Lucie Kraicová and Jozef Baruník*

Estimation of long memory in volatility using wavelets

DOI 10.1515/snde-2016-0101

Abstract: This work studies wavelet-based Whittle estimator of the fractionally integrated exponential generalized autoregressive conditional heteroscedasticity (FIEGARCH) model often used for modeling long memory in volatility of financial assets. The newly proposed estimator approximates the spectral density using wavelet transform, which makes it more robust to certain types of irregularities in data. Based on an extensive Monte Carlo study, both behavior of the proposed estimator and its relative performance with respect to traditional estimators are assessed. In addition, we study properties of the estimators in presence of jumps, which brings interesting discussion. We find that wavelet-based estimator may become an attractive robust and fast alternative to the traditional methods of estimation. In particular, a localized version of our estimator becomes attractive in small samples.

Keywords: FIEGARCH; long memory; Monte Carlo; volatility; wavelets; Whittle.

1 Introduction

During past decades, volatility has become one of the most extensively studied variables in finance. This enormous interest has mainly been spurred by the importance of volatility as a measure of risk for both academics and practitioners. Despite numerous modeling and estimation approaches developed in the literature, there are many interesting aspects of estimation waiting for further research. One area of lively discussions is estimation of parameters in long memory models that capture persistence of volatility time series. This persistence belongs to the important stylized facts, as it implies that shock in the volatility will impact future volatility over a long horizon. The FI(E)GARCH extension (Bollerslev and Mikkelsen 1996) to the original (G)ARCH modeling framework (Engle 1982; Bollerslev 1986) was shown to capture this empirically observed correlation well. In our work, we contribute to the discussion with interesting alternative estimation framework for the FIEGARCH model based on wavelet approximation of likelihood function.

Although traditional maximum likelihood (ML) framework for parameters estimation is desirable due to its efficiency, an alternative approach, Whittle estimator can be employed (Zaffaroni 2009). The Whittle estimator is obtained by maximizing frequency domain approximation of the Gaussian likelihood function, the so-called Whittle function (Whittle 1962), and although it can not attain better efficiency, it may serve as a computationally fast alternative to ML for complex optimization problems.

Traditionally, Whittle estimators use likelihood approximations based on Fourier transform. Whereas this is accurate alternative to be used in many applications, in finance, non-stationarities and significant time-localized patterns in data can emerge. Jensen (1999) provides an alternative type of estimation based on approximation of likelihood function using wavelets, which are time localized and can better approximate spectral density.

***Corresponding author: Jozef Baruník,** Institute of Economic Studies, Charles University, Opletalova 26, 110 00 Prague, Czech Republic; and Department of Econometrics, IITA, The Czech Academy of Sciences, Pod Vodarenskou Vezi 4, 18200 Prague, Czech Republic, e-mail: barunik@utia.cas.cz
Lucie Kraicová: Institute of Economic Studies, Charles University, Opletalova 26, 110 00 Prague, Czech Republic; and Department of Econometrics, IITA, The Czech Academy of Sciences, Pod Vodarenskou Vezi 4, 18200 Prague, Czech Republic

Favorable properties of wavelets has been increasingly used for estimation as well as testing strategies in economics and finance. Gençay and Gradojevic (2011) use wavelets to address error-in-variables problem in a classical linear regression setting, Tseng and Gençay (2014) further estimate linear models with a time-varying parameter. Using spectral properties of time series, Gencay and Signori (2015) proposes a new family of portmanteau tests for serial correlation based on wavelet decomposition, and Fan and Gençay (2010) new wavelet approach to testing the presence of a unit root in a stochastic process. In a high frequency econometrics literature, Fan and Wang (2007), Xue, Gençay, and Fagan (2014), and Barunik and Vacha (2015) use wavelets successfully in jump detection, and estimation of realized volatility at different scales. Barunik, Krehlik, and Vacha (2016) build a multi-scale model with jumps to forecast volatility, and Barunik and Vacha (2016) further the research in estimation of wavelet realized covariation as well as co-jumps.

Compared to the wide range of studies on semi-parametric Wavelet Whittle estimators [for relative performance of local FWE and WWE of ARFIMA model see e.g. Faÿ et al. (2009) or Frederiksen and Nielsen (2005) and related works], literature assessing performance of their parametric counterparts is not extensive. Though, results of the studies on parametric WWE completed so far are promising. Jensen (1999) introduces wavelet Whittle estimation (WWE) of ARFIMA process, and compares its performance with traditional Fourier-based Whittle estimator. He finds that estimators perform similarly, with an exception of MA coefficients being close to boundary of invertibility of the process. In this case, Fourier-based estimation deteriorates, whereas wavelet-based estimation retains its accuracy. Percival and Walden (2000) describe a wavelet-based approximate MLE for both stationary and non-stationary fractionally differenced processes, and demonstrates its relatively good performance on very short samples (128 observations). Whitcher (2004) applies WWE based on a discrete wavelet packet transform (DWPT) to a seasonal persistent process and again finds good performance of this estimation strategy. Heni and Mohamed (2011) apply this strategy on a FIGARCH-GARMA model, further application can be seen in Gonzaga and Hauser (2011).

Literature focusing on WWE studies various models, but estimation of FIEGARCH has not been fully explored yet with exception of Perez and Zaffaroni (2008) and Zaffaroni (2009). These authors successfully applied traditional Fourier-based Whittle estimators of FIEGARCH models, and found that Whittle estimates perform better in comparison to ML in cases of processes close to being non-stationary. Authors found that while ML is often more efficient alternative, FWE outperforms it in terms of bias mainly in case of high persistence of the processes. Hence Whittle type of estimators seem to offer lower bias at cost of lower efficiency.

