}} v_{1,j,m^{(k_s)}} v_{1,j,m^{(k_i)}} \right). \\
\lim_{n\to\infty} \frac{t_{n,1,1}^2(x)}{x_{11}^2} &= \lim_{n\to\infty} \sum_{i=1}^{K} \sum_{j=1}^{m^{(k_i)}} \frac{\gamma_i^2}{m^{(k_i)}} \frac{v_{1,j,m^{(k_i)}}^2}{m^{(k_i)}} \\
&\quad + \lim_{n\to\infty} 2 \sum_{i=2}^{K} \sum_{s=1}^{i-1} \sum_{j=1}^{m^{(k_i)}} \frac{\gamma_s \gamma_i}{m^{(k_s)}} \frac{v_{1,j,m^{(k_s)}} v_{1,j,m^{(k_i)}}}{m^{(k_i)}} \\
&= \lim_{n\to\infty} t_{n,1,1,1}^2 + \lim_{n\to\infty} t_{n,1,1,2}^2,
\end{aligned}$$

and approximating sums by integrals (as this is elementary calculus, although tedious, we omit the intermediate steps)

$$\lim_{n\to\infty} t_{n,1,1,1}^2 = 4 \sum_{i=1}^{K} \frac{\gamma_i^2}{m^{(k_i)}}, \quad \lim_{n\to\infty} t_{n,1,1,2}^2 = 8 \sum_{i=2}^{K} \sum_{s=1}^{i-1} \frac{\gamma_s \gamma_i}{m^{(k_s)}}.$$

Thus,

$$\lim_{n\to\infty} \frac{t_{n,1,1}^2(x)}{x_{11}^2} = 4 \left(\sum_{i=1}^{K} \frac{\gamma_i^2}{m^{(k_i)}} + 2 \sum_{i=2}^{K} \sum_{s=1}^{i-1} \frac{\gamma_s \gamma_i}{m^{(k_s)}} \right).$$

Applying the same procedure, and therefore omitted, to the other elements of (30), we get that $\lim_{n\to\infty} t_n^2(x) = x'\alpha\Omega x$.

Now we only need to prove that T_n is asymptotically standard Gaussian. In order to apply the results of Hurvich et al. (2005, Proposition A.2), we need

$$\sum_{j=1}^{m^{(k_1)}} k_{n,j}(x) = 0, \quad \sum_{j=1}^{m^{(k_1)}} k_{n,j}^2(x) = 1, \tag{31}$$

and

$$\lim_{n\to\infty}\left(\sum_{j=1}^{m^{(k_1)}} |k_{n,j}(x) - k_{n,j+1}(x)| + \left|k_{n,m^{(k_1)}}(x)\right|\right)^2 \log(n) = 0. \tag{32}$$

Notice that (31) is fulfilled by the construction of c_j and $k_{n,j}(x)$, i.e.

$$\sum_{j=1}^{m^{(k_1)}} k_{n,j}(x) = \sum_{j=1}^{m^{(k_1)}} \frac{xc_j}{\sqrt{\sum_{j=1}^{m^{(k_1)}} (xc_j)^2}} = \frac{\sum_{j=1}^{m^{(k_1)}} xc_j}{\sqrt{\sum_{j=1}^{m^{(k_1)}} (xc_j)^2}} = 0,$$

since $\sum_{j=1}^{m^{(k_1)}} c_j = 0$ by construction, and

$$\sum_{j=1}^{m^{(k_1)}} k_{n,j}^2(x) = \sum_{j=1}^{m^{(k_1)}} \frac{(xc_j)^2}{\sqrt{\sum_{j=1}^{m^{(k_1)}} (xc_j)^2}^2} = 1.$$

Furthermore, since

$$|c_j| = O\left(2\sum_{i=1}^{K}\gamma_i \frac{1}{m^{(k_i)}}\log m^{(k_i)}\right) \quad \text{and} \quad |c_j - c_{j+1}| = O\left(2\sum_{i=1}^{K}\gamma_i\frac{1}{m^{(k_i)}}j^{-1}\right),$$

(32) is fulfilled. Thus, we can apply Hurvich et al. (2005, Proposition A.2) to state that T_n is asymptotically standard Gaussian.

To finish the proof, we show that the sum of the weighted (and scaled) Hessian converges in probability to $\alpha\Omega_{R_y,R_w}$. Now, if we write the weighted sum of the scaled derivatives vector as

$$\begin{aligned}
&\sum_{i=1}^{K}\gamma_i m^{(k_i)}\widetilde{B}^{-1}_{m^{(k_i)}}\nabla Q_{m^{(k_i)}}(d_i,\theta_i)\\
=\ &\sum_{i=1}^{K}\sum_{j=1}^{m^{(k_i)}}\gamma_i\left(S_{m^{(k_i)}}(d_i,\theta_i)\right)^{-1}\widehat{B}^{-1}_{m^{(k_i)}}N_{j,m^{(k_i)}}(d_i,\theta_i)\varphi_j(d_i,\theta_i)j^{2d_i-2d_0}\varepsilon_j^{(k_i)}\\
=\ &\left(S_{m^{(k_1)}}(d_1,\theta_1)\right)^{-1}\sum_{j=1}^{m^{(k_1)}}\sum_{i=1}^{K}\gamma_i\widehat{B}^{-1}_{m^{(k_i)}}N_{j,m^{(k_i)}}(d_i,\theta_i)\varphi_j(d_i,\theta_i)j^{2d_i-2d_0}\varepsilon_j^{(k_1)} + o_p(1),
\end{aligned}$$

we can write the scaled Hessian as

$$
\begin{aligned}
& \left(S_{m^{(k_1)}}(d_1,\theta_1)\right)^{-1}\sum_{j=1}^{m^{(k_1)}}\sum_{i=1}^{K}\gamma_i\widehat{B}^{-1}_{m^{(k_i)}}N_{j,m^{(k_i)}}(d_i,\theta_i)\times \\
& \qquad\qquad \{\sum_{i=1}^{K}\gamma_i\widehat{B}^{-1}_{m^{(k_i)}}\nabla(\varphi_j(d_i,\theta_i)j^{2d_i-2d_0})'\}\varepsilon_j^{(k_i)} \\
& +\left(S_{m^{(k_1)}}(d_1,\theta_1)\right)^{-1}\sum_{j=1}^{m^{(k_1)}}\sum_{i=1}^{K}\gamma_i\widehat{B}^{-1}_{m^{(k_i)}}\times \\
& \qquad\qquad (\sum_{i=1}^{K}\gamma_i\widehat{B}^{-1}_{m^{(k_i)}}\nabla N_{j,m^{(k_i)}}(d_i,\theta_i))\varphi_j(d_i,\theta_i)j^{2d_i-2d_0}\varepsilon_j^{(k_i)} \\
& -\left(S_{m^{(k_1)}}(d_1,\theta_1)\right)^{-2}\sum_{j=1}^{m^{(k_1)}}\sum_{i=1}^{K}\gamma_i\widehat{B}^{-1}_{m^{(k_i)}}N_{j,m^{(k_i)}}(d_i,\theta_i)\varphi_j(d_i,\theta_i)j^{2d_i-2d_0}\varepsilon_j^{(k_i)}\times \\
& \qquad\qquad (\sum_{i=1}^{K}\gamma_i\widehat{B}^{-1}_{m^{(k_i)}}\nabla S_{m^{(k_i)}}(d_i,\theta_i))' \\
= \; & \left(S_{m^{(k_1)}}(d_1,\theta_1)\right)^{-1}M_{1,m^{(k_i)}}(d_i,\theta_i)+\left(S_{m^{(k_1)}}(d_1,\theta_1)\right)^{-1}M_{2,m^{(k_i)}}(d_i,\theta_i) \\
& -\left(S_{m^{(k_1)}}(d_1,\theta_1)\right)^{-2}M_{3,m^{(k_i)}}(d_i,\theta_i).
\end{aligned}
$$

