



Fast computation and practical use of amplitudes at non-Fourier frequencies

Erhard Reschenhofer¹ · Manveer K. Mangat¹

Received: 14 April 2020 / Accepted: 19 December 2020 / Published online: 3 February 2021
© The Author(s) 2021

Abstract

In this paper, it is shown that the performance of various frequency-domain estimators of the memory parameter can be boosted by the inclusion of non-Fourier frequencies in addition to the regular Fourier frequencies. A fast two-stage algorithm for the efficient computation of the amplitudes at these additional frequencies is presented. In the first stage, the naïve sine and cosine transforms are computed with a modified version of the Fast Fourier Transform. In the second stage, these transforms are amended by taking the violation of the standard orthogonality conditions into account. A considerable number of auxiliary quantities, which are required in the second stage, do not depend on the data and therefore only need to be computed once. The superior performance (in terms of root-mean-square error) of the estimators based also on non-Fourier frequencies is demonstrated by extensive simulations. Finally, the empirical results obtained by applying these estimators to financial high-frequency data show that significant long-range dependence is present only in the absolute intraday returns but not in the signed intraday returns.

Keywords Long-range dependence · Periodogram · Fractional Fourier frequencies · Financial high-frequency data

1 Introduction

Long memory refers to statistical dependence between observations far apart in time. Indications of long memory have been found in various fields such as hydrology, meteorology, geophysics, psychology, economics, and finance (for surveys see, e.g., Beran 1994; Graves et al. 2017). A popular measure of the degree of long-range dependence is the memory parameter d , which can be defined for a given stationary time series in terms of the rate of decay of its autocorrelation function

✉ Erhard Reschenhofer
erhard.reschenhofer@univie.ac.at

¹ Department of Statistics and Operations Research, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria

or the steepness of its spectral density near frequency zero. The most widely used long memory models are the autoregressive fractionally integrated moving average (ARFIMA) models. Pötscher (2002) showed that drawing inferences about d is only possible when severely restricted subsets of ARFIMA models are used. Unfortunately, restrictions of this type are not plausible in most applications, with series of stock returns being notable exceptions because of the absence of significant short-term autocorrelation.

A popular semiparametric method of estimating d , which allows to estimate d without simultaneously estimating the other model parameters, is based on the log periodogram. Since the periodogram is a very erratic estimator for the spectral density, it seems natural to smooth it prior to its use in the estimation of the memory parameter d (Hassler 1993; Peiris and Court 1993; Reisen 1994). While smoothing has the expected positive effect on the variance of the estimator, it may also increase the bias (Chen et al. 1994; Reschenhofer et al. 2020), which is especially problematic when individual estimates are first obtained with a rolling estimation window and then combined by averaging. In this paper, we take an opposite approach. Instead of reducing the resolution through smoothing, we increase it by doubling the number of frequencies. However, there are two problems. First, the Fast Fourier Transform allows only the fast computation of the periodogram at the Fourier frequencies. Second, the periodogram ordinates can in the case of non-Fourier frequencies no longer be interpreted as the squared amplitudes of simple sinusoids. We address the latter problem by modifying the periodogram and the former by introducing a two-stage algorithm for the fast computation of the modified periodogram. In the first stage, a modified version of the Fast Fourier Transform is used for the computation of the standard periodogram at non-standard frequencies, which is subsequently subjected to a simple and efficient transformation in order to rectify the distortions caused by the violation of the standard orthogonality conditions in the case of non-Fourier frequencies. Both a simulation study and an empirical study with financial high-frequency data are carried out to verify the relevance of our approach.

The paper is structured as follows. Section 2 briefly reviews various frequency-domain estimators for the memory parameter d and explains how they can be modified by the inclusion of non-Fourier frequencies. In Sect. 3, an efficient algorithm for the computation of the modified estimators is described. Sections 4 and 5 present the results of the simulation study and the empirical study, respectively. Section 6 concludes.

2 Frequency-domain estimation of the memory parameter

A parametric approach (see Fox and Taqqu 1986) to estimate the memory parameter (or fractional differencing parameter) d of the autoregressive fractionally integrated moving average (ARFIMA) model

$$y_t = (1 - \phi_1 L - \dots - \phi_p L^p)^{-1} (1 - L)^{-d} (1 - \theta_1 L - \dots - \theta_q L^q) u_t \quad (1)$$

(Granger and Joyeux 1980; Hosking 1981) with spectral density

$$f(\omega) = \frac{\sigma^2}{2\pi} \left| 1 - e^{-i\omega} \right|^{-2d} \left| 1 - \sum_{j=1}^q \theta_j e^{-i\omega j} \right|^2 \left| 1 - \sum_{j=1}^p \phi_j e^{-i\omega j} \right|^{-2} \quad (2)$$

only makes sense when we are sure that the parametric model is correctly specified (see, e.g., Robinson 1995; Reschenhofer 2013). The uncertainty regarding the order (p, q) of the ARMA model can be avoided by adopting Geweke and Porter-Hudak's (1983) approach. Their semiparametric estimator \hat{d}_{GPH} is obtained as the slope of a simple linear regression which explains the log periodogram

$$\log I(\omega_k) = \log \frac{1}{2\pi n} \left| \sum_{t=1}^n y_t e^{-i\omega_k t} \right|^2 \quad (3)$$

of the observations y_1, \dots, y_n by the deterministic regressor

$$-\log \left| 1 - e^{-i\omega_j} \right|^2 = -2 \log (2 \sin (\omega_j/2)), \quad (4)$$

where

$$\omega_j = 2\pi j/n, \quad j = 1, \dots, K \ll \lfloor n/2 \rfloor, \quad (5)$$

are the first K Fourier frequencies. Hurvich et al. (1998) showed that this estimator is consistent when only Fourier frequencies in the neighborhood of frequency zero are used, more precisely when $K = o(n^{4/5})$ and $\log^2(n) = o(K)$.

Because of the irregular behavior of the spectral density (2) in the neighborhood of frequency zero, the standard asymptotic results for the normalized periodogram ordinates $J(\omega_j) = I(\omega_j)/f(\omega_j)$ do not hold when the frequencies ω_j are too low. However, Künsch (1986) was able to show that $J(\omega_j)$, $j = H + 1, \dots, H + K$, are still asymptotically i.i.d. standard exponential provided that $(H + 1)/\sqrt{n} \rightarrow \infty$ and $(H + K)/n \rightarrow 0$. Robinson (1995) therefore proposed to remove the very lowest Fourier frequencies from the log periodogram regression. The most conservative variant \hat{d}_{tr} of Robinson's estimator is obtained by trimming out only the contribution from the lowest frequency. Clearly, removing further frequencies will only be an option if the removal of the first does not already lead to a deterioration in the estimation properties.