In our work, we contribute to the literature by extending the study of Perez and Zaffaroni (2008) using wavelet-based Whittle estimator (Jensen 1999). The newly introduced WWE is based on two alternative approximations of likelihood function. Following the work of Jensen (1999), we propose to use discrete wavelet transform (DWT) in approximation of FIEGARCH likelihood function, and alternatively, we use maximal overlap discrete wavelet transform (MODWT). Moreover, we also study the localized version of WWE. In an experiment setup mirroring that of Perez and Zaffaroni (2008), we focus on studying small sample performance of the newly proposed estimators, and guiding potential users of the estimators through practical aspects of estimation. To study both small sample properties of the estimator and its relative performance to traditional estimation techniques under different situations, we run extensive Monte Carlo experiments. Competing estimators are Fourier-based Whittle estimator (FWE), and traditional maximum likelihood estimator (MLE). In addition, we also study the performance of estimators under the presence of jumps in the processes.

Our results show that even in the case of simulated data, which follow a pure FIEGARCH process, and thus do not allow to fully utilize the advantages of WWE over its traditional counterparts, the estimator performs reasonably well. When we focus on the individual parameters estimation, in terms of bias the performance is comparable to traditional estimators, in some cases outperforming FWE. Localized version of our estimator using partial decomposition up to five scales gives the best results in small samples, whereas it is preferable to use the estimator with full information in large samples. In terms of forecasting performance, the differences are even smaller. The exact MLE mostly outperforms both of the Whittle estimators in terms of efficiency, with just rare exceptions. Yet, due to the computational complexity of the MLE in case of large data sets, FWE and WWE thus represent an attractive fast alternatives for parameter estimation.

2 Usual estimation frameworks for FIEGARCH(q, d, p)

2.1 FIEGARCH(q, d, p) process

Observation of time variation in volatility and consecutive development of models capturing the conditional volatility became one of the most important steps in understanding risk in stock markets. Original autoregressive conditional heteroskedastic (ARCH) class of models introduced in the seminal Nobel Prize winning paper by Engle (1982) spurred race in development of new and better procedures for modeling and forecasting time-varying financial market volatility [see e.g. Bollerslev (2008) for a glossary]. The main aim of the literature was to incorporate important stylized facts about volatility, long memory being one of the most pronounced ones.

In our study we focus on one of the important generalizations capturing long memory. Fractionally integrated exponential generalized autoregressive conditional heteroscedasticity, FIEGARCH(q, d, p) models log-returns $\{\varepsilon_t\}_{t=1}^{T}$ conditionally on their past realizations as:

$$\varepsilon_t = z_t h_t^{1/2} \tag{1}$$

$$\ln(h_t) = \omega + \Phi(L) g(z_{t-1}) \tag{2}$$

$$g(z_t) = \theta z_t + \gamma[|z_t| - E(|z_t|)], \tag{3}$$

where z_t is an $\mathcal{N}(0, 1)$ independent identically distributed (*i.i.d.*) unobservable innovations process, ε_t is observable discrete-time real valued process with conditional log-variance process dependent on the past innovations $E_{t-1}(\varepsilon_t^2) = h_t$, and L is a lag operator $L^i g_t = g_{t-i}$ in $\Phi(L) = (1-L)^{-d}[1+\alpha(L)][\beta(L)]^{-1}$. The polynomials $\alpha(L) = 1 + \alpha_{[2]}(L) = 1 + \sum_{i=2}^{p} \alpha_i L^{i-1}$ and $\beta(L) = 1 - \sum_{i=1}^{q} \beta_i L^i$ have no zeros in common, their roots are outside the unit circle, $\theta\gamma \neq 0$ and $d < 0.5$. $(1-L)^d = 1 - d\sum_{k=1}^{\infty} \Gamma(k-d)\Gamma(1-d)^{-1}\Gamma(k+1)^{-1} L^k$ with Γ (.) being gamma function.

The model is able to generate important stylized facts about real financial time series data including long memory, volatility clustering, leverage effect and fat tailed distribution of returns. While correct model specification is important for capturing all the empirical features of the data, feasibility of estimation of its parameters is crucial. Below, estimation methods are described together with practical aspects of their application.

2.2 (Quasi) maximum likelihood estimator

As a natural benchmark estimation framework, maximum likelihood estimation will serve to us in the comparison exercise. For a general zero mean, stationary Gaussian process $\{x_t\}_{t=1}^{T}$, the maximum likelihood estimator (MLE) minimizes following (negative) log-likelihood function $\mathcal{L}_{\mathrm{MLE}}(\zeta)$ with respect to vector of parameters ζ

$$\mathcal{L}_{\mathrm{MLE}}(\zeta) = \frac{T}{2}\ln(2\pi) + \frac{1}{2}\ln|\Sigma_T| + \frac{1}{2}(x_t' \Sigma_T^{-1} x_t), \tag{4}$$

where Σ_T is the covariance matrix of x_t, $|\Sigma_T|$ is its determinant and ζ is the vector of parameters to be estimated.

While MLE is the most efficient estimator in the class of available efficient estimators, its practical applicability may be limited in some cases. For long memory processes with dense covariance matrices, it may be extremely time demanding, or even unfeasible with large datasets to deal with inversion of the covariance matrix. Moreover, solution may be even unstable in the presence of long memory [(Beran 1994), chapter 5], when the covariance matrix is close to singularity. In addition, empirical data often does not to have zero mean, hence the mean has to be estimated and deducted. The efficiency and bias of the estimator of the mean then contributes to the efficiency and bias of the MLE. In case of long-memory processes it can cause significant deterioration of the MLE (Cheung and Diebold 1994). Both these issues have motivated construction of alternative estimators, usually formulated as approximate MLE and defined by an approximated log-likelihood function (Beran 1994; Nielsen and Frederiksen 2005).