Since $S_{m^{(k_1)}}$ converges uniformly to 1, we only need to prove that $M_{1,m^{(k_i)}}$ converges in probability to $\alpha\Omega_{R_y,R_w}$ uniformly w.r.t. $(d,\theta)\in D\times\Theta$, where $\alpha=\left(\sum_{i=1}^{K}\frac{\gamma_i^2}{m^{(k_i)}}+2\sum_{i=2}^{K}\sum_{s=1}^{i-1}\frac{\gamma_s\gamma_i}{m^{(k_s)}}\right)^{-1/2}$, and that $M_{2,m^{(k_i)}}$ and $M_{3,m^{(k_i)}}$ converge to 0. We prove the first fact, the other being routine applications of the same technique, and therefore omitted.

Let $B_{l,m^{(k_i)}}$ be the lth diagonal element of $B_{m^{(k_i)}}$. Thus, the derivatives wrt. θ_i are (dette er forkert!!!, se Hurvich + hvordan vi definerer tingene!)

$$
\begin{aligned}
& M^{(l,s)}_{1,m^{(k_i)}}(d_i,\theta_i) \\
= \; & -\sum_{j=1}^{m^{(k_1)}}\sum_{i=1}^{K}\gamma_i\widehat{B}^{-1}_{l,m^{(k_i)}}v_{l,j,m^{(k_i)}}\sum_{i=1}^{K}\gamma_i\widehat{B}^{-1}_{r,m^{(k_i)}}\frac{\partial_{\theta_r}h\left(d_i,\theta_i,\lambda_j\right)}{1+h\left(d_i,\theta_i,\lambda_j\right)}\varphi_j(d_i,\theta_i)j^{2d_i-2d_0}\varepsilon_j^{(k_1)},
\end{aligned}
$$

for $l=1,...,R_y+R_w+2$ and $r=2,...,R_y+R_w+2$. Since $\sum_{j=1}^{m^{(k_i)}}v_{l,j,m^{(k_i)}}=0$, this expression can be rewritten as

$$
\begin{aligned}
& M^{(l,s)}_{1,m^{(k_i)}}(d_i,\theta_i) \\
= & -\sum_{j=1}^{m^{(k_1)}} (\sum_{i=1}^{K} \gamma_i \widehat{B}^{-1}_{l,m^{(k_i)}} v_{l,j,m^{(k_i)}} \sum_{i=1}^{K} \gamma_i \widehat{B}^{-1}_{r,m^{(k_i)}} v_{s,j,m^{(k_i)}}) & (33) \\
& -\sum_{j=1}^{m^{(k_1)}} \sum_{i=1}^{K} \gamma_i \widehat{B}^{-1}_{l,m^{(k_i)}} v_{l,j,m^{(k_i)}} \sum_{i=1}^{K} \gamma_i \widehat{B}^{-1}_{r,m^{(k_i)}} \frac{\partial_{\theta_r} h(d_i,\theta_i,\lambda_j)}{1+h(d_i,\theta_i,\lambda_j)} (\varphi_j(d_i,\theta_i) j^{2d_i-2d_0} - 1) & (34) \\
& -\sum_{i=1}^{K} \sum_{j=1}^{m^{(k_i)}} \gamma_i \widehat{B}^{-1}_{l,m^{(k_i)}} v_{l,j,m^{(k_i)}} \sum_{i=1}^{K} \gamma_i \widehat{B}^{-1}_{r,m^{(k_i)}} \frac{\partial_{\theta_r} h(d_i,\theta_i,\lambda_j)}{1+h(d_i,\theta_i,\lambda_j)} \varphi_j(d_i,\theta_i) j^{2d_i-2d_0} (\varepsilon_j^{(k_1)} - 1) & (35)
\end{aligned}
$$

Approximating sums by integrals (33) converges to the (l,s)th element of $\alpha\Omega_{R_y,R_w}$. From the consistency of the LPWN estimates and that $\left|\widehat{B}^{-1}_{l,m^{(k_i)}} v_{l,j,m^{(k_i)}}\right| \leq C\ln(n)\left(m^{(k_i)}\right)^{-1}$, we know that (34) converges to 0. Finally, the term (35) is $o_p(1)$ by Hurvich et al. (2005, Proposition A.1) since $\sum_{j=1}^{m^{(k_i)}} v_{l,j,m^{(k_i)}} = 0$ and $\varphi_j(d_i,\theta_i) j^{2d_i-2d_0} \xrightarrow{p} 1$.

Concerning the derivatives wrt. d_i

$$
\begin{aligned}
& M^{(l,1)}_{1,m^{(k_i)}}(d_i,\theta_i) \\
= & \sum_{j=1}^{m^{(k_1)}} \sum_{i=1}^{K} \gamma_i \widehat{B}^{-1}_{l,m^{(k_i)}} v_{l,j,m^{(k_i)}} \sum_{i=1}^{K} \gamma_i \widehat{B}^{-1}_{1,m^{(k_i)}} \mu_j^i \varphi_j(d_i,\theta_i) j^{2d_i-2d_0} \varepsilon_j^{(k_1)},
\end{aligned}
$$

where

$$
\mu_j^i = 2\ln(j) - 2\lambda_j^{2d_i} \ln(\lambda_j)\theta_i,
$$

we have

$$
\begin{aligned}
& M^{(l,1)}_{1,m^{(k_i)}}(d_i,\theta_i) \\
= & \sum_{j=1}^{m^{(k_1)}} (\sum_{i=1}^{K} \gamma_i \widehat{B}^{-1}_{l,m^{(k_i)}} v_{l,j,m^{(k_i)}} \sum_{i=1}^{K} \gamma_i \widehat{B}^{-1}_{1,m^{(k_i)}} v_{1,j,m^{(k_i)}}) & (36) \\
& + \sum_{j=1}^{m^{(k_1)}} \sum_{i=1}^{K} \gamma_i \widehat{B}^{-1}_{l,m^{(k_i)}} v_{l,j,m^{(k_i)}} \sum_{i=1}^{K} \gamma_i \widehat{B}^{-1}_{1,m^{(k_i)}} \mu_j^i (\varphi_j(d_i,\theta_i) j^{2d_i-2d_0} - 1) & (37) \\
& + \sum_{j=1}^{m^{(k_1)}} \sum_{i=1}^{K} \gamma_i \widehat{B}^{-1}_{l,m^{(k_i)}} v_{l,j,m^{(k_i)}} \sum_{i=1}^{K} \gamma_i \widehat{B}^{-1}_{1,m^{(k_i)}} \mu_j^i \varphi_j(d_i,\theta_i) j^{2d_i-2d_0} (\varepsilon_j^{(k_i)} - 1), & (38)
\end{aligned}
$$

since $\sum_{j=1}^{m^{(k_i)}} v_{l,j,m^{(k_i)}} = 0$.