Another modification of the log periodogram regression is based on smoothing. Hassler (1993), Peiris and Court (1993), Reisen (1994) proposed to smooth the periodogram before it is used for the estimation of the memory parameter d . This can be achieved by cutting off the sum in the representation

$$I(\omega_j) = \frac{1}{2\pi} \sum_{s=-(n-1)}^{n-1} \hat{\gamma}(s) e^{-i\omega_j s} \quad (6)$$

at a much smaller value than $n - 1$ and downweighting the higher-order autocovariances $\hat{\gamma}(s)$ with a suitable lag window such as the Bartlett window or the Parzen window. In general, the resulting estimators have typically a smaller variance and a

larger bias than those based on the raw periodogram (Chen et al. 1994; Reschenhofer et al. 2020). However, the squared bias is not always small relative to the variance. For example, in the case of high-frequency financial data, preliminary estimates of the memory parameter d must first be obtained separately from the different trading sessions because of the overnight gaps and then be combined by averaging. Thus, the variance of the global estimator decreases with the number of trading sessions but the bias remains fixed. Estimators with a large bias are therefore of no use. We may keep the bias small by using an estimator based on minimal smoothing, e.g.,

$$\tilde{I}(\omega_j) = (I(\omega_{j-1}) + I(\omega_j) + I(\omega_{j+1}))/3, \quad (7)$$

or by using a more sophisticated smoothing method that does not systematically increase the bias. Reschenhofer and Mangat (2020) managed to achieve a reduction in the variance which does not come at the expense of an increase in the bias. Their estimator is based on running a log periodogram regression repeatedly for different partitions of the data. In the simplest case, only the whole sample and the first and second halves are used but the frequency range $(0, \omega_K]$, is kept constant, which implies that reducing the sample size by half results in half the number of included Fourier frequencies. The simple estimator

$$\hat{d}_2 = \frac{1}{2}\hat{d}_1 + \frac{1}{4}\hat{d}_{21} + \frac{1}{4}\hat{d}_{22}, \quad (8)$$

where $\hat{d}_1, \hat{d}_{21}, \hat{d}_{22}$ are the OLS estimators for d based on the log periodograms L^1, L^{21}, L^{22} of the whole sample and the first and second halves, respectively, can easily be extended to the more general estimator

$$\hat{d}_k = \frac{1}{k} \left(\hat{d}_1 + \sum_{j=2}^k \frac{1}{j} (\hat{d}_{j1} + \dots + \hat{d}_{jj}) \right), \quad (9)$$

which is based on k partitions.

Reschenhofer (2013) observed that the root mean square error (RMSE), which is defined as the square root of the average of squared errors, of Geweke and Porter-Hudak's (1983) semiparametric estimator \hat{d}_{GPH} can be reduced significantly by including additional frequencies. However, he did not examine the effect on other estimators and he also did not address the problem of efficiently computing the periodogram at non-Fourier frequencies.

For an arbitrary frequency ω_k , where k is not necessarily an integer, it is more appropriate to define the periodogram by

$$I_k = \frac{n}{8\pi} (\hat{A}_k^2 + \hat{B}_k^2), \quad (10)$$

where

$$\hat{R}_k^2 = \hat{A}_k^2 + \hat{B}_k^2 \quad (11)$$

is the least squares (LS) estimate of the squared amplitude of a (normalized) sinusoid

$$\begin{aligned} y_t &= \mu + R_k \sin(\omega_k t + \varphi) + u_t \\ &= \mu + \underbrace{R_k \sin(\varphi)}_{A_k} \cos(\omega_k t) + \underbrace{R_k \cos(\varphi)}_{B_k} \sin(\omega_k t) + u_t \end{aligned} \quad (12)$$

fitted to the data. If

$$\sum_{t=1}^n \cos(\omega_k t) = \sum_{t=1}^n \sin(\omega_k t) = 0, \quad (13)$$

$$\sum_{t=1}^n \cos(\omega_k t) \sin(\omega_k t) = 0, \quad (14)$$

and

$$\sum_{t=1}^n \cos^2(\omega_k t) = \sum_{t=1}^n \sin^2(\omega_k t) = n/2, \quad (15)$$

then the LS estimates \hat{A}_k, \hat{B}_k simplify to

$$\tilde{A}_k = \frac{2}{n} \sum_{t=1}^n y_t \cos(\omega_k t), \quad \tilde{B}_k = \frac{2}{n} \sum_{t=1}^n y_t \sin(\omega_k t) \quad (16)$$

or, because of (13), to

$$\tilde{A}_k = \frac{2}{n} \sum_{t=1}^n \underbrace{(y_t - \bar{y})}_{\underline{y}_t} \cos(\omega_k t), \quad \tilde{B}_k = \frac{2}{n} \sum_{t=1}^n \underline{y}_t \sin(\omega_k t). \quad (17)$$

The definition (10) therefore matches the conventional definition of the periodogram. The conditions (13), (14), (15) are satisfied for the regular Fourier frequencies $\omega_k = 2\pi k/n$, $1 \leq k \leq m = \lfloor n/2 \rfloor$ (except the last one if n is an even number), hence \hat{A}_k and \hat{B}_k can easily be obtained from

$$\tilde{A}_k = \frac{2}{n} \operatorname{Re}(\tilde{C}_k), \quad \tilde{B}_k = -\frac{2}{n} \operatorname{Im}(\tilde{C}_k), \quad (18)$$

where

$$\tilde{C}_k = \sum_{t=1}^n \underline{y}_t \exp(-i\omega_k t), \quad k = 1, \dots, m = \lfloor n/2 \rfloor, \quad (19)$$

is the Discrete Fourier Transform (DFT) of y_1, \dots, y_n , which can be efficiently computed with the Fast Fourier Transform (FFT; see Cooley and Tukey 1965).

Since condition (13) is not satisfied for the fractional Fourier frequencies $\omega_k = 2\pi k/n, k = 1/2, 3/2, \dots, m - 1/2$, we introduce a two-stage algorithm for the efficient computation of the modified periodogram (10) at these frequencies in the next section. This algorithm allows to carry out an extensive simulation study in order to examine the effect of including additional frequencies in the log periodogram regression on the performance of various estimators. In Sect. 5, we compare the standard estimator \hat{d}_{GPH} based on I_1, I_2, \dots, I_K with the analogous estimator \hat{d}_{GPH}^+ based on $I_1, I_{1.5}, \dots, I_K$, the estimator \hat{d}_r based on I_2, I_3, \dots, I_K with the estimator \hat{d}_r^+ based on $I_2, I_{2.5}, \dots, I_K$, the estimator \hat{d}_{sm} based on $(I_1 + I_2 + I_3)/3, (I_2 + I_3 + I_4)/3, \dots, (I_{K-2} + I_{K-1} + I_K)/3$ with the estimator \hat{d}_{sm}^+ based on $(I_1 + I_{1.5} + I_2)/3, (I_{1.5} + I_2 + I_{2.5})/3, \dots, (I_{K-1} + I_{K-0.5} + I_K)/3$, and the estimator \hat{d}_2 with the estimator

$$\hat{d}_2^+ = \frac{1}{2}\hat{d}_1^+ + \frac{1}{4}\hat{d}_{21}^+ + \frac{1}{4}\hat{d}_{22}^+. \tag{20}$$

3 Algorithm

In case of the regular Fourier frequencies $\omega_k = 2\pi k/n, 1 \leq k \leq m = [n/2]$, \hat{A}_k and \hat{B}_k are identical to \tilde{A}_k and \tilde{B}_k and can therefore be computed with the Fast Fourier Transform. For the fractional Fourier frequencies $\omega_k = 2\pi k/n, k = 1/2, 3/2, \dots, m - 1/2$, an analogous approach can be used for the computation of \tilde{A}_k and \tilde{B}_k . Suppose that n is even. Then

$$\begin{aligned} \tilde{C}_k &= \sum_{t=1}^n y_{-t} \exp(-i\omega_k t) \\ &= \sum_{t=1}^m y_{-2t-1} \exp(-i\omega_k(2t-1)) + \sum_{t=1}^m y_{-2t} \exp(-i\omega_k 2t) \\ &= \exp(i\omega_k) \underbrace{\sum_{t=1}^m y_{-2t-1} \exp(-i\omega_k 2t)}_{\tilde{C}_k(1)} + \underbrace{\sum_{t=1}^m y_{-2t} \exp(-i\omega_k 2t)}_{\tilde{C}_k(2)}, \end{aligned} \tag{21}$$

where $\tilde{C}_k(1)$ and $\tilde{C}_k(2)$ need to be computed only for $2k \leq m$ because

$$\exp(-i\omega_{n-2k}t) = \exp(2\pi - i\omega_{2k}t) = \exp(-i\omega_{2k}t). \tag{22}$$

Thus, the number of required calculations is approximately reduced by half. If n is divisible by 4 or, in the best case, is even a power of 2, this technique can be applied repeatedly. Fortunately, by appending zeroes to the end of the time series (zero padding), it can always be achieved that n is a power of 2.