Most important, MLE can lead to inconsistent estimates of model parameters if the distribution of innovation is misspecified. Alternatively, quasi-maximum likelihood estimator (QMLE) is often considered, as it provides consistent estimates even if the true distribution is far from Gaussian, provided existence of fourth moment. Under high-level assumptions, Bollerslev and Wooldridge (1992) studied the theory for GARCH(p, q), although asymptotic theory for FIEGARCH process is not available.

The reduced-form negative log-likelihood function assuming log-returns ε_t to follow a Gaussian, zero-mean process of independent variables, Σ_T being diagonal with conditional variances h_t as its elements, and determinant reducing to a sum of its diagonal terms, can be written as:

$$\mathcal{L}_{(\mathrm{Q})\mathrm{MLE}}(\zeta)=\sum_{t=1}^{T}\left(\ln h_t(\zeta)+\frac{\varepsilon_t^2}{h_t(\zeta)}\right), \tag{5}$$

Then the (Q)MLE estimator is defined as $\hat{\zeta}_{(\mathrm{Q})\mathrm{MLE}}=\operatorname{argmin}_{\zeta\in\Theta}\mathcal{L}_{(\mathrm{Q})\mathrm{MLE}}(\zeta)$, where Θ is the parameter space.

While QMLE is feasible estimator in case of short-memory processes, when long memory is present, relatively large truncation is necessary to prevent a significant loss of information about long-run dependencies in the process declining slowly. In our Monte Carlo experiment, we follow Bollerslev and Mikkelsen (1996) and use sample volatility as pre-sample conditional volatility with truncation at lag 1000. Given the complexity of this procedure, the method remains significantly time consuming.

2.3 Fourier-based Whittle estimator

Fourier-based Whittle estimator (FWE) serves as spectral-based alternative, where the problematic terms in the log-likelihood function $|\Sigma_T|$ and $x_t'\Sigma_T^{-1}x_t$, are replaced by their asymptotic frequency domain representations. Orthogonality of the Fourier transform projection matrix ensures diagonalization of the covariance matrix and allows to achieve the approximation by means of multiplications by identity matrices, simple rearrangements and approximation of integrals by Riemann sums [see e.g. Beran (1994)]. The reduced-form approximated Whittle negative log-likelihood function for estimation of parameters under Gaussianity assumption is:

$$\mathcal{L}_W(\zeta)=\frac{1}{T}\sum_{j=1}^{m^*}\left(\ln f(\lambda_j,\zeta)+\frac{I(\lambda_j)}{f(\lambda_j,\zeta)}\right), \tag{6}$$

where $f(\lambda_j, \zeta)$ is the spectral density of process x_t evaluated at frequencies $\lambda_j=j/T$ (i.e. $2\pi j/T$ in terms of angular frequencies) for $j=1, 2, \dots m^*$ and $m^*=\max\{m\in Z; m\leq(T-1)/2\}$, i.e. $\lambda_j<1/2$, and its link to the variance-covariance matrix of the process x_t is:

$$\operatorname{cov}(x_t, x_s)=\int_{-1/2}^{1/2}f(\lambda,\zeta)e^{i2\pi\lambda(s-t)}d\lambda=2\int_0^{1/2}f(\lambda,\zeta)e^{i2\pi\lambda(s-t)}d\lambda; \tag{7}$$

see Percival and Walden (2000) for details. The $I(\lambda_j)$ is the value of periodogram of x_t at jth Fourier frequency:

$$I(\lambda_j)=(2\pi T)^{-1}\left|\sum_{t=1}^{T}x_t e^{i2\pi\lambda_j t}\right|^2, \tag{8}$$

and the respective Fourier-based Whittle estimator is defined as $\hat{\zeta}_W=\underset{\zeta\in\Theta}{\operatorname{argmin}}\,\mathcal{L}_W(\zeta)$. [1]

1 [For a detailed FWE treatment see e.g. Beran (1994)].

It can be shown that the FWE has the same asymptotic distribution as the exact MLE, hence is asymptotically efficient for Gaussian processes (Fox and Taqqu 1986; Dahlhaus 1989, 2006). In the literature, FWE is frequently applied to both Gaussian and non-Gaussian processes (equivalent to QMLE), whereas even in the later case, both finite sample and asymptotic properties of the estimator are often shown to be very favorable and the complexity of the computation depends on the form of the spectral density of the process. Next to a significant reduction in estimation time, the FWE also offers an efficient solution for long-memory processes with an unknown mean, which can impair efficiency of the MLE. By elimination of the zero frequency coefficient FWE becomes robust to addition of constant terms to the series, and thus in case, when no efficient estimator of the mean is available, FWE can become an appropriate choice even for time series where the MLE is still computable within reasonable time.