Again, it is seen by approximating sums by integrals that (36) converges to the $(l,1)$th element of $\alpha\Omega_{R_y,R_w}$, and from the arguments above that (37) converges to 0. That (38) is $o_p(1)$ can be seen since $\sum_{j=1}^{m^{(k_i)}} v_{l,j,m^{(k_i)}} = 0$ and $\varphi_j(d_i,\theta_i)j^{2d_i-2d_0} \xrightarrow{p} 1$ such that the assumptions to apply Hurvich et al. (2005, Proposition A.1) are satisfied.

Hence, the sum of the weighted (and scaled) Hessian converges in probability to $\alpha\Omega_{R_y,R_w}$ such that $\sum_{i=1}^{K}\gamma_i\tilde{B}_{m^{(k_i)}}\left(\hat{d}_i - d_0, \hat{\theta}_i - \theta_0\right)'$ by standard arguments is asymptotically Gaussian with mean zero and variance $(\alpha\Omega_{R_y,R_w})^{-1}$. By defining $\alpha = \left(\sum_{i=1}^{K}\frac{\gamma_i^2}{m^{(k_i)}} + 2\sum_{i=2}^{K}\sum_{s=1}^{i-1}\frac{\gamma_s\gamma_i}{m^{(k_s)}}\right)^{-1/2}$, know from the derivations above that the scaled derivatives vector $\alpha t_n(x)\, T_n$ is asymptotically centered Gaussian with variance Ω_{R_y,R_w}, and that the scaled Hessian converges in probability also to Ω_{R_y,R_w}. This concludes the proof.

8.8 Proof of Proposition 3.1

Here we follow the proof of Lemma 2 of Ohanissian et al. (2008) and applying the above distribution result for the simple sum of two estimators, i.e. $\gamma_i = \gamma_j = 1$ for $1 \le i < j \le K$ and $\gamma_k = 0$ for all $1 \le k \le K$ such that $k \ne i,j$, we get

$$\lim_{n\to\infty} m^{(k_i)} cov\left(\hat{d}_i, \hat{d}_j\right) = \Omega_{R_y,R_w}.$$

This implies that the asymptotic covariance between any LPWN estimates obtained using temporal aggregated series equals the variance of the lesser aggregated series, i.e.

$$\lim_{n\to\infty}\left(m^{(k_i)}\left(Cov\left(\hat{d}_i,\hat{d}_j\right) - Var\left(\hat{d}_i\right)\right)\right) = 0.$$

This concludes the proof.

8.9 Proof of Proposition 3.2

Following Hurvich & Ray (2003), we conjecture that the covariance matrix is well approximated by $\mathrm{cov}(m\nabla Q(d,\theta))^{-1}$.

For any m (and corresponding aggregation level), consider the objective function

$$Q(d,\theta) = \log\left(\frac{1}{m}\sum_{j=1}^{m}\frac{\lambda_j^{2d} I_z(\lambda_j)}{1+h(d,\theta,\lambda_j)}\right) + \frac{1}{m}\sum_{j=1}^{m}\log\left(\lambda_j^{-2d}\left(1+h(d,\theta,\lambda_j)\right)\right). \tag{39}$$

Thus, the score multiplied by m can be written as

$$m\nabla Q_m(d,\theta) = \hat{G}(d,\theta)^{-1}\sum_{j=1}^{m}\left(\frac{\lambda_j^{2d} I_z(\lambda_j)}{1+h(d,\theta,\lambda_j)} - \frac{1}{m}\sum_{q=1}^{m}\frac{\lambda_q^{2d} I_z(\lambda_q)}{1+h(d,\theta,\lambda_q)}\right) X_j,$$

where

$$
\begin{aligned}
X_j &= (X_{1j}, X_{2j}', X_{3j}')', \\
X_{1j} &= 2\log\lambda_j - \frac{2h_w(\theta_w,\lambda_j)\lambda_j^{2d}\log\lambda_j}{1+h(d,\theta,\lambda_j)}, \\
X_{2j} &= \left(\frac{-\lambda_j^2}{1+h(d,\theta,\lambda_j)}, \ldots, \frac{-\lambda_j^{2R_y}}{1+h(d,\theta,\lambda_j)}\right)', \\
X_{3j} &= \left(\frac{-\lambda_j^{2d}}{1+h(d,\theta,\lambda_j)}, \ldots, \frac{-\lambda_j^{2d+2R_w}}{1+h(d,\theta,\lambda_j)}\right)',
\end{aligned}
$$

where X_j is the vector of partial derivatives of $-\log\left(\lambda_j^{-2d} G\left(1+h(d,\theta,\lambda_j)\right)\right)$. For simplicity lets look at the score wrt. d evaluated at the true parameters $\left(d_0\,,\theta_0'\right)'$.

$$
\begin{aligned}
m\nabla_d Q_m(d_0,\theta_0) &= \begin{array}{c}\left(\frac{1}{m}\sum_{j=1}^m \frac{\lambda_j^{2d_0} I_z(\lambda_j)}{1+h(d_0,\theta_0,\lambda_j)}\right)^{-1}\sum_{j=1}^m\left(\frac{\lambda_j^{2d_0} I_z(\lambda_j)}{1+h(d_0,\theta_0,\lambda_j)} - \frac{1}{m}\sum_{q=1}^m \frac{\lambda_q^{2d_0} I_z(\lambda_q)}{1+h(d_0,\theta_0,\lambda_q)}\right) \\ \times\left(2\log\lambda_j - \frac{2h_w(\theta_w,\lambda_j)\lambda_j^{2d_0}\log\lambda_j}{1+h(d_0,\theta_0,\lambda_j)}\right)\end{array} \\
&= 2m\left(\frac{\sum_{j=1}^m v_{1,j}\frac{\lambda_j^{2d_0} I_z(\lambda_j)}{1+h(d_0,\theta_0,\lambda_j)}}{\sum_{j=1}^m \frac{\lambda_j^{2d_0} I_z(\lambda_j)}{1+h(d_0,\theta_0,\lambda_j)}}\right),
\end{aligned}
$$

where $v_{1,j} = \left(\log\lambda_j - \frac{h_w(\theta_w,\lambda_j)\lambda_j^{2d_0}\log\lambda_j}{1+h(d_0,\theta_0,\lambda_j)}\right) - \frac{1}{m}\sum_{q=1}^m\left(\log\lambda_q - \frac{h_w(\theta_w,\lambda_q)\lambda_q^{2d_0}\log\lambda_q}{1+h(d_0,\theta_0,\lambda_q)}\right)$. Then assuming that $I_z(\lambda_j)$ is independently distributed as $\frac{1}{2}\lambda_j^{-2d_0}\frac{1}{1+h(d_0,\theta_0,\lambda_j)}G\chi_2^2$ it follows that