However, in case of the fractional Fourier frequencies, we obtain with the help of standard formulas for sums of trigonometric functions (see Gradshteyn and Ryzhik 2007, p. 37)

$$\sum_{t=1}^n \cos(\omega_k t) = \cos(\omega_k(n+1)/2) \sin(\omega_k n/2) / \sin(\omega_k/2) = -1 \tag{23}$$

and

$$\sum_{t=1}^n \sin(\omega_k t) = \sin(\omega_k(n+1)/2) \sin(\omega_k n/2) / \sin(\omega_k/2) = \cot(\omega_k/2), \tag{24}$$

hence condition (13) is violated and we must therefore still try to get the values of \hat{A}_k and \hat{B}_k in an efficient manner from \tilde{C}_k or, equivalently, from \tilde{A}_k and \tilde{B}_k . Eliminating the constant from (12) gives

$$y_{-t} = \underbrace{A_k \left(\cos(\omega_k t) + \frac{1}{n} \right)}_{\cos(\omega_k t)} + B_k \underbrace{\left(\sin(\omega_k t) - \overbrace{\cot(\omega_k/2)/n}^{\tau_k} \right)}_{\sin(\omega_k t)} + u_t. \tag{25}$$

Thus,

$$\begin{aligned} \hat{A}_k &= \left(\sum_{t=1}^n \sin^2(\omega_k t) \sum_{t=1}^n \cos(\omega_k t) y_{-t} - \sum_{t=1}^n \cos(\omega_k t) \sin(\omega_k t) \sum_{t=1}^n \sin(\omega_k t) y_{-t} \right) / D_k \\ &= \left(\frac{n^2}{2} \left(\frac{1}{2} - \tau_k^2 \right) \tilde{A}_k - \frac{n^2}{2} \frac{\tau_k}{n} \tilde{B}_k \right) / D_k \\ &= \tilde{A}_k + \frac{2}{n^2(1 - 2\tau_k^2) - 2} (\tilde{A}_k - n\tau_k \tilde{B}_k) \end{aligned} \tag{26}$$

and

$$\begin{aligned} \hat{B}_k &= \left(\sum_{t=1}^n \cos^2(\omega_k t) \sum_{t=1}^n \sin(\omega_k t) y_{-t} - \sum_{t=1}^n \cos(\omega_k t) \sin(\omega_k t) \sum_{t=1}^n \cos(\omega_k t) y_{-t} \right) / D_k \\ &= \left(\frac{n^2}{2} \left(\frac{1}{2} - \frac{1}{n^2} \right) \tilde{B}_k - \frac{n^2}{2} \frac{\tau_k}{n} \tilde{A}_k \right) / D_k \\ &= \tilde{B}_k + \frac{2}{n^2(1 - 2\tau_k^2) - 2} (n^2 \tau_k^2 \tilde{B}_k - n\tau_k \tilde{A}_k), \end{aligned} \tag{27}$$

where the auxiliary quantities

$$\begin{aligned} D_k &= \sum_{t=1}^n \cos^2(\omega_k t) \sum_{t=1}^n \sin^2(\omega_k t) y_{-t} - \left(\sum_{t=1}^n \cos(\omega_k t) \sin(\omega_k t) \right)^2 \\ &= \frac{n^2}{2} \left(\frac{1}{2} - \tau_k^2 - \frac{1}{n^2} \right) \end{aligned} \tag{28}$$

and

$$\tau_k^2 = (\cot(\pi k/n)/n)^2 \sim (\pi k)^{-2} \quad (29)$$

do not depend on the data and therefore only need to be computed once. After the computation of \tilde{C}_k , \tilde{A}_k , and \tilde{B}_k with the modified Fast Fourier Transform in the first stage, the quantities of interest, \hat{A}_k and \hat{B}_k , can easily be obtained with the formulas (26) and (27) in the second stage.

4 Simulations

To investigate the bias-variance tradeoff for different estimators of the memory parameter d , we carry out an extensive simulation study. In general, one would expect that the bias decreases and the variance increases as the complexity of the estimation procedure increases. However, things get more complicated when more sophisticated procedures are involved. E.g., Geman et al. (1992) found evidence for the bias-variance tradeoff in nonparametric procedures such as the k -nearest neighbors algorithm and kernel regression whereas Neal et al. (2020) found in the context of neural networks that both bias and variance can decrease as the number of parameters grows. In the case of the log periodogram regression, the number of parameters is of dubious relevance. The performance of this procedure critically depends on the highest frequency used. Smoothing the periodogram before carrying out the regression requires the specification of an additional parameter, namely the cut-off point of the lag window. Thus, smoothing increases the number of parameters but at the same time also reduces the spectral resolution and thereby increases the bias and decreases the variance rather than the other way round. When dealing with the bias-variance tradeoff, the complexity of a procedure should therefore not be judged solely on the basis of the number of parameters. More important is the flexibility to fit the data. Overfitting typically tends to increase the variance while underfitting tends to increase the bias. Since the periodogram may be regarded as a representation of the raw data in the frequency domain, it provides a perfect fit. Clearly, smoothing compromises this fit. On the other hand, it is also not possible to improve the perfect fit by the introduction of additional periodogram ordinates at non-Fourier frequencies. These new ordinates are, in a certain sense, just interpolations of the original periodogram ordinates. Their inclusion therefore brings about some kind of minimal local smoothing which has the big advantage that it does not systematically have a negative impact on the bias. This advantage is particularly important when estimates obtained from separate samples of fixed size are combined by averaging because in this case only the variance decreases as the number of samples increases but the bias remains fixed.

With the help of the R-package ‘fracdiff’, 40,000 realizations of length $n = 390,3000$ of ARFIMA(1,d,0) processes with standard normal innovations and parameter values $d = -0.25, -0.1, 0, 0.1, 0.25$ and $\phi_1 = -0.25, -0.1, 0, 0.1, 0.25$, respectively, are generated using a burn-in period of 40,000. For each realization,

the memory parameter d is estimated with the estimators \hat{d}_{GPH} , \hat{d}_{tr} , \hat{d}_{sm} , and \hat{d}_2 , which use only the conventional periodogram, and their counterparts \hat{d}_{GPH}^+ , \hat{d}_{tr}^+ , \hat{d}_{sm}^+ , and \hat{d}_2^+ , which use also the modified periodogram at fractional Fourier frequencies. Tables 1 and 2, 3 and 4, 5 and 6 give the sample bias, the sample variance, and the root mean square error for the sample sizes $n = 390$ and $n = 3000$, respectively. The smaller sample size corresponds to the number of minutes in a regular trading session for U.S. stocks, which starts at 9:30 a.m. and ends at 4:00 p.m. The parameter K , which determines the highest frequency to be used for estimation, is set to $K = n^\alpha$, where $\alpha = 0.5$.