Concerning the FIEGARCH estimation, the FIEGARCH-FWE is, to the authors' best knowledge, the only estimator, for which an asymptotic theory is currently available. Strong consistency and asymptotic normality are established in Zaffaroni (2009) for a whole class of exponential volatility processes, even though the estimator works as an approximate QMLE of a process with an asymmetric distribution, rather than an approximate MLE. This is due to the need to adjust the model to enable derivation of the spectral density of the estimated process. More specifically, it is necessary to rewrite the model in a signal plus noise form [for details see Perez and Zaffaroni (2008) and Zaffaroni (2009)]:

$$x_t=\ln(\varepsilon_t^2)=\ln(z_t^2)+\omega+\sum_{s=0}^{\infty}\Phi_s g(z_{t-s-1}) \tag{9}$$

$$g(z_t)=\theta z_t+\gamma[|z_t|-E(|z_t|)] \tag{10}$$

$$\Phi(L)=(1-L)^{-d}[1+\alpha_{[2]}(L)][\beta(L)]^{-1}. \tag{11}$$

where for FIEGARCH(1, d, 2), it holds that $\alpha_{[2]}(L)=\alpha L$, and $\beta(L)=1-\beta L$. The process x_t then enters the FWE objective function instead of the process ε_t. The transformed process is derived together with its spectral density in the on-line appendix.

3 Wavelet Whittle estimation of FIEGARCH(q, d, p)

Although FWE seems to be a advantageous alternative for estimation of FIEGARCH parameters (Perez and Zaffaroni 2008), its use on real data may be problematic in some cases. FWE performance depends on the accuracy of the spectral density estimation, which may be impaired by various time-localized patterns in the data diverging from the underlying FIEGARCH process due to a fourier base. Motivated by the advances in the spectral density estimation using wavelets, we propose a wavelet-based estimator, the wavelet Whittle estimator (WWE), as an alternative to FWE. As in the case of FWE, the WWE effectively overcomes the problem with the $|\Sigma_T|$ and $x_t'\Sigma_T^{-1}x_t$ by means of transformation. The difference is that instead of using discrete Fourier transform (DFT), it uses discrete wavelet transform (DWT).[2]

3.1 Wavelet Whittle estimator

Analogically to the FWE, we use the relationship between wavelet coefficients and the spectral density of x_t to approximate the likelihood function. The main advantage is, compared to the FWE, that the wavelets have limited support, and thus, the coefficients are not determined by the whole time series, but by a limited number of observations only. This increases the robustness of the resulting estimator to irregularities in the

2 For the reader's convenience, the discrete wavelet transform (DWT) is briefly introduced in an on-line appendix.

data well localized in time, such as jumps. These may be poorly detectable in the data, especially in the case of strong long memory that itself creates jump-like patterns, but at the same time, their presence can significantly impair the FWE performance. On the other hand, the main disadvantages of using the DWT are the restriction to sample lengths 2^j and the low number of coefficients at the highest levels of decomposition j.

Skipping the details of wavelet-based approximation of the covariance matrix and the detailed WWE derivation, which can be found e.g. in Percival and Walden (2000), the reduced-form wavelet-Whittle objective function can be defined as:

$$\mathcal{L}_{WW}(\zeta)=\ln|\Lambda_T|+(W'_{j,k}\Lambda_T^{-1}W_{j,k}) \tag{12}$$

$$=\sum_{j=1}^{J}\left[N_j\ln\left(2\int_{1/2^{j+1}}^{1/2^j}2^j f(\lambda,\zeta)d\lambda\right)+\sum_{k=1}^{N_J}\frac{W_{j,k}^2}{2\int_{1/2^{j+1}}^{1/2^j}2^j f(\lambda,\zeta)d\lambda}\right], \tag{13}$$

where $W_{j,k}$ are the wavelet (detail) coefficients, and Λ_T is a diagonal matrix with elements $\{C_1, C_1, \ldots, C_1, C_2, \ldots, C_J\}$, where for each level j, we have N_j elements $\left(C_j=2\int_{1/2^{j+1}}^{1/2^j}2^j f(\lambda,\zeta)d\lambda\right)$, where N_j is the number of DWT coefficients at level j. The wavelet Whittle estimator can then be defined as $\hat{\zeta}_{WW}=\underset{\zeta\in\Theta}{\operatorname{argmin}}\,\mathcal{L}_{WW}(\zeta)$.

Similarly to the fourier-based Whittle, the estimator is equivalent to a (Q)MLE of parameters in the probability density function of wavelet coefficients under normality assumption. The negative log-likelihood function can be rewritten as a sum of partial negative log-likelihood functions respective to individual levels of decomposition, whereas at each level, the coefficients are assumed to be homoskedastic, while across levels the variances differ. All wavelet coefficients are assumed to be (approximately) uncorrelated (the DWT approximately diagonalizes the covariance matrix), which requires an appropriate filter choice. Next, in our work the variance of scaling coefficients is excluded. This is possible due to the WWE construction, the only result is that the part of the spectrum respective to this variance is neglected in the estimation. This is optimal especially in cases of long-memory processes, where the spectral density goes to infinity at zero frequency, and where the sample variance of scaling coefficients may be significantly inaccurate estimate of its true counterpart due to the embedded estimation of the process mean.

3.2 Full vs. partial decomposition: a route to optimal decomposition level

Similarly to the omitted scaling coefficients, we can exclude any number of the sets of wavelet coefficients at the highest and/or lowest levels of decomposition. What we get is a parametric analogy to the local wavelet Whittle estimator (LWWE) developed in Wornell and Oppenheim (1992) and studied by Moulines, Roueff, and Taqqu (2008), who derive the asymptotic theory for LWWE with general upper and lower bound for levels of decomposition $\{j\in\langle L, U\rangle; 1\leq L\leq U\leq J\}$, where J is the maximal level of decomposition available given the sample length.