$$
\begin{aligned}
E\left[m\nabla_d Q_m(d_0,\theta_0)\right] &= E\left[2m\left(\frac{\sum_{j=1}^m v_{1,j}\frac{\lambda_j^{2d_0} I_z(\lambda_j)}{1+h(d_0,\theta_0,\lambda_j)}}{\sum_{j=1}^m \frac{\lambda_j^{2d_0} I_z(\lambda_j)}{1+h(d_0,\theta_0,\lambda_j)}}\right)\right] \\
&= 2\left(\frac{m\sum_{j=1}^m E\left[v_{1,j}\frac{\lambda_j^{2d_0} I_z(\lambda_j)}{1+h(d_0,\theta_0,\lambda_j)}\right]}{mG}\right) \\
&= 2\left(\sum_{j=1}^m v_{1,j} E\left[\frac{\lambda_j^{2d_0} I_z(\lambda_j)}{1+h(d_0,\theta_0,\lambda_j)}G^{-1}\right]\right) = 0,
\end{aligned}
$$

since $E\left[\frac{\lambda_j^{2d_0} I_z(\lambda_j)}{1+h(d_0,\theta_0,\lambda_j)}G^{-1}\right] = 1$ and $\sum_{j=1}^m v_{1,j} = 0$. The same holds for the other entries of the score, i.e. $E\left[m\nabla Q_m(d_0,\theta_0)\right] = 0$, as $E\left[\frac{\lambda_j^{2d_0} I_z(\lambda_j)}{1+h(d_0,\theta_0,\lambda_j)}G^{-1}\right] = 1$ and $\sum_{j=1}^m v_{s,j} = 0$ for $s = 1, ..., R_y + R_w + 2$. Hence, using $\mathrm{cov}(m\nabla Q_m(d_0,\theta_0)) = E\left[\left(m\nabla Q_m(d_0,\theta_0)\right)\left(m\nabla Q_m(d_0,\theta_0)\right)'\right]$, the covariance matrix is approximately given by Γ^{-1}. This concludes the proof.

9 Appendix B: Specification of the spurious long memory models

Stationary and Nonstationary Random Level Shift Models (SRLS and NSRLS)

$$z_t = \mu_t + \varepsilon_t,\ \mu_t = (1 - j_t \mathbb{I})\,\mu_{t-1} + j_t\eta_t,\ j_t = \begin{cases} P\,(j_t = 0) = 1 - p \\ P\,(j_t = 1) = p \end{cases},$$

where ε_t is $NID\left(0, \sigma_\varepsilon^2\right)$, η_t is $NID\left(0, \sigma_\eta^2\right)$ and $\mathbb{I}$ is the indicator function taking the value 1 in the stationary case (SRLS).

Markow Switching Models (MS-IID and MS-GARCH)

$$(MS - IID) : z_t = \begin{cases} N\left(\mu_0, \sigma^2\right) & \text{if } s_t = 0 \\ N\left(\mu_1, \sigma^2\right) & \text{if } s_t = 1 \end{cases}$$

and

$$(MS - GARCH) : z_t = \sigma_t\epsilon_t,\ \sigma_t^2 = \omega_0 + (\omega_1 - \omega_0)\,s_t + \alpha z_{t-1}^2 + \beta\sigma_{t-1}^2,\ \epsilon_t \sim N\,(0,1)$$

with (transition) probability p_0 of remaining in regime 0 ($s_t = 0$) and (transition) probability p_1 of remaining in regime 1 ($s_t = 1$).

White Noise with Trend Model (WNT)

$$z_t = ct^{\beta - 1/2} + u_t,$$

where c is a constant and u_t is $NID\left(0, \sigma_u^2\right)$.

Table 9 shows the parameter values we use in the power study.

[Table 9 about her]

References

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Ebens, H. (2001), 'The distribution of realized stock return volatility', *Journal of Financial Economics* **61**, 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2001), 'The distribution of realized exchange rate volatility', *Journal of the American Statistical Association* **96**, 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), 'Modelling and forecasting realized volatility', *Econometrica* **71**, 579–625.

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*a*), Augmented log-periodogram regression in long memory signal plus noise models, *in* '2004 Hawaii International Conference on Statistics, Mathematics and Related Fields', pp. 108–119.

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

Arteche, J. (2006), 'Semiparametric estimation in perturbed long memory series', *Computational Statistics & Data Analysis* **51**, 2118–2141.

Baillie, R. T. (1996), 'Long memory processes and fractional integration in econometrics', *Journal of Econometrics* **73**, 5–59.

Baillie, R. T., Bollerslev, T. & Mikkelsen, H. O. (1996), 'Fractionally integrated generalized autoregressive conditional heteroscedasticity', *Journal of Econometrics* **74**, 3–30.

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

Beran, J. & Ocker, D. (2000), 'Temporal aggregation of stationary and nonstationary farima(p,d,0) models', *CoFE Discussion Paper, number* **00-22**.

Bhattacharya, R., Gupta, V. & Waymire, E. (1983), 'The hurst effect under trends', *Journal of Applied Probability* **20**, 649–662.

Bloomfield, P. (1973), 'An exponential model for the spectrum of a scalar time series', *Biometrika* **60**, 217–226.

Calvet, L. E., Fisher, A. J. & Mandelbrot, B. B. (1997), 'Large deviations and the distribution of price changes', *Cowles Foundation Discussion Paper* **1165**.

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

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

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

Deo, R. S. & Hurvich, C. M. (2002), Estimation of long memory in volatility, *in* P. Doukhan, G. Oppenhein & M. Taqqu, eds, 'Theory and Applications of Long-Range Dependence', Boston: Birkhauser.

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

Ding, Z., Granger, C. W. J. & Engle, R. F. (1993), 'A long memory property of stock returns and a new model', *Journal of Empirical Finance* **1**, 83–106.

Fisher, A. J., Calvet, L. E. & Mandelbrot, B. B. (1997), 'Multifractality of deutschemark / us dollar exchange rates', *Cowles Foundation Discussion Paper* **1166**.

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* .

Frederiksen, P. H. & Nielsen, M. Ø. (2008), 'Bias-reduced estimation of long memory stochastic volatility', *Forthcoming in Journal of Financial Econometrics* .

Fuller, W. A. (1996), *Introduction to statistical time series*, Wiley, New York.

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

Gil-Alana, L. A. (2003), 'Fractional integration with bloomfield disturbances in the specification of real output in the g7 countries', *Topics in Macroeconomics* **3**. Article 11.

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. & Joyeux, R. (1980), 'An introduction to long-memory time series models and fractional differencing', *Journal of Time Series Analysis* **1**, 15–29.

Harvey, A. (1989), *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press, Cambridge.

Hausman, J. (1978), 'Specification tests in econometrics', *Econometrica* **46**, 1251–1271.

Henry, M. & Zaffaroni, P. (2003), The long range paradigm for macroeconomics and finance, *in* P. Doukhan, G. Oppenheim & M. S. Taqqu, eds, 'Theory and Applications of Long-Range Dependence', Birkhäuser, Boston, pp. 417–438.

Hosking, J. R. M. (1981), 'Fractional differencing', *Biometrika* **68**, 165–176.

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.

Kim, C. & Phillips, P. (2006), 'Log periodogram regression in the nonstationary case', *Cowles Foundation Discussion Papers* **1587**.

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.

Man, K. & Tiao, G. (2001), 'Fractional gaussian noise and arfima processes', *Unpublished manuscript, Syracuse University and University of Chicago* .