The positive effect of including the fractional Fourier frequencies is obvious. In each of the 200 cases (25 models, 4 types of estimators, 2 sample sizes), the

Table 1 Sample bias of the estimators \hat{d}_{GPH} (log periodogram regression), \hat{d}_{GPH}^+ (log periodogram regression additionally with non-Fourier frequencies), \hat{d}_{tr} (trimming), \hat{d}_{tr}^+ (trimming additionally with non-Fourier frequencies), \hat{d}_{sm} (simple smoothing), \hat{d}_{sm}^+ (smoothing additionally with non-Fourier frequencies), \hat{d}_2 (2 partitions) and \hat{d}_2^+ (2 partitions additionally with non-Fourier frequencies) obtained from 40,000 realizations (length: $n = 390$, number of used Fourier frequencies: $K = 20$) of Gaussian ARFIMA(1, d ,0) with $d = -0.25, -0.1, 0, 0.1, 0.25$ and $\phi_1 = -0.25, -0.1, 0, 0.1, 0.25$

d	ϕ_1	\hat{d}_{GPH}	\hat{d}_{GPH}^+	\hat{d}_{tr}	\hat{d}_{tr}^+	\hat{d}_{sm}	\hat{d}_{sm}^+	\hat{d}_2	\hat{d}_2^+
-0.25	-0.25	0.0073	0.0098	0.0039	0.0055	0.0002	0.0075	0.0097	0.0126
	-0.1	0.0082	0.0099	0.0053	0.0068	0.0016	0.0078	0.0094	0.0125
	0	0.0087	0.0100	0.0050	0.0055	0.0011	0.0072	0.0099	0.0123
	0.1	0.0108	0.0125	0.0085	0.0100	0.0049	0.0102	0.0124	0.0151
	0.25	0.0135	0.0162	0.0135	0.0145	0.0086	0.0146	0.0154	0.0188
-0.1	-0.25	-0.0012	0.0011	-0.0018	-0.0006	-0.0045	-0.0006	-0.0009	0.0023
	-0.1	0.0008	0.0024	-0.0010	-0.0002	-0.0036	0.0002	0.0005	0.0034
	0	0.0026	0.0051	0.0019	0.0029	-0.0007	0.0035	0.0029	0.0061
	0.1	0.0040	0.0068	0.0051	0.0060	0.0019	0.0057	0.0044	0.0081
	0.25	0.0099	0.0122	0.0114	0.0128	0.0082	0.0109	0.0095	0.0134
0	-0.25	-0.0041	-0.0009	-0.0061	-0.0037	-0.0043	-0.0022	-0.0034	-0.0002
	-0.1	-0.0020	0.0005	-0.0021	-0.0004	-0.0023	-0.0005	-0.002	0.0011
	0	0.0003	0.0030	0.0013	0.0023	0.0004	0.0021	0.0003	0.0037
	0.1	0.0032	0.0057	0.0038	0.0051	0.0036	0.0050	0.0035	0.0070
	0.25	0.0078	0.0107	0.0099	0.0116	0.0096	0.0107	0.0084	0.0123
0.1	-0.25	-0.0035	-0.0015	-0.0047	-0.0035	-0.0005	-0.0016	-0.0045	-0.0014
	-0.1	-0.0004	0.0012	0.0007	0.0006	0.0026	0.0011	-0.0013	0.0015
	0	-0.0014	0.0013	-0.0010	0.0000	0.0028	0.0018	-0.0011	0.0021
	0.1	0.0032	0.0059	0.0048	0.0055	0.0073	0.0064	0.0037	0.0073
	0.25	0.0085	0.0109	0.0121	0.0129	0.0128	0.0117	0.0085	0.0120
0.25	-0.25	0.0019	0.0034	0.0013	0.0019	0.0114	0.0053	0.0018	0.0039
	-0.1	0.0038	0.0051	0.0030	0.0032	0.0140	0.0075	0.0041	0.0063
	0	0.0042	0.0060	0.0033	0.0038	0.0148	0.0082	0.0048	0.0072
	0.1	0.0073	0.0084	0.0066	0.0071	0.0166	0.0103	0.0071	0.0092
	0.25	0.0113	0.0135	0.0133	0.0143	0.0228	0.0164	0.0126	0.0153

Table 2 Sample bias of the estimators \hat{d}_{GPH} (log periodogram regression), \hat{d}_{GPH}^+ (log periodogram regression additionally with non-Fourier frequencies), \hat{d}_{tr} (trimming), \hat{d}_{tr}^+ (trimming additionally with non-Fourier frequencies), \hat{d}_{sm} (simple smoothing), \hat{d}_{sm}^+ (smoothing additionally with non-Fourier frequencies), \hat{d}_2 (2 partitions) and \hat{d}_2^+ (2 partitions additionally with non-Fourier frequencies) obtained from 40,000 realizations (length: $n = 3000$, number of used Fourier frequencies: $K = 55$) of Gaussian ARFIMA(1, d ,0) with $d = -0.25, -0.1, 0, 0.1, 0.25$ and $\phi_1 = -0.25, -0.1, 0, 0.1, 0.25$

d	ϕ_1	\hat{d}_{GPH}	\hat{d}_{GPH}^+	\hat{d}_{tr}	\hat{d}_{tr}^+	\hat{d}_{sm}	\hat{d}_{sm}^+	\hat{d}_2	\hat{d}_2^+
-0.25	-0.25	0.0040	0.0051	0.0022	0.0030	0.0010	0.0041	0.0055	0.0069
	-0.1	0.0053	0.0061	0.0037	0.0040	0.0020	0.0049	0.0063	0.0075
	0	0.0058	0.0066	0.0043	0.0045	0.0022	0.0053	0.0068	0.0081
	0.1	0.0055	0.0061	0.0042	0.0042	0.0017	0.0046	0.0063	0.0075
	0.25	0.0062	0.0071	0.0042	0.0049	0.0028	0.0058	0.0071	0.0083
-0.1	-0.25	0.0006	0.0018	0.0003	0.0009	-0.0009	0.0007	0.0007	0.0021
	-0.1	0.0009	0.0020	0.0000	0.0007	-0.0006	0.0011	0.0007	0.0025
	0	0.0005	0.0016	0.0003	0.0008	-0.0006	0.0009	0.0007	0.0023
	0.1	0.0017	0.0024	0.0018	0.0018	0.0002	0.0017	0.0018	0.0032
	0.25	0.0019	0.003	0.0018	0.0022	0.0005	0.0021	0.002	0.0035
0	-0.25	-0.0004	0.0007	-0.0005	-0.0001	-0.0004	0.0004	-0.0002	0.0014
	-0.1	-0.0003	0.0010	-0.0004	0.0004	-0.0002	0.0004	-0.0004	0.0012
	0	-0.0005	0.0010	-0.0010	-0.0001	-0.0005	0.0005	-0.0002	0.0015
	0.1	-0.0001	0.0007	-0.0003	0.0000	-0.0007	0.0000	-0.0004	0.0010
	0.25	0.0011	0.0020	0.0010	0.0014	0.0011	0.0016	0.0009	0.0025
0.1	-0.25	0.0012	0.0017	0.0009	0.0007	0.0016	0.0012	0.0007	0.0019
	-0.1	-0.0001	0.0007	-0.0006	-0.0002	0.0011	0.0006	-0.0003	0.0009
	0	0.0007	0.0018	0.0011	0.0014	0.0022	0.0018	0.0009	0.0024
	0.1	0.0014	0.0024	0.0018	0.0022	0.0028	0.0022	0.0013	0.0029
	0.25	0.0010	0.0023	0.0012	0.0020	0.0026	0.0023	0.0013	0.0029
0.25	-0.25	0.0024	0.0031	0.0022	0.0025	0.0062	0.0038	0.0027	0.0037
	-0.1	0.0023	0.0026	0.0024	0.002	0.0064	0.0037	0.0026	0.0033
	0	0.0029	0.0034	0.0028	0.0026	0.0069	0.0043	0.0030	0.0040
	0.1	0.0027	0.0033	0.0023	0.0025	0.0068	0.0043	0.0028	0.0039
	0.25	0.0023	0.0036	0.0019	0.0029	0.0067	0.0044	0.0031	0.0044

sample variance of the modified estimator is less than the sample variance of the original estimator (see Tables 3, 4). This improvement comes at the cost of a slightly increased sample bias (see Tables 1, 2). When both the sample bias and the sample variance are taken into account, the effect is still positive. In all 200 cases, the RMSE of the original estimator is larger than that of its modified counterpart. Clearly, the latter estimator will become less competitive when averages of several estimates are considered because the variance term in the RMSE will become less important as the number of individual estimates increases.