Although, in the parametric context, it seems to be natural to use the full decomposition, there are several features of the WWE causing that it may not be optimal. To see this, let's rewrite the WWE objective function as:

$$\mathcal{L}_{WW}(\zeta)=\sum_{j=1}^{J}N_j\left(\ln\sigma^2_{W,j,\mathrm{DWT}}(\zeta)+\frac{\hat{\sigma}^2_{W,j,\mathrm{DWT}}}{\sigma^2_{W,j,\mathrm{DWT}}(\zeta)}\right), \tag{14}$$

where $\sigma^2_{W,j,\mathrm{DWT}}(\zeta)$ is the theoretical variance of jth level DWT coefficients and $\hat{\sigma}^2_{W,j,\mathrm{DWT}}$ is its sample counterpart, ζ is the vector of parameters in $\sigma^2_{W,j,\mathrm{DWT}}(\zeta)$ and $\{W_{j,\mathrm{DWT}}; j=1, \ldots J\}$ are vectors of DWT coefficients used to calculate $\hat{\sigma}^2_{W,j,\mathrm{DWT}}$. Using the definition of wavelet variance $\upsilon_j^2=2\int_{1/2^{j+1}}^{1/2^j}f(\lambda,\zeta)d\lambda=\frac{\sigma^2_{W,j,\mathrm{DWT}}}{2^j}$; $j=1, 2, \ldots J$ and using the fact that the optimization problem does not change by dividing the right-hand side term by N^*, the total number of coefficients used in the estimation, the $\mathcal{L}_{WW}(\zeta)$ above is equivalent to

$$\mathcal{L}^*_{WW}(\zeta)=\sum_{j=1}^{J}\frac{N_j}{N^*}\left(\ln\sigma^2_{W,j,\mathrm{DWT}}(\zeta)+\frac{\hat{\upsilon}^2_{W,j,\mathrm{DWT}}}{\upsilon^2_{W,j,\mathrm{DWT}}(\zeta)}\right), \tag{15}$$

where $\upsilon^2_{W,j,\mathrm{DWT}}(\zeta)$ is the theoretical jth level wavelet variance and $\hat{\upsilon}^2_{W,j,\mathrm{DWT}}$ is its estimate using DWT coefficients.

The quality of our estimate of ζ depends on the the quality of our estimates of $\sigma^2_{W,j,\mathrm{DWT}}(\zeta)$ using sample variance of DWT coefficients, or equivalently, on the quality of our estimates of $\upsilon^2_{W,j,\mathrm{DWT}}(\zeta)$ using the rescaled sample variance of DWT coefficients, whereas each level of decomposition has a different weight (N_j/N^*) in the objective function. The weights reflect the number of DWT coefficients at individual levels of decomposition and, asymptotically, the width of the intervals of frequencies (scales) which they represent [i.e. the intervals $(2^{-(j+1)}, 2^{-j})$].

The problem, and one of the motivations for the partial decomposition, stems from the decreasing number of coefficients at subsequent levels of decomposition. With the declining number of coefficients, the averages of their squares are becoming poor estimates of their variances. Consequently, at these levels, the estimator is trying to match inaccurate approximations of the spectral density, and the quality of estimates is impaired. Then the full decomposition, that uses even the highest levels with just a few coefficients, may not be optimal. The importance of this effect should increase with the total energy concentrated at the lowest frequencies used for the estimation and with the level of inaccuracy of the variance estimates. To get a preliminary notion of the magnitude of the problem in the case of FIEGARCH model, see Table 1, and Figure 2 in Appendix A, where integrals of the spectral density (for several sets of coefficients) over intervals respective to individual levels are presented, together with the implied theoretical variances of the DWT coefficients. By their nature, the variances of the DWT coefficients reflect not only the shape of the spectral density (the integral of the spectral density multiplied by two), but also the decline in their number at subsequent levels (the 2^j term). This results in the interesting patterns observable in Figure 2, which suggest to think about both the direct effect of the decreasing number of coefficients on the variance estimates and about the indirect effect that changes their theoretical magnitudes. This indirect effect can be especially important in case of long-memory processes, where a significant portion of energy is located at low frequencies, the respective wavelet coefficients variances to be estimated become very high, while the accuracy of their estimates is poor. In general, dealing with this problem can be very important in case of small samples, where the share of the coefficients at "biased levels" is significant, but the effect should die out with increasing sample size.

One of the possible means of dealing with the latter problem is to use a partial decomposition, which leads to a local estimator similar to that in Moulines, Roueff, and Taqqu (2008). The idea is to set a minimal required number of coefficients at the highest level of decomposition considered in the estimation and discard all levels with lower number of coefficients. Under such a setting, the number of levels is increasing with the sample size, as in the case of full decomposition, but levels with small number of coefficients are cut off. According to Percival and Walden (2000), the convergence of the wavelet variance estimator is relatively fast, so that 128 (2^7) coefficients should already ensure a reasonable accuracy.[3] Though, for small samples (such as 2^9) this means a significant cut leading to estimation based on high frequencies only, which may cause even larger problems than the inaccuracy of wavelet variances estimates itself. The point is that every truncation implies a loss of information about the shape of the spectral density, whose quality depends on the accuracy of the estimates of wavelet variances. Especially for small samples, this means a tradeoff between inaccuracy due to poor variance estimation and inaccuracy due to insufficient level of decomposition. As far as our results for FIEGARCH model, based on partial decomposition suggest, somewhat inaccurate information may be still better than no information at all, and consequently, the use of truncation of six lags ensuring 128

3 Accuracy of the wavelet variance estimate, not the parameters in approximate MLE.

coefficients at the highest level of decomposition may not be optimal. The optimal level, will be discussed together with the experiment results.