Man, K. & Tiao, G. (2006), 'Aggregation effect and forecasting temporal aggregates of long memory processes', *International Journal of Forecasting* **22**, 267–281.

Mandelbrot, B. B., Fisher, A. J. & Calvet, L. E. (1997), 'A multifractal model of asset returns', *Cowles Foundation Discussion Paper* **1164**.

Mandelbrot, B. B. & van Ness, J. W. (1968), 'Fractional brownian motion fractional noises and applications', *SIAM Rev.* **10**, 422–437.

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

Moulines, E. & Soulier, P. (2003), Semiparametric estimation for fractional processes, *in* P. Doukhan, G. Oppenheim & M. Taqqu, eds, 'Long Range Dependence: Theory and Applications', Birkhauser, Boston.

Nielsen, M. Ø. (2004), 'Efficient likelihood inference in nonstationary univariate models', *Econometric Theory* **20**, 116–146.

Ohanissian, A., Russell, J. R. & Tsay, R. S. (2003), 'True or spurious long memory in volatility: Does it matter for pricing options?', *Working paper, University of Chicago* .

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.

Perron, P. & Qu, Z. (2004), 'An analytical evaluation of the log-periodogram estimate in the presence of level shifts and its implication for stock returns', *Working Paper, Department of Economics, Boston University* .

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

Priestley, M. B. (1981), 'Spectral analysis of time series', *Academic Press (London)* .

Ray, B. K. & Tsay, R. (2000), 'Long-range dependence in daily stock volatility', *Journal of Business and Economic Statistics* **18**, 254–262.

Robinson, P. M. (1994*a*), 'Efficient tests of nonstationary hypotheses', *Journal of the American Statistical Association* **89**, 1420–1437.

Robinson, P. M. (1994b), Time series with strong dependence, *in* C. Sims, ed., 'Advances in Econometrics', Cambridge University Press, Cambridge, pp. 47–95.

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

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

Soulier, P. (2001), 'Moment bounds and central limit theorems for functions of gaussian vectors', *Statistics & Probability Letters* **54**, 193–203.

Souza, L. R. (2003a), 'The aliasing effect, the fejer kernel and temporally aggregated long memory processes', *Working Paper, EPGE - Getulio Vargas Foundation* .

Souza, L. R. (2003b), 'Temporal aggregation and bandwidth selection in estimating long memory', *Working Paper, EPGE - Getulio Vargas Foundation* .

Souza, L. R. (2004), 'Spectral properties of temporally aggregated long memory processes', *Working Paper, EPGE - Getulio Vargas Foundation* .

Sun, Y. & Phillips, P. C. B. (2003), 'Nonlinear log-periodogram regression for perturbed fractional processes', *Journal of Econometrics* **115**, 355–389.

Taqqu, M. S. (2003), Fractional brownian motion and longe-range dependence, *in* P. Doukhan, G. Oppenheim & M. S. Taqqu, eds, 'Theory ans Applications of Longe-Range Dependence', Birkhäuser, pp. 5–38.

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

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

Velasco, C. (2000), 'Non-gaussian log-periodogram regression', *Econometric Theory* **16**, 44–79.

Woodward, W. A., Cheng, Q. C. & Gray, H. L. (1998), 'A k-factor garma long-memory model', *Journal of Time Series Analysis* **19**, 485–504.

Wright, J. H. (2002), 'Log periodogram estimation of long memory volatility dependencies with conditionally heavy tailed returns', *Econometric Reviews* **21**, 397–417.

Table 1: Empirical size properties for Model I

	$d = 0.4$				$d = 0.6$			
K	2	3	4	5	2	3	4	5
Panel A: $m = \lfloor n^{0.5} \rfloor$								
LPR	0.048	0.059	0.067	0.063	0.051	0.055	0.059	0.061
$TLPR$	0.050	0.058	0.063	0.068	0.054	0.059	0.062	0.059
LW	0.059	0.061	0.075	0.091	0.074	0.077	0.080	0.093
LPW	0.066	0.070	0.089	0.106	0.047	0.051	0.064	0.061
LWN	0.035	0.039	0.052	0.063	0.048	0.059	0.063	0.071
$LPWN$	0.039	0.041	0.059	0.068	0.042	0.057	0.065	0.075
Panel B: $m = \lfloor n^{0.6} \rfloor$								
LPR	0.048	0.057	0.065	0.061	0.049	0.054	0.057	0.061
$TLPR$	0.049	0.057	0.063	0.063	0.054	0.055	0.061	0.059
LW	0.058	0.061	0.074	0.088	0.064	0.069	0.070	0.084
LPW	0.062	0.065	0.079	0.091	0.047	0.050	0.064	0.063
LWN	0.039	0.039	0.052	0.059	0.048	0.065	0.065	0.071
$LPWN$	0.041	0.044	0.057	0.066	0.045	0.061	0.068	0.072
Panel C: $m = \lfloor n^{0.7} \rfloor$								
LPR	0.049	0.055	0.055	0.056	0.049	0.053	0.055	0.056
$TLPR$	0.049	0.056	0.055	0.058	0.051	0.054	0.057	0.059
LW	0.057	0.059	0.070	0.081	0.064	0.061	0.062	0.074
LPW	0.058	0.061	0.076	0.084	0.049	0.051	0.057	0.058
LWN	0.040	0.041	0.052	0.058	0.048	0.056	0.057	0.062
$LPWN$	0.041	0.045	0.056	0.061	0.045	0.053	0.062	0.067

Notes: The table shows the empirical size properties for $\hat{W}_n$ for 5% theoretical size. The number of observations of the original series is 4096 and the aggregated series have $4096/2^{i-1}$ for $i = 1, 2, \ldots, K = 5$ observations.

Table 2: Empirical size properties for Model II with $\mathbb{I} = 1$ and $(\alpha_y, \beta_y 0) = (0, 0)$.

nsr	1				5				10			
K	2	3	4	5	2	3	4	5	2	3	4	5
Panel A: $m = \lfloor n^{0.5} \rfloor$												
LPR	0.051	0.048	0.054	0.052	0.053	0.055	0.064	0.062	0.061	0.068	0.069	0.062
$TLPR$	0.054	0.046	0.055	0.062	0.064	0.059	0.065	0.073	0.067	0.066	0.071	0.072
LW	0.055	0.059	0.060	0.061	0.061	0.064	0.063	0.079	0.064	0.067	0.066	0.075
LPW	0.059	0.056	0.073	0.074	0.059	0.063	0.077	0.083	0.069	0.076	0.081	0.094
LWN	0.037	0.045	0.049	0.055	0.042	0.045	0.051	0.055	0.047	0.049	0.059	0.062
$LPWN$	0.051	0.057	0.059	0.065	0.052	0.061	0.064	0.066	0.051	0.065	0.068	0.071
Panel B: $m = \lfloor n^{0.6} \rfloor$												
LPR	0.059	0.061	0.064	0.063	0.062	0.067	0.075	0.079	0.069	0.078	0.078	0.092
$TLPR$	0.058	0.059	0.061	0.062	0.061	0.067	0.077	0.081	0.069	0.077	0.081	0.089
LW	0.061	0.060	0.065	0.069	0.066	0.068	0.075	0.078	0.071	0.069	0.082	0.089
LPW	0.062	0.064	0.076	0.077	0.069	0.075	0.083	0.089	0.072	0.085	0.099	0.107
LWN	0.046	0.049	0.055	0.056	0.052	0.051	0.055	0.059	0.051	0.053	0.059	0.060
$LPWN$	0.051	0.054	0.058	0.063	0.055	0.053	0.061	0.061	0.055	0.063	0.065	0.069
Panel C: $m = \lfloor n^{0.7} \rfloor$												
LPR	0.089	0.092	0.102	0.108	0.101	0.095	0.124	0.120	0.161	0.145	0.112	0.102
$TLPR$	0.091	0.100	0.099	0.113	0.092	0.112	0.121	0.115	0.159	0.138	0.120	0.111
LW	0.088	0.095	0.097	0.102	0.105	0.121	0.130	0.129	0.162	0.161	0.142	0.101
LPW	0.105	0.114	0.126	0.131	0.133	0.141	0.152	0.149	0.193	0.184	0.152	0.131
LWN	0.053	0.052	0.057	0.058	0.055	0.054	0.059	0.061	0.069	0.061	0.058	0.055
$LPWN$	0.058	0.058	0.060	0.071	0.069	0.071	0.068	0.075	0.081	0.079	0.078	0.064