Table 3 Sample variance of the estimators \hat{d}_{GPH} (log periodogram regression), \hat{d}_{GPH}^+ (log periodogram regression additionally with non-Fourier frequencies), \hat{d}_{tr} (trimming), \hat{d}_{tr}^+ (trimming additionally with non-Fourier frequencies), \hat{d}_{sm} (simple smoothing), \hat{d}_{sm}^+ (smoothing additionally with non-Fourier frequencies), \hat{d}_2 (2 partitions) and \hat{d}_2^+ (2 partitions additionally with non-Fourier frequencies) obtained from 40,000 realizations (length: $n = 390$, number of used Fourier frequencies: $K = 20$) of Gaussian ARFIMA(1, d ,0) with $d = -0.25, -0.1, 0, 0.1, 0.25$ and $\phi_1 = -0.25, -0.1, 0, 0.1, 0.25$

d	ϕ_1	\hat{d}_{GPH}	\hat{d}_{GPH}^+	\hat{d}_{tr}	\hat{d}_{tr}^+	\hat{d}_{sm}	\hat{d}_{sm}^+	\hat{d}_2	\hat{d}_2^+
-0.25	-0.25	0.0334	0.0290	0.0535	0.0460	0.0331	0.0299	0.0291	0.0282
	-0.1	0.0328	0.0285	0.0525	0.0453	0.0327	0.0296	0.0288	0.0280
	0	0.0335	0.0290	0.0534	0.0457	0.0333	0.0300	0.0294	0.0284
	0.1	0.0328	0.0284	0.0530	0.0450	0.0327	0.0294	0.0288	0.0278
	0.25	0.0331	0.0288	0.0526	0.0455	0.0328	0.0298	0.0291	0.0282
-0.1	-0.25	0.0329	0.0292	0.0524	0.0454	0.0322	0.0297	0.0291	0.0286
	-0.1	0.0333	0.0294	0.0533	0.0459	0.0327	0.0300	0.0296	0.0289
	0	0.0329	0.0289	0.0528	0.0455	0.0320	0.0295	0.0288	0.0282
	0.1	0.0329	0.0291	0.0527	0.0458	0.0323	0.0298	0.0291	0.0287
	0.25	0.0323	0.0289	0.0517	0.0452	0.0320	0.0296	0.0291	0.0287
0	-0.25	0.0328	0.0292	0.0524	0.0455	0.0320	0.0297	0.0294	0.0289
	-0.1	0.0325	0.0291	0.0517	0.0452	0.0316	0.0294	0.0290	0.0287
	0	0.0328	0.0294	0.0526	0.0458	0.0323	0.0300	0.0295	0.0290
	0.1	0.0326	0.0289	0.0518	0.0445	0.0316	0.0293	0.0292	0.0286
	0.25	0.0327	0.0292	0.0529	0.0458	0.0318	0.0297	0.0293	0.0289
0.1	-0.25	0.0331	0.0296	0.0528	0.0462	0.0322	0.0301	0.0298	0.0294
	-0.1	0.0325	0.0289	0.0523	0.0452	0.0315	0.0293	0.0292	0.0288
	0	0.0331	0.0298	0.0526	0.0455	0.0318	0.0299	0.0297	0.0293
	0.1	0.0330	0.0293	0.0527	0.0456	0.0317	0.0296	0.0294	0.0290
	0.25	0.0331	0.0296	0.0527	0.0460	0.0321	0.0301	0.0295	0.0293
0.25	-0.25	0.0327	0.0293	0.0528	0.0456	0.0319	0.0297	0.0295	0.0292
	-0.1	0.0326	0.0290	0.0520	0.0450	0.0314	0.0293	0.0292	0.0288
	0	0.0328	0.0290	0.0531	0.0457	0.0317	0.0294	0.0293	0.0288
	0.1	0.0326	0.0293	0.0524	0.0455	0.0318	0.0297	0.0298	0.0294
	0.25	0.0332	0.0296	0.0534	0.0463	0.0322	0.0300	0.0298	0.0294

5 Empirical results

Greene and Fielitz (1977) used the R/S statistic (see Hurst 1951; Mandelbrot and Wallis 1969; Mandelbrot 1972, 1975), which is defined as the ratio of the range (R) of all partial sums of a time series to its standard deviation (S), in order to verify Mandelbrot’s (1971) suspicion that the strength of the statistical dependence of stock prices may decrease very slowly. Indeed, they found indications of long-range dependence when they investigated daily stock return series. However, Lo (1991) argued that the R/S statistic cannot distinguish between short-range dependence and long-range dependence and proposed therefore a simple generalization of the R/S statistic that is robust to general forms of short-range

Table 4 Sample variance of the estimators \hat{d}_{GPH} (log periodogram regression), \hat{d}_{GPH}^+ (log periodogram regression additionally with non-Fourier frequencies), \hat{d}_{tr} (trimming), \hat{d}_{tr}^+ (trimming additionally with non-Fourier frequencies), \hat{d}_{sm} (simple smoothing), \hat{d}_{sm}^+ (smoothing additionally with non-Fourier frequencies), \hat{d}_2 (2 partitions) and \hat{d}_2^+ (2 partitions additionally with non-Fourier frequencies) obtained from 40,000 realizations (length: $n = 3000$, number of used Fourier frequencies: $K = 55$) of Gaussian ARFIMA(1,d,0) with $d = -0.25, -0.1, 0, 0.1, 0.25$ and $\phi_1 = -0.25, -0.1, 0, 0.1, 0.25$