Next possible solution to the problem can be based on a direct improvement of the variances estimates at the high levels of decomposition (low frequencies). Based on the theoretical results on wavelet variance estimation provided in Percival (1995) and summarized in Percival and Walden (2000), this should be possible by applying maximal overlap discrete wavelet transform (MODWT) instead of DWT. The main difference between the two transforms is that there is no sub-sampling in the case of MODWT. The number of coefficients at each level of decomposition is equal to the sample size, which can improve our estimates of the coefficients' variance. Generally, it is a highly redundant non-orthogonal transform, but in our case this is not an issue. Since the MODWT can be used for wavelet variance estimation, it can be used also for the estimation of the variances of DWT coefficients, and thus, it can be used as a substitute for the DWT in the WWE. Using the definitions of variances of DWT and MODWT coefficients at level j and their relation to the original data spectral density $f(\lambda, \zeta)$ described in Percival and Walden (2000)

$$\hat{\sigma}^2_{W,j,\mathrm{DWT}} = \frac{\sum_{k=1}^{N_j} W^2_{j,k,\mathrm{DWT}}}{N_j} = 2^{j+1}\int_{1/2^{j+1}}^{1/2^j} f(\lambda, \zeta)d\lambda \tag{16}$$

$$\hat{\sigma}^2_{W,j,\mathrm{MODWT}} = \frac{\sum_{k=1}^{T} W^2_{j,k,\mathrm{MODWT}}}{T} = 2\int_{1/2^{j+1}}^{1/2^j} f(\lambda, \zeta)d\lambda, \tag{17}$$

where $N_j = T/2^j$, it follows that

$$\hat{\sigma}^2_{W,j,\mathrm{DWT}} = 2^j \hat{\sigma}^2_{W,j,\mathrm{MODWT}}. \tag{18}$$

Then the MODWT-based approximation of the negative log-likelihood function can thus be defined as

$$\mathcal{L}^*_{WW,\mathrm{MODWT}} = \sum_{j=1}^{J} \frac{N_j}{N^*}\left(\ln \sigma^2_{W,j}(\zeta) + \frac{2^j \hat{\sigma}^2_{W,j,\mathrm{MODWT}}}{\sigma^2_{W,j}(\zeta)}\right), \tag{19}$$

and the MODWT-based WWE estimator as $\hat{\zeta}_{WW,\mathrm{MODWT}} = \operatorname*{argmin}_{\zeta\in\Theta} \mathcal{L}^*_{WW,\mathrm{MODWT}}$.

According to Percival (1995), in theory, the estimates of wavelet variance using MODWT can never be less efficient than those provided by the DWT, and thus the approach described above should improve the estimates.

Next interesting question related to the optimal level of decomposition concerns the possibility to make the estimation faster by using a part of the spectrum only. The idea is based on the shape of the spectral density determining the energy at every single interval of frequencies. As can be seen in Table 1 and Figure 2 in Appendix A, for FIEGARCH model, under a wide range of parameter sets most of the energy is concentrated at the upper intervals. Therefore, whenever it is reasonable to assume that the data-generating process is not an extreme case with parameters implying extremely strong long memory, estimation using a part of the spectrum only may be reasonable. In general, this method should be both better applicable and more useful in case of very long time-series compared to the short ones, especially when fast real-time estimation is required. In case of small samples the partial decomposition can be used as a simple solution to the inaccurate variance estimates at the highest levels of decomposition, but in most cases it is not reasonable to apply it just to speed up the estimation.

At this point the questions raised above represent just preliminary notions based mostly on common sense and the results of Moulines, Roueff, and Taqqu (2008) in the semi-parametric setup. To treat them properly, an asymptotic theory, in our case for the FIEGARCH-WWE, needs to be derived. This should enable to study all the various patterns in detail, decompose the overall convergence of the estimates into convergence with increasing sample size and convergence with increasing level of decomposition and to optimize the estimation setup respectively. Yet, leaving this for future research, we study the optimal decomposition with Monte Carlo simulations to see if we can provide any guidelines.

3.3 Fourier vs. wavelet approximation of spectral density

Since the relative accuracy of the Fourier- and wavelet-based spectral density estimates determine the relative performance of the parameters estimators, it is interesting to see how the sample Fourier- and wavelet-based approximations of the spectral density match its true shape. Figure 1 shows the true shape of a FIEGARCH spectral density under three different parameter sets, demonstrating the smoothness of this function and the importance of the long memory. Figure 1, then provides the wavelet-based approximations based on the simple assumption that the spectral density is constant over the whole intervals, equal to the estimated averages. Using this specification is relevant given the definition of the WWE. Wavelet-based approximations are compared with the respective true spectral densities, true averages of these spectral densities over intervals of frequencies, as well as with two Fourier-based approximations, one providing point estimates and the second estimating the averages over whole intervals. The figures show a good fit of both Fourier-based and wavelet-based approximations at most of the intervals, some problems can be seen at the lowest frequencies, which supports the idea of partial decomposition. In general, the wavelet-based approximation works well especially for processes with well behaved spectral densities without significant patterns well localized in the frequency domain, when the average energy over the whole intervals of frequencies represents a sufficient information about the shape of the true spectral density. For these processes, the wavelet transform can be effectively used for visual data analysis and both parametric and semi-parametric estimation of parameters in the spectral density function. More figures for the spectral density approximation are available in the online appendix.

4 Monte Carlo study: optimal decomposition

In order to study how the WWE performs compared to the two benchmark estimators (MLE and FWE), we have carried out a Monte Carlo experiment. Each round consisted of 1000 simulations of a FIEGARCH process at a fixed set of parameters, and estimation of these parameters by all methods of interest. To maintain coherency with previous results, our experiment setup mirrors that of Perez and Zaffaroni (2008). In addition, we need to make several choices concerning the WWE application, and we extend the setup with longer data sets as it may bring interesting insights [Jensen (1999), Percival and Walden (2000)]. Most importantly, we focus on studying the estimators that use partial decomposition to see, if we can gain some advantage from it.