Notes: The table shows the empirical size properties of $\hat{W}_n$ for 5% theoretical size. The number of observations of the original series is 4096 and the aggregated series have $4096/2^{i-1}$ for $i = 1, 2, \ldots, K = 5$ observations.

Table 3: Empirical size properties for Model III with $\mathbb{I} = 1$ and $(\alpha_y, \beta_y 0) = (0.8, 0)$.

nsr	1				5				10			
K	2	3	4	5	2	3	4	5	2	3	4	5
Panel A: $m = \lfloor n^{0.5} \rfloor$												
LPR	0.062	0.068	0.069	0.073	0.063	0.081	0.098	0.112	0.068	0.071	0.078	0.111
$TLPR$	0.063	0.066	0.071	0.072	0.065	0.080	0.101	0.109	0.067	0.074	0.081	0.116
LW	0.061	0.069	0.064	0.069	0.064	0.078	0.102	0.111	0.065	0.075	0.076	0.114
LPW	0.059	0.061	0.067	0.069	0.070	0.088	0.101	0.119	0.076	0.081	0.081	0.101
LWN	0.049	0.055	0.056	0.059	0.059	0.060	0.062	0.064	0.058	0.059	0.060	0.065
$LPWN$	0.051	0.055	0.054	0.057	0.064	0.063	0.065	0.069	0.061	0.060	0.063	0.070
Panel B: $m = \lfloor n^{0.6} \rfloor$												
LPR	0.053	0.061	0.070	0.078	0.077	0.079	0.105	0.119	0.088	0.100	0.101	0.111
$TLPR$	0.060	0.064	0.071	0.080	0.075	0.081	0.112	0.120	0.088	0.095	0.108	0.117
LW	0.054	0.056	0.070	0.078	0.081	0.083	0.112	0.121	0.086	0.099	0.101	0.114
LPW	0.058	0.058	0.067	0.072	0.085	0.089	0.100	0.119	0.091	0.099	0.105	0.119
LWN	0.054	0.057	0.058	0.061	0.064	0.066	0.066	0.069	0.062	0.062	0.066	0.071
$LPWN$	0.051	0.055	0.055	0.059	0.059	0.058	0.059	0.066	0.060	0.057	0.059	0.068
Panel C: $m = \lfloor n^{0.7} \rfloor$												
LPR	0.114	0.112	0.110	0.113	0.179	0.161	0.132	0.136	0.152	0.102	0.108	0.127
$TLPR$	0.109	0.111	0.125	0.119	0.157	0.162	0.135	0.130	0.170	0.111	0.109	0.129
LW	0.113	0.107	0.104	0.117	0.181	0.144	0.133	0.139	0.163	0.119	0.110	0.123
LPW	0.089	0.078	0.079	0.082	0.141	0.120	0.117	0.120	0.121	0.135	0.129	0.101
LWN	0.059	0.061	0.064	0.063	0.076	0.071	0.065	0.066	0.089	0.079	0.074	0.071
$LPWN$	0.055	0.053	0.054	0.061	0.058	0.055	0.060	0.064	0.071	0.055	0.059	0.063

Notes: The table shows the empirical size properties of $\hat{W}_n$ for 5% theoretical size. The number of observations of the original series is 4096 and the aggregated series have $4096/2^{i-1}$ for $i = 1, 2, \ldots, K = 5$ observations.

Table 4: The power against the nonstationary random level shift (NSRLS), the stationary random level shift (SRLS), and white noise with trend (WNT) models with $m = \lfloor n^{0.5} \rfloor$.

	Generic Parameter Values				Estimated Parameter Values			
	Panel A: NSRLS							
K	2	3	4	5	2	3	4	5
LPR	**94.0%**	**96.5%**	97.6%	98.0%	93.4%	93.2%	93.1%	94.3%
TLPR	99.0%	99.3%	98.7%	98.5%	93.2%	94.1%	93.7%	94.1%
LW	94.2%	95.7%	**97.3%**	**97.4%**	90.4%	91.2%	91.0%	91.4%
LPW	97.9%	98.0%	98.0%	98.3%	**89.5%**	**89.9%**	**90.6%**	**91.1%**
LWN	100.0%	100.0%	100.0%	100.0%	99.5%	99.7%	99.8%	99.8%
LPWN	98.5%	99.0%	99.8%	100.0%	99.5%	99.5%	100.0%	100.0%
	Panel B: SRLS							
K	2	3	4	5	2	3	4	5
LPR	**95.6%**	**98.1%**	**98.7%**	99.4%	97.3%	**98.7%**	**99.5%**	**99.3%**
TLPR	99.9%	99.8%	99.6%	99.3%	100.0%	99.7%	99.5%	99.3%
LW	96.1%	98.6%	99.3%	**99.3%**	**97.2%**	98.9%	99.5%	99.8%
LPW	99.1%	99.8%	99.3%	99.5%	99.2%	99.6%	99.9%	99.9%
LWN	100.0%	100.0%	100.0%	99.9%	100.0%	100.0%	100.0%	100.0%
LPWN	99.0%	99.3%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
	Panel C: WNT							
K	2	3	4	5	2	3	4	5
LPR	**95.8%**	98.4%	99.5%	99.6%	77.8%	87.7%	91.6%	96.1%
TLPR	99.8%	99.9%	100.0%	99.6%	98.3%	98.9%	98.0%	98.0%
LW	96.3%	**98.1%**	**98.8%**	**99.3%**	89.3%	91.6%	95.6%	98.9%
LPW	99.3%	99.5%	100.0%	99.7%	**75.5%**	**80.6%**	**83.4%**	**86.6%**
LWN	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
LPWN	100.0%	99.6%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Notes: The table shows the power properties of $\hat{W}_n$ against nonstationary random level shift (Panel A), stationary random level shift (Panel B), and white noise with trend (Panel C) models. The number of observations of the original series is 4096 and the aggregated series have $4096/2^{i-1}$ for $i = 1, 2, \ldots, K = 5$ observations.