d	ϕ_1	\hat{d}_{GPH}	\hat{d}_{GPH}^+	\hat{d}_{tr}	\hat{d}_{tr}^+	\hat{d}_{sm}	\hat{d}_{sm}^+	\hat{d}_2	\hat{d}_2^+
-0.25	-0.25	0.0097	0.0081	0.0123	0.0104	0.0080	0.0077	0.0080	0.0075
	-0.1	0.0095	0.0081	0.0122	0.0103	0.0080	0.0076	0.0080	0.0075
	0	0.0095	0.0080	0.0121	0.0102	0.0079	0.0076	0.0079	0.0075
	0.1	0.0095	0.0081	0.0121	0.0103	0.0081	0.0077	0.0081	0.0076
	0.25	0.0097	0.0082	0.0123	0.0104	0.0080	0.0077	0.0081	0.0076
-0.1	-0.25	0.0095	0.0082	0.0121	0.0103	0.0079	0.0077	0.0081	0.0076
	-0.1	0.0096	0.0083	0.0123	0.0104	0.0080	0.0077	0.0082	0.0077
	0	0.0094	0.0082	0.0120	0.0103	0.0079	0.0077	0.0080	0.0076
	0.1	0.0095	0.0082	0.0122	0.0104	0.0079	0.0077	0.0081	0.0077
	0.25	0.0095	0.0081	0.0121	0.0103	0.0078	0.0076	0.0080	0.0076
0	-0.25	0.0094	0.0082	0.0121	0.0103	0.0078	0.0077	0.0081	0.0077
	-0.1	0.0095	0.0084	0.0123	0.0105	0.0079	0.0078	0.0082	0.0078
	0	0.0095	0.0082	0.0121	0.0104	0.0078	0.0077	0.0081	0.0077
	0.1	0.0095	0.0083	0.0122	0.0104	0.0078	0.0077	0.0081	0.0077
	0.25	0.0094	0.0082	0.0121	0.0104	0.0078	0.0077	0.0081	0.0077
0.1	-0.25	0.0093	0.0082	0.0120	0.0103	0.0077	0.0077	0.0081	0.0077
	-0.1	0.0096	0.0084	0.0123	0.0105	0.0079	0.0078	0.0082	0.0078
	0	0.0096	0.0084	0.0122	0.0104	0.0078	0.0077	0.0082	0.0077
	0.1	0.0095	0.0083	0.0121	0.0104	0.0078	0.0077	0.0081	0.0077
	0.25	0.0095	0.0083	0.0122	0.0104	0.0079	0.0077	0.0082	0.0077
0.25	-0.25	0.0095	0.0082	0.0122	0.0103	0.0077	0.0076	0.0081	0.0076
	-0.1	0.0095	0.0082	0.0120	0.0102	0.0078	0.0076	0.0081	0.0077
	0	0.0096	0.0082	0.0120	0.0103	0.0078	0.0076	0.0082	0.0077
	0.1	0.0095	0.0082	0.0123	0.0104	0.0078	0.0076	0.0082	0.0077
	0.25	0.0096	0.0083	0.0124	0.0105	0.0079	0.0077	0.0083	0.0078

dependence. Using this modified R/S statistic, he found little evidence in favor of long-range dependence in daily and monthly index returns. Similarly, Barkoulas and Baum (1996) found no convincing evidence of long-range dependence when they applied the log periodogram regression method of Geweke and Porter-Hudak (1983) to daily U.S. stock returns and index returns. Negative results were also obtained for international data by Cheung and Lai (1995) with both modified R/S analysis and log periodogram regression and by Crato (1994) with parametric methods (ARFIMA models). In contrast to signed returns, absolute returns and squared returns actually do appear to exhibit significant long-range dependence (see Crato and de Lima 1994; Lobato and Savin 1998; Grau-Carles 2000).

Table 5 RMSE of the estimators \hat{d}_{GPH} (log periodogram regression), \hat{d}_{GPH}^+ (log periodogram regression additionally with non-Fourier frequencies), \hat{d}_{tr} (trimming), \hat{d}_{tr}^+ (trimming additionally with non-Fourier frequencies), \hat{d}_{sm} (simple smoothing), \hat{d}_{sm}^+ (smoothing additionally with non-Fourier frequencies), \hat{d}_2 (2 partitions) and \hat{d}_2^+ (2 partitions additionally with non-Fourier frequencies) obtained from 40,000 realizations (length: $n = 390$, number of used Fourier frequencies: $K = 20$) of Gaussian ARFIMA(1, d ,0) with $d = -0.25, -0.1, 0, 0.1, 0.25$ and $\phi_1 = -0.25, -0.1, 0, 0.1, 0.25$

d	ϕ_1	\hat{d}_{GPH}	\hat{d}_{GPH}^+	\hat{d}_{tr}	\hat{d}_{tr}^+	\hat{d}_{sm}	\hat{d}_{sm}^+	\hat{d}_2	\hat{d}_2^+
-0.25	-0.25	0.1828	0.1705	0.2313	0.2145	0.1818	0.1730	0.1709	0.1683
	-0.1	0.1814	0.1691	0.2292	0.2129	0.1809	0.1721	0.1700	0.1679
	0	0.1832	0.1705	0.2311	0.2138	0.1825	0.1734	0.1717	0.1690
	0.1	0.1814	0.1688	0.2304	0.2125	0.1809	0.1718	0.1701	0.1675
	0.25	0.1823	0.1706	0.2298	0.2137	0.1813	0.1732	0.1711	0.1690
-0.1	-0.25	0.1814	0.1709	0.2290	0.2131	0.1796	0.1724	0.1706	0.1690
	-0.1	0.1825	0.1714	0.2310	0.2142	0.1808	0.1732	0.1721	0.1700
	0	0.1813	0.1702	0.2299	0.2134	0.1790	0.1717	0.1697	0.1682
	0.1	0.1813	0.1708	0.2296	0.2141	0.1796	0.1727	0.1707	0.1697
	0.25	0.1801	0.1705	0.2276	0.2130	0.1789	0.1725	0.1709	0.1698
0	-0.25	0.1812	0.1710	0.2289	0.2134	0.1789	0.1724	0.1714	0.1699
	-0.1	0.1803	0.1706	0.2275	0.2127	0.1779	0.1713	0.1703	0.1693
	0	0.1811	0.1716	0.2293	0.2140	0.1798	0.1731	0.1716	0.1704
	0.1	0.1806	0.1701	0.2277	0.2111	0.1777	0.1714	0.1709	0.1693
	0.25	0.1810	0.1712	0.2303	0.2143	0.1785	0.1727	0.1713	0.1705
0.1	-0.25	0.1820	0.1719	0.2299	0.2150	0.1794	0.1735	0.1726	0.1714
	-0.1	0.1803	0.1700	0.2286	0.2126	0.1775	0.1713	0.1710	0.1696
	0	0.1819	0.1725	0.2293	0.2134	0.1784	0.1730	0.1723	0.1713
	0.1	0.1818	0.1714	0.2296	0.2137	0.1781	0.1722	0.1715	0.1703
	0.25	0.1822	0.1724	0.2299	0.2149	0.1797	0.1738	0.1721	0.1715
0.25	-0.25	0.1809	0.1711	0.2299	0.2137	0.1789	0.1724	0.1718	0.1710
	-0.1	0.1805	0.1705	0.2281	0.2121	0.1778	0.1713	0.1710	0.1698
	0	0.1810	0.1705	0.2304	0.2137	0.1786	0.1717	0.1712	0.1699
	0.1	0.1806	0.1714	0.2290	0.2133	0.1792	0.1727	0.1727	0.1718
	0.25	0.1826	0.1727	0.2315	0.2156	0.1808	0.1738	0.1730	0.1721

When high-frequency data are available, realized variances (sums of squared intraday returns) can be used instead of squared daily returns. Alternatively, the unaggregated squared intraday returns can be investigated right away. The latter approach has the advantage that the sample size is many times greater than the number of trading days. A disadvantage is that there are usually gaps between the individual trading sessions, which make it necessary to estimate the memory parameter d separately for each trading session and compute the final estimate by averaging the individual estimates (see Reschenhofer and Mangat 2020). In this case, it is important to use an estimator with a small bias because the variance decreases with the number of trading sessions but the bias remains fixed. Of course, it might also be interesting to adopt the same approach for the signed intraday returns. In Fig. 1, the