4.1 Practical considerations for WWE application

First, using WWE, the same transformation of the data as in the case of the FWE is necessary. Second, due to the flexibility of the DWT, important choices have to be made before the WWE can be applied. The filters chosen for the Monte Carlo experiment in our work are the same as those chosen in Percival and Walden (2000), i.e. Haar wavelet, D4 (Daubechies) wavelet and LA8 (Least asymmetric) wavelet, but the need of a detailed study focusing on the optimal wavelet choice for FIEGARCH WWE is apparent. The only property of the filters that was tested before the estimation was their ability to decorrelate the FIEGARCH process, that is important for the WWE derivation and its performance [see Percival and Walden (2000), Jensen (1999), Jensen (2000) or Johnstone and Silverman (1997)]. Further, we assess the quality of the DWT-based decorrelation based on the dependencies among the resulting wavelet coefficients. We study estimates of autocorrelation functions (ACFs) of wavelet coefficients respective to FIEGARCH processes for ($T=2^{11}$; $d=0.25$,

$d = 0.45$, $d = -0.25$) and filters Haar, D4 and LA8. Both sample mean and 95% confidence intervals based on 500 FIEGARCH simulations are provided for each lag available.[4] Next, to avoid the problem with boundary coefficients, they are excluded from the analysis; sample sizes considered are: 2^k; $k = 9, 10, 11, 12, 13, 14$ and concerning the level of decomposition, both full and partial decomposition are used, the respective results are compared. Making all these choices, the WWE is fully specified and the objective function is ready for parameters estimation.

4.2 Results for partial decomposition

A look at comparison of the MLE, FWE and DWT-based WWE using Haar, D4 and LA8 wavelets and full decomposition tells us that WWE works fairly well in all setups, with smaller bias in comparison to FWE, although small loss in efficiency in terms of RMSE. The overall performance of the wavelet-based estimators (WWE using various filters) in the experiment suggests using D4 for 2^{10} and 2^{11} and switching to LA8 for 2^9 and $\{2^j; j > 11\}$ in case of long memory in the data (a simple ACF analysis before estimation should reveal this pattern). For negative dependence the optimal choice seems to be Haar for 2^9 and D4 otherwise (with possible shift to LA8 for samples longer than 2^{14}). While our aim is mainly in studying estimator with partial decomposition, we deffer these results to an online appendix.

Encouraged by the overall performance, we focus on varying number of levels used for the estimation. For all sample lengths of (2^M, $M = 9, 10, \ldots, 14$) experiments for levels $J \in (4, 5, \ldots M)$ have been carried out. Results are available for both processes with long memory ($d = 0.25$ and $d = 45$), which are of the most interest for practical applications, the case of $d = -0.25$ is omitted to keep the extent of simulations reasonable. Figures 4–6 show the main results. For the results including mean estimates, respective levels of bias and RMSE see tables in online appendix.

As the results suggest, for small samples with length of 2^9–2^{10}, estimation under the restriction to first five levels of decomposition leads to better estimates of both $d = 0.25$ and $d = 0.45$ in terms of both bias and RMSE in comparison to situation when full decomposition is used. With increasing sample size the performance of the estimator under partial decomposition deteriorates relatively to that using full decomposition. WWE also works better relatively to FWE for all filter specifications.

Comparing the performance of individual filters, in most cases LA8 provides the best sets of estimates for both $d = 0.25$ and $d = 0.45$, except for the case of 2^{10}–2^{13} sample sizes with $d = 0.25$, where D4 seems to be preferred.

We conclude that the results well demonstrate the effects mentioned when discussing the partial decomposition in 3.2. We can see how the partial decomposition helps in the case of short samples and how the benefits from truncation (no use of inaccurate information) decrease relative to the costs (more weight on the high-frequency part of the spectra and no information at all about the spectral density shape at lower frequencies) as the sample size increases, as the long-memory strengthens and as the truncation becomes excessive. Moreover, the effect becomes negligible with longer samples, as the share of problematic coefficients goes to zero. This is highlighted by Figure 3, where the approximation of the spectral density by wavelets is compared to fourier transform. In small samples, approximation is more precise supporting our findings.

Yet, the optimal setup choice for small samples is a non-trivial problem that cannot be reduced to a simple method of cutting a fixed number of highest levels of decomposition to ensure some minimal number of coefficients at each level. Although in case of long samples a nice convergence with both sample size and level of decomposition can be seen for all specifications, the results for small samples are mixed. In this latter case the convergence with sample size still works relatively well, but the increase in level of decomposition does not always improve the estimates.

4 These results can be found in the online appendix.

5 Monte Carlo study: jumps and forecasting

Although WWE does not seem to significantly outperform other estimation frameworks in the simple simulation setting, we should not make premature conclusions. Wavelets have been used successfully in detection of jumps in the literature (Fan and Wang 2007; Xue, Gençay, and Fagan 2014; Barunik and Vacha 2015, 2016; Barunik, Krehlik, and Vacha 2016), hence we assume more realistic scenario for data generating process including jumps. Since the evaluation based on individual parameters estimation only may not be the best practice when forecasting is the main concern, we analyze also the relative forecasting performance.

5.1 FIEGARCH-jump model

Jumps are one of the several well known stylized features of log-returns and/or realized volatility time series and there is a lot of studies on incorporating this pattern in volatility models [for a discussion see e.g. Mancini and Calvori (2012)].