Table 5: The power against the Markov switching models with iid (MS-IID) and GARCH (MS-GARCH) regimes with $m = \lfloor n^{0.5} \rfloor$.

	Generic Parameter Values				Estimated Parameter Values			
	Panel A: MS-IID							
K	2	3	4	5	2	3	4	5
LPR	**93.9%**	96.9%	96.3%	97.0%	94.8%	97.2%	97.8%	**97.7%**
TLPR	95.7%	**95.7%**	96.6%	97.1%	99.1%	99.5%	98.9%	98.9%
LW	94.2%	96.6%	**96.1%**	**96.1%**	**94.1%**	**96.7%**	**97.4%**	97.9%
LPW	96.4%	96.9%	98.8%	99.1%	98.6%	99.1%	99.0%	98.7%
LWN	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
LPWN	100.0%	100.0%	99.6%	100.0%	100.0%	99.6%	100.0%	99.6%
	Panel B: MS-GARCH							
K	2	3	4	5	2	3	4	5
LPR	97.9%	98.0%	98.0%	98.0%	95.4%	**97.8%**	99.2%	99.5%
TLPR	97.7%	97.7%	97.5%	97.7%	99.6%	100.0%	99.5%	99.6%
LW	95.3%	95.3%	96.7%	96.7%	**94.8%**	97.9%	**98.7%**	**99.0%**
LPW	**94.7%**	**94.6%**	**95.0%**	**95.5%**	99.4%	99.6%	99.7%	99.8%
LWN	100.0%	99.8%	99.8%	99.9%	100.0%	100.0%	100.0%	100.0%
LPWN	99.9%	99.5%	99.9%	99.9%	100.0%	100.0%	100.0%	100.0%

Notes: The table shows the power properties of $\hat{W}_n$ against Markov switching models with iid (Panel A) and GARCH (Panel B) regimes. The number of observations of the original series is 4096 and the aggregated series have $4096/2^{i-1}$ for $i = 1, 2, \ldots, K = 5$ observations.

Table 6: Empirical results for the DEM/USD, YEN/USD, and USD/GBP exchange rates with $m = \lfloor n^{0.5} \rfloor$.

	LPR		Long Memory Estimates				TLPR		Long Memory Estimates			
Currency	$\hat{W}$	*p-values*	$k=1$	$k=2$	$k=3$	$k=4$	$\hat{W}$	*p-values*	$k=1$	$k=2$	$k=3$	$k=4$
DEM/USD	23.02	0.00	0.47	0.35	0.37	-0.02	9.47	0.02	0.41	0.22	0.55	0.16
YEN/USD	9.14	0.02	0.24	0.26	0.45	0.31	0.92	0.81	0.45	0.54	0.64	0.58
USD/GBP	2.44	0.48	0.40	0.44	0.49	0.37	2.44	0.48	0.41	0.42	0.47	0.38
	LW		Long Memory Estimates				LPW		Long Memory Estimates			
Currency	$\hat{W}$	*p-values*	$k=1$	$k=2$	$k=3$	$k=4$	$\hat{W}$	*p-values*	$k=1$	$k=2$	$k=3$	$k=4$
DEM/USD	18.80	0.00	0.39	0.27	0.32	0.08	4.46	0.21	0.15	0.22	0.26	0.02
YEN/USD	5.02	0.16	0.30	0.30	0.38	0.26	0.61	0.89	0.30	0.28	0.21	0.16
USD/GBP	1.58	0.66	0.32	0.36	0.40	0.38	1.05	0.78	0.43	0.36	0.37	0.45
	LWN		Long Memory Estimates				LPWN		Long Memory Estimates			
Currency	$\hat{W}$	*p-values*	$k=1$	$k=2$	$k=3$	$k=4$	$\hat{W}$	*p-values*	$k=1$	$k=2$	$k=3$	$k=4$
DEM/USD	0.46	0.92	0.32	0.15	0.17	0.01	0.02	0.99	0.41	0.28	0.35	0.01
YEN/USD	2.84	0.41	0.28	0.24	0.41	0.01	0.22	0.97	0.16	0.24	0.07	0.31
USD/GBP	0.82	0.84	0.53	0.44	0.51	0.57	0.25	0.96	0.28	0.73	0.73	0.74

Note: The LPWN denotes the local polynomial Whittle estimator implemented with $(R_y, R_w) = (1, 0)$.

Table 7: Empirical results for the DEM/USD, YEN/USD, and USD/GBP exchange rates with $m = \lfloor n^{0.7} \rfloor$.

	LPR		Long Memory Estimates				TLPR		Long Memory Estimates			
Currency	$\hat{W}$	*p-values*	$k=1$	$k=2$	$k=3$	$k=4$	$\hat{W}$	*p-values*	$k=1$	$k=2$	$k=3$	$k=4$
DEM/USD	30.58	0.00	0.24	0.34	0.37	0.18	14.69	0.00	0.21	0.35	0.42	0.20
YEN/USD	16.15	0.00	0.18	0.26	0.36	0.37	5.77	0.12	0.17	0.28	0.35	0.31
USD/GBP	7.08	0.07	0.25	0.32	0.32	0.34	8.14	0.04	0.27	0.38	0.29	0.17

	LW		Long Memory Estimates				LPW		Long Memory Estimates			
Currency	$\hat{W}$	*p-values*	$k=1$	$k=2$	$k=3$	$k=4$	$\hat{W}$	*p-values*	$k=1$	$k=2$	$k=3$	$k=4$
DEM/USD	13.87	0.00	0.22	0.30	0.32	0.28	9.69	0.02	0.32	0.36	0.35	0.17
YEN/USD	14.63	0.00	0.23	0.29	0.33	0.41	2.99	0.39	0.29	0.32	0.31	0.22
USD/GBP	11.63	0.01	0.24	0.30	0.31	0.38	1.76	0.62	0.31	0.30	0.36	0.33

	LWN		Long Memory Estimates				LPWN		Long Memory Estimates			
Currency	$\hat{W}$	*p-values*	$k=1$	$k=2$	$k=3$	$k=4$	$\hat{W}$	*p-values*	$k=1$	$k=2$	$k=3$	$k=4$
DEM/USD	7.10	0.07	0.48	0.38	0.38	0.37	0.37	0.94	0.36	0.32	0.35	0.19
YEN/USD	7.57	0.06	0.41	0.33	0.30	0.49	5.11	0.16	0.33	0.18	0.38	0.24
USD/GBP	4.12	0.24	0.44	0.36	0.45	0.42	0.69	0.87	0.38	0.51	0.50	0.54

Note: The LPWN denotes the local polynomial Whittle estimator implemented with $(R_y, R_w) = (1, 0)$.

Table 8: Empirical results for the DJIA30 stocks with $m = \lfloor n^{0.5} \rfloor$.