Table 6 RMSE of the estimators \hat{d}_{GPH} (log periodogram regression), \hat{d}_{GPH}^+ (log periodogram regression additionally with non-Fourier frequencies), \hat{d}_{tr} (trimming), \hat{d}_{tr}^+ (trimming additionally with non-Fourier frequencies), \hat{d}_{sm} (simple smoothing), \hat{d}_{sm}^+ (smoothing additionally with non-Fourier frequencies), \hat{d}_2 (2 partitions) and \hat{d}_2^+ (2 partitions additionally with non-Fourier frequencies) obtained from 40,000 realizations (length: $n = 3000$, number of used Fourier frequencies: $K = 55$) of Gaussian ARFIMA(1, d ,0) with $d = -0.25, -0.1, 0, 0.1, 0.25$ and $\phi_1 = -0.25, -0.1, 0, 0.1, 0.25$

d	ϕ_1	\hat{d}_{GPH}	\hat{d}_{GPH}^+	\hat{d}_{tr}	\hat{d}_{tr}^+	\hat{d}_{sm}	\hat{d}_{sm}^+	\hat{d}_2	\hat{d}_2^+
-0.25	-0.25	0.0984	0.0904	0.1110	0.1019	0.0897	0.0877	0.0898	0.0871
	-0.1	0.0978	0.0901	0.1104	0.1013	0.0894	0.0875	0.0896	0.0868
	0	0.0975	0.0899	0.1103	0.1012	0.089	0.0873	0.0893	0.0867
	0.1	0.0976	0.0903	0.1103	0.1017	0.0898	0.0881	0.0900	0.0873
	0.25	0.0985	0.0910	0.1109	0.1022	0.0898	0.0881	0.0904	0.0876
-0.1	-0.25	0.0976	0.0906	0.1101	0.1015	0.0887	0.0877	0.0899	0.0874
	-0.1	0.0979	0.0909	0.1109	0.1020	0.0892	0.0880	0.0907	0.0880
	0	0.0972	0.0904	0.1094	0.1014	0.0886	0.0875	0.0895	0.0872
	0.1	0.0977	0.0909	0.1106	0.1021	0.0889	0.0880	0.0901	0.0876
	0.25	0.0974	0.0902	0.1101	0.1013	0.0884	0.0875	0.0896	0.0871
0	-0.25	0.0967	0.0904	0.1099	0.1017	0.0884	0.0877	0.0899	0.0877
	-0.1	0.0977	0.0914	0.1108	0.1025	0.0889	0.0882	0.0905	0.0882
	0	0.0973	0.0904	0.1102	0.1018	0.0883	0.0876	0.0899	0.0875
	0.1	0.0974	0.0908	0.1106	0.1018	0.0883	0.0877	0.0900	0.0877
	0.25	0.0968	0.0904	0.1101	0.1019	0.0886	0.0878	0.0900	0.0878
0.1	-0.25	0.0967	0.0907	0.1097	0.1017	0.0880	0.0875	0.0898	0.0879
	-0.1	0.0981	0.0917	0.1108	0.1024	0.0890	0.0883	0.0907	0.0883
	0	0.0980	0.0914	0.1106	0.1020	0.0881	0.0877	0.0906	0.0881
	0.1	0.0973	0.0910	0.1102	0.1020	0.0882	0.0877	0.0900	0.0879
	0.25	0.0976	0.0909	0.1105	0.1019	0.0886	0.0879	0.0903	0.0880
0.25	-0.25	0.0973	0.0905	0.1103	0.1013	0.0881	0.0870	0.0901	0.0875
	-0.1	0.0973	0.0906	0.1096	0.1010	0.0883	0.0873	0.0903	0.0878
	0	0.0978	0.0908	0.1098	0.1015	0.0885	0.0875	0.0904	0.0878
	0.1	0.0976	0.0906	0.1110	0.1019	0.0886	0.0876	0.0905	0.0879
	0.25	0.0982	0.0914	0.1112	0.1023	0.0890	0.0880	0.0909	0.0884

estimates obtained from the individual trading sessions are plotted cumulatively. By dividing each estimate by the total number of trading sessions, we make sure that the final value corresponds to the mean of all estimates. Each subsample consists of 390 1-min intraday returns of the General Electric Company (GE) stock, which are obtained from the first mid-quotes (midpoints of the best bid and ask quotes) in each minute. The limit order book data from 2007-06-27 to 2019-04-30 have been downloaded from Lobster (<https://lobsterdata.com>).

Figure 1a, b suggests that the memory parameter is close to zero for the signed returns and somewhere between 0.2 and 0.3 for the transformed absolute returns. The discrepancies between the results obtained with the original estimators and the modified estimators, respectively, are relatively small. In case of the GPH-estimator,

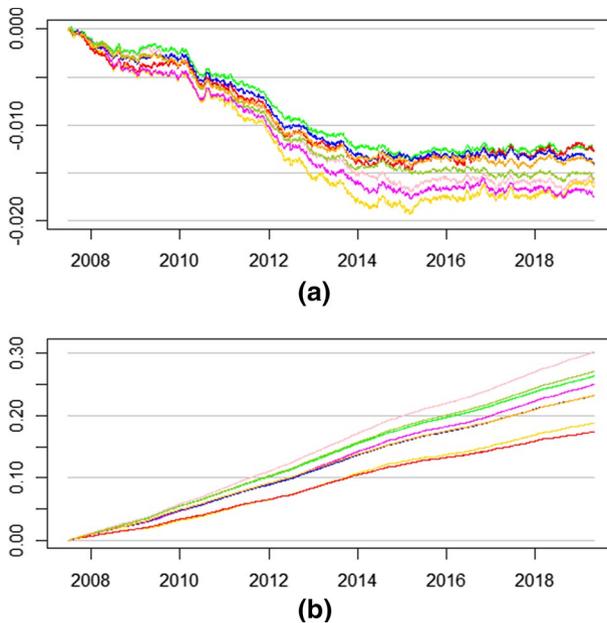


Fig. 1 Cumulative plots of the estimates obtained by applying \hat{d}_{GPH} (pink), \hat{d}_{GPH}^+ (green), \hat{d}_{lr} (gold), \hat{d}_{lr}^+ (red), \hat{d}_{sm} (magenta), \hat{d}_{sm}^+ (blue), \hat{d}_2 (yellowgreen) and \hat{d}_2^+ (orange) to the **a** 1-min intraday log returns $r_t(s), s = 1, \dots, 390$, **b** transformed absolute 1-min intraday log returns $\log(\varepsilon + |r_t(s)|)$ with $\varepsilon = 10^{-6}$ of GE stock

the difference between the means is -0.0034 (see Table 7) and the difference between the biases is -0.0027 for $d = 0, \phi_1 = 0$ (see Table 1). The agreement is less good in case of the other estimators. For the transformed absolute returns, the estimates produced by the modified estimators are generally greater than those produced by the original estimators, which is inconsistent with the results of the simulation study. This discrepancy may indicate that the transformation of the absolute returns does not achieve approximate normality. However, Table 7 shows that the inclusion of the fractional Fourier frequencies always leads to a reduction in the variance, which is of crucial importance in any conventional study with only a single sample of observations.

6 Discussion

A standard approach to improve the performance of frequency-domain methods for the estimation of the memory parameter d is to smooth the periodogram before it is put to use. While this approach usually leads to a reduction of the variance of the estimator, it has the opposite effect on the size of the bias. Alternatively, a reduction in the variance can also be achieved by additionally including fractional Fourier frequencies. Unfortunately, there are two problems. Firstly, the Fast Fourier Transform cannot be used for the calculation of the periodogram at these frequencies.