To test the performance of the individual estimators in the case of FIEGARCH-Jump processes, an additional Monte Carlo experiment has been conducted. The simulations are augmented by additional jumps, which do not enter the conditional volatility process, but the log-returns process only. This represents the situation, when the jumps are not resulting from the long memory in the volatility process, which can produce patterns similar to jumps in some cases, as well as they do not determine the volatility process in any way. The log-return process is then specified as:

$$\varepsilon_t = z_t h_t^{1/2} + J_t(\lambda), \tag{20}$$

where the process ht remains the same as in the original FIEGARCH model (Eq. 1) and $J_t; t=1, 2, \dots, T$ is a Jump process modeled as a sum of intraday jumps, whereas the number of intraday jumps in 1 day follows a Poisson process with parameter $\lambda = 0.028$ and their size is drawn from a normal distribution $N(0, 0.2)$. The Jump process is based on Mancini and Calvori (2012), with parameters slightly adjusted [originally $\lambda = 0.014$ and sizes follow $N(0, 0.25)$] based on analysis of resulting simulations and comparison with real data. Moreover, unlike in the previous Monte Carlo experiment, a non-zero constant is assumed. Since we would like to keep consistency in the research (keep the parameters the same throughout this paper) and at the same time to simulate time series as close to the real ones as possible, we have compared our simulated time series with real data and found a good match.

5.2 Forecasting

For each simulation the fitted values and a 1 day ahead forecast per each estimator are calculated. We present the mean error, mean absolute deviation and root mean squared error for both the in-sample and out-of-sample forecasts from 1000 simulations.

Before we report the results, there are two important issues to be discussed. First, one needs to note that the process entering the estimation is transformed by logarithm, hence the magnitute of jumps became very small relative to the process. Second, to obtain unbiased parameter estimates, we need to first detect and extract the jumps from the process. To deal with the jumps we apply one of the well performing wavelet-based jump estimators that is based on a universal threshold of Donoho and Johnstone (1994) and that is described in detail and successfully applied in Fan and Wang (2007), and Barunik and Vacha (2015) to detect jumps in high frequency data, and further utilized in Barunik, Krehlik, and Vacha (2016) in forecasting and Barunik and Vacha (2016) in co-jumps detection. When detected, the jumps are replaced by average of the two adjacent values. This, of course, is not the best practice in case of large double-jumps, where this transformation leads to two smaller jumps instead of getting rid of them. Yet, in case of long memory that can produce jump-like patterns, which are usually clustered in high volatility intervals, getting rid of the multiple jumps may not be the best alternative. So we use this simple transform for our data with moderate jumps. Thus, it is important to distinguish between the jump detection and model estimation as two separable tasks.

5.3 Results: FIEGARCH-Jump

The results of the Monte Carlo experiment with adding jumps to the process are summarized in Tables 2–8. Tables 2 and 3 compare MLE, FWE and MODWT-based WWE in terms of individual parameters estimation performance. Note we report MODWT instead of DWT in forecasting excercise, as the overall performance of the MODWT-WWE is better than that of the DWT-WWE both in terms of bias and RMSE and considering also the loss of sample size limitation, the MODWT-WWE is preferred.

Next, focusing on the MLE, FWE and MODWT-WWE relative performance in terms of RMSE for jumps and $d=0.25$, the MLE, despite being affected by the residual jump effects remains the best estimator followed by the two Whittles, which perform comparably, with FWE delivering slightly better results. Yet, the bias of the MLE is significant and we would prefer the use of FWE considering both the bias and the RMSE. Moreover, in case of longer time series, WWE seems to be the best option due to the faster bias decay. Next, for $d=0.45$, the MLE performance is very poor and the use of WE is preferable. As expected, the bias and RMSE in case of individual parameters estimates as well as the mean absolute deviation and RMSE of the out-of-sample forecasts decline and the overall in-sample fit improves with sample size increase and long memory weakening. Next, the constant term estimation performance is worth mentioning, since it is very poor in the case of MLE and strong long memory, and therefore an ex ante estimation as in the case of FWE and WWE is appropriate.

On the other hand, when we look at the forecasting performance, the results are much more clear. The best in all scenarios and by all indicators is the MLE, followed by the FWE and a little less accurate WWE. The impact of jumps depends, of course, on the jump estimator performance and in our case, for forecasting, it is very limited, although the same cannot be said about the impact on individual parameters estimates.

6 Conclusion

In this paper, we introduce a new, wavelet-based estimator (wavelet Whittle estimator, WWE) of a FIEGARCH model, ARCH-family model allowing for long-memory and asymmetry in volatility, and study its properties. Based on several Monte Carlo experiments its accuracy and empirical convergence are examined, as well as its relative performance with respect to two traditional estimators: Fourier-based Whittle estimator (FWE) and maximum likelihood estimator (MLE). It is shown that even in the case of simulated pure FIEGARCH processes, which do not allow to fully utilize the advantages of the WWE, the estimator can work reasonably well. In terms of bias, it often outperforms the FWE, while in terms of RMSE the FWE is better. Yet, the absolute differences are usually small. As expected, MLE in most casest performs best in terms of efficiency. The Whittle estimators outperform the MLE in some cases, usually in situations with negative memory. The forecasting performance analysis has a similar conclusion, yielding the differences across estimators even smaller. Yet, since the Whittle estimators are significantly faster and the differences in the performance are small, they are an attractive alternative to the MLE for large samples. Concerning the optimal WWE settings studied, the strength of long memory, sample size and parameter concerned seem to be important for the optimal filter (wavelet) choice.

Next, practical aspects of the WWE application are discussed. The main focus is on the problem of declining number of wavelet coefficients at subsequent levels of decomposition, which impairs the estimates accuracy. Two solutions to this problem are suggested. One is based on a partial decomposition (parametric