	LPR		Long Memory Estimates				LWN		Long Memory Estimates			
Ticker	$\hat{W}$	*p-values*	$k=1$	$k=2$	$k=3$	$k=4$	$\hat{W}$	*p-values*	$k=1$	$k=2$	$k=3$	$k=4$
AA	4.05	0.26	0.42	0.50	0.39	0.32	0.84	0.84	0.55	0.60	0.74	0.58
AIG	4.22	0.24	0.42	0.50	0.39	0.38	0.92	0.82	0.66	0.68	0.75	0.56
AXP	5.59	0.13	0.53	0.51	0.35	0.47	0.01	1.00	0.75	0.73	0.74	0.73
BA	**15.89**	0.00	0.34	0.42	0.28	0.60	2.65	0.44	0.74	0.57	0.58	0.50
C	7.55	0.06	0.53	0.51	0.32	0.21	**14.80**	0.00	0.66	0.76	0.77	0.01
CAT	**17.71**	0.00	0.28	0.43	0.40	0.73	1.87	0.59	0.72	0.52	0.42	0.49
DD	1.40	0.70	0.24	0.21	0.18	0.08	0.46	0.92	0.28	0.33	0.54	0.08
DIS	**13.33**	0.00	0.49	0.49	0.27	0.47	0.23	0.97	0.68	0.78	0.80	0.47
GE	**15.08**	1.00	0.55	0.55	0.27	0.41	**10.86**	0.01	0.62	0.88	0.81	0.31
GM	5.69	0.13	0.43	0.43	0.30	0.14	**23.43**	0.00	0.41	0.28	0.94	0.97
HD	5.48	0.14	0.53	0.60	0.44	0.41	1.20	0.75	0.61	0.56	0.45	0.68
HON	**11.80**	0.01	0.33	0.50	0.46	0.24	1.51	0.68	0.72	0.81	0.75	0.74
HPQ	2.18	0.53	0.50	0.52	0.42	0.47	0.34	0.95	0.72	0.71	0.82	0.74
IBM	2.37	0.49	0.42	0.51	0.50	0.44	0.72	0.87	0.67	0.73	0.80	0.66
INTC	5.21	0.15	0.47	0.54	0.39	0.37	1.32	0.72	0.71	0.71	0.77	0.32
JNJ	**9.51**	0.02	0.45	0.49	0.29	0.42	3.52	0.31	0.61	0.61	0.34	0.46
JPM	**19.04**	0.00	0.56	0.51	0.23	0.45	6.81	0.07	0.54	0.71	0.72	0.01
KO	7.73	0.06	0.46	0.52	0.49	0.23	0.29	0.96	0.61	0.69	0.77	0.78
MCD	**11.89**	0.00	0.49	0.57	0.34	0.44	4.08	0.25	0.44	0.45	0.74	0.42
MMM	7.20	0.07	0.27	0.40	0.28	0.30	0.74	0.86	0.56	0.64	0.68	0.65
MO	1.99	0.57	0.35	0.40	0.35	0.27	1.13	0.77	0.66	0.62	0.64	0.73
MRK	5.88	0.11	0.29	0.40	0.39	0.22	0.98	0.80	0.56	0.52	0.35	0.20
MSFT	5.90	0.12	0.55	0.55	0.37	0.33	2.55	0.46	0.68	0.67	0.18	0.01
PFE	4.48	0.21	0.36	0.37	0.47	0.30	0.29	0.96	0.64	0.64	0.56	0.75
PG	7.63	0.06	0.44	0.40	0.39	0.11	2.52	0.47	0.59	0.74	0.85	0.66
SBC	3.43	0.33	0.49	0.45	0.32	0.28	7.03	0.07	0.61	0.66	0.76	0.75
UTX	**8.03**	0.05	0.41	0.45	0.31	0.11	1.25	0.74	0.66	0.57	0.63	0.59
VZ	7.63	0.06	0.46	0.63	0.49	0.43	3.97	0.26	0.92	0.88	0.58	0.41
WMT	6.33	0.10	0.59	0.64	0.47	0.39	5.04	0.16	0.71	0.63	0.31	0.31
XOM	**11.83**	0.01	0.37	0.41	0.21	0.41	1.02	0.79	0.49	0.47	0.88	0.91

Table 9: The Simulated Spurious Long Memory Models.

	Generic Parameter Values	Estimated Parameter Values
NSRLS	$\sigma^2_\varepsilon = 5$, $\sigma^2_\eta = 1$, $p = 0.00001$	$\sigma^2_\varepsilon = 4.4$, $\sigma^2_\eta = 2.3$, $p = 0.000004$
SRLS	$\sigma^2_\varepsilon = \sigma^2_\eta = 1$, $p = 0.003$	$\sigma^2_\varepsilon = 2.2$, $\sigma^2_\eta = 5.7$, $p = 0.002$
MS-IID	$\mu_0 = -\ \mu_1 = 1$, $\sigma^2 = 1$, $p_0 = p_1 = 0.001$	$\mu_0 = -0.4$, $\mu_1 = -4$, $\sigma^2 = 2.5$, $p_0 = p_1 = 0.001$
MS-GARCH	$\omega_0 = 1$, $\omega_1 = 3$, $\alpha = 0.4$, $\beta = 0.3$, $p = 0.001$	$\omega_0 = 0.08$, $\omega_1 = 0.23$, $\alpha = 0.28$, $\beta = 0.56$, $p = 0.0008$
WNT	$c = 3$, $\beta = -0.1$, $\sigma^2_u = 1$	$c = -4.3$, $\beta = -0.063$, $\sigma^2_u = 8.1$

Notes: The table displays the generic and empirical parameter values obtained from the foreign exchange rate analysis in Ohanissian et al. (2008).

Research Papers 2008

CREATES

Center for Research in Econometric Analysis of Time Series

2008-43: Martin Møller Andreasen: Explaining Macroeconomic and Term Structure Dynamics Jointly in a Non-linear DSGE Model

2008-44: Christian M. Dahl, Henrik Hansen and John Smidt: The cyclical component factor model

2008-45: Christian M. Dahl and Yu Qin: The limiting behavior of the estimated parameters in a misspecified random field regression model

2008-46: Bent Jesper Christensen, Christian M. Dahl and Emma M. Iglesias: Semiparametric Inference in a GARCH-in-Mean Model

2008-47: Charlotte Christiansen: Mean Reversion in US and International Short Rates

2008-48: Tim Bollerslev, George Tauchen and Hao Zhou: Expected Stock Returns and Variance Risk Premia

2008-49: Tim Bollerslev: Glossary to ARCH (GARCH)

2008-50: Giuseppe Cavaliere, Anders Rahbek and A.M.Robert Taylor: Testing for Co-integration in Vector Autoregressions with Non-Stationary Volatility

2008-51: Bent Jesper Christensen and Michael Sørensen: Optimal inference in dynamic models with conditional moment restrictions

2008-52: Katarzyna Lasak: Likelihood based testing for no fractional cointegration

2008-53: Katarzyna Lasak: Maximum likelihood estimation of fractionally cointegrated systems

2008-54: Andrew J. Patton and Allan Timmermann: The Resolution of Macroeconomic Uncertainty: Evidence from Survey Forecast

2008-55: Carlos Capistrán and Allan Timmermann: Forecast Combination With Entry and Exit of Experts

2008-56: Carlos Capistrán and Allan Timmermann: Disagreement and Biases in Inflation Expectations

2008-57: Almut E. D. Veraart: Impact of time–inhomogeneous jumps and leverage type effects on returns and realised variances

2008-58: Dennis Kristensen and Yongseok Shin: Estimation of Dynamic Models with Nonparametric Simulated Maximum Likelihood

2008-59: Per Frederiksen and Frank S. Nielsen: Testing for long memory in potentially nonstationary perturbed fractional processes