Table 7 Sample mean and sample variance of the daily estimates of the memory parameter d obtained for the signed 1-min returns and the transformed absolute 1-min returns of the GE stock with the estimators \hat{d}_{GPH} (log periodogram regression), \hat{d}_{GPH}^+ (log periodogram regression additionally with non-Fourier frequencies), \hat{d}_{tr} (trimming), \hat{d}_{tr}^+ (trimming additionally with non-Fourier frequencies), \hat{d}_{sm} (simple smoothing), \hat{d}_{sm}^+ (smoothing additionally with non-Fourier frequencies), \hat{d}_2 (2 partitions) and \hat{d}_2^+ (2 partitions additionally with non-Fourier frequencies)

Times series	Statistic	\hat{d}_{GPH}	\hat{d}_{GPH}^+	\hat{d}_{tr}	\hat{d}_{tr}^+	\hat{d}_{sm}	\hat{d}_{sm}^+	\hat{d}_2	\hat{d}_2^+
Returns	$mean(\hat{d})$	-0.0161	-0.0127	-0.0163	-0.0126	-0.0174	-0.0139	-0.0161	-0.0157
	$var(\hat{d})$	0.0445	0.0361	0.0674	0.0547	0.0459	0.0361	0.0445	0.0355
Transform. abs. returns	$mean(\hat{d})$	0.3028	0.2635	0.1884	0.1741	0.2500	0.2322	0.3028	0.2712
	$var(\hat{d})$	0.0301	0.0275	0.0549	0.0454	0.0328	0.0285	0.0301	0.0288

Secondly, the usual orthogonality conditions are only satisfied for the regular Fourier frequencies, which raises the question whether the standard definition of the periodogram is still meaningful in the general case. In this article, we address both problems. We propose to first compute the naïve sine and cosine transforms for the fractional Fourier frequencies with a modified version of the Fast Fourier Transform and then to amend these transforms by taking the violation of the standard orthogonality conditions into account. The computational efficiency of the second stage is due to the fact that a large number of auxiliary quantities only need to be computed once because they do not depend on the data.

We have carried out an extensive simulation study in order to investigate the effect of the inclusion of the fractional Fourier frequencies on the performance of various estimators. The results show that the overall effect in terms of RMSE is always positive which is due to a significant reduction in the variance and only a slight increase in the bias. These findings are corroborated by the results of an empirical study of financial high-frequency data. In this study, the variances of the modified estimators, which use also fractional Fourier frequencies, are always smaller than those of their conventional counterparts. Clearly, little can be said about the size of the bias because the true value of the memory parameter is unknown in practice. However, all competing estimators, regardless which frequencies are used, unanimously agree that significant long-range dependence is present only in the intraday volatility but not in the signed intraday returns.

Funding Open Access funding provided by University of Vienna.

Data availability The data were downloaded from Lobster (<https://lobsterdata.com>).

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barkoulas J, Baum C (1996) Long-term dependence in stock returns. *Econ Lett* 53:253–259
- Beran J (1994) *Statistics for long-memory processes*. Chapman & Hall, New York
- Chen G, Abraham B, Peiris S (1994) Lag window estimation of the degree of differencing in fractionally integrated time series models. *J Time Ser Anal* 15:473–487

- Cheung Y, Lai K (1995) A search for long memory in international stock market returns. *J Int Money Finance* 14:597–615
- Cooley JW, Tukey JW (1965) An algorithm for the machine calculation of complex Fourier series. *Math Comput* 19:297–301
- Crato N (1994) Some international evidence regarding the stochastic behaviour of stock returns. *Appl Financ Econ* 4:33–39
- Crato N, de Lima P (1994) Long-range dependence in the conditional variance of stock returns. *Econ Lett* 45:281–285
- Fox R, Taqqu MS (1986) Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann Stat* 14:57–532
- Geman S, Bienenstock E, Doursat R (1992) Neural networks and the bias/variance dilemma. *Neural Comput* 4:1–58
- Geweke J, Porter-Hudak S (1983) The estimation and application of long memory time series models. *J Time Ser Anal* 4:221–238
- Gradshteyn IS, Ryzhik IM (2007) Table of integrals, series, and products, 7th edn. Elsevier/Academic Press, Amsterdam
- Granger CWJ, Joyeux R (1980) An introduction to long-memory time series models and fractional differencing. *J Time Ser Anal* 1:15–29
- Grau-Carles P (2000) Empirical evidence of long-range correlations in stock returns. *Phys A* 287:396–404
- Graves T, Gramacy R, Watkins N, Franzke C (2017) A brief history of long memory: Hurst, Mandelbrot and the road to ARFIMA, 1951–1980. *Entropy* 19:437
- Greene MT, Fielitz BD (1977) Long term dependence in common stock returns. *J Financ Econ* 4:339–349
- Hassler U (1993) Regression of spectral estimators with fractionally integrated time series. *J Time Ser Anal* 14:369–380
- Hosking JRM (1981) Fractional differencing. *Biometrika* 68:165–176
- Hurst HE (1951) Long-term storage capacity of reservoirs. *Trans Am Soc Civ Eng* 116:770–799
- Hurvich CM, Deo R, Brodsky J (1998) The mean square error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series. *J Time Ser Anal* 19:19–46
- Künsch HR (1986) Discrimination between monotonic trends and long-range dependence. *J Appl Probab* 23:1025–1030
- Lo A (1991) Long-term memory in stock market prices. *Econometrica* 59:1279–1313
- Lobato IN, Savin NE (1998) Real and spurious long-memory properties of stock-market data. *J Bus Econ Stat* 16:261–268
- Mandelbrot B (1971) When can price be arbitrated efficiently? A limit to the validity of the random walk and martingale models. *Rev Econ Stat* 53:225–236
- Mandelbrot B (1972) Statistical methodology for non-periodic cycles: from the covariance to R/S analysis. *Ann Econ Soc Meas* 1:259–290
- Mandelbrot B (1975) Limit theorems on the self-normalized range for weakly and strongly dependent processes. *Z Wahrscheinlichkeits Verw Geb* 31:271–285
- Mandelbrot B, Wallis J (1969) Computer experiments with fractional Gaussian noises. Parts 1, 2, 3. *Water Resour Res* 4:909–918
- Neal B, Mittal S, Baratin A, Tantia V, Sciluna M, Lacoste-Julien S, Mitliagkas I (2020) A modern take on the bias-variance tradeoff in neural networks. arXiv preprint [arXiv:1810.08591](https://arxiv.org/abs/1810.08591)
- Peiris MS, Court JR (1993) A note on the estimation of degree of differencing in long memory time series analysis. *Probab Math Stat* 14:223–229
- Pötscher BM (2002) Lower risk bounds and properties of confidence sets for ill-posed estimation problems with applications to spectral density and persistence estimation, unit roots, and estimation of long memory parameters. *Econometrica* 70:1035–1065
- Reisen VA (1994) Estimation of the fractional difference parameter in the ARIMA(p, d, q) model using the smoothed periodogram. *J Time Ser Anal* 15:335–350
- Reschenhofer E (2013) Log-periodogram regression with odd Fourier frequencies. *InterStat: July 2013#004*
- Reschenhofer E, Mangat MK (2020) Reducing the bias of the smoothed log periodogram regression for financial high-frequency data. *Econometrics* 8(40):1–16
- Reschenhofer E, Mangat MK, Stark T (2020) Improved estimation of the memory parameter. *Theor Econ Lett* 10:47–68

Robinson PM (1995) Log-periodogram regression of time series with long range dependence. *Ann Stat* 23:1048–1072

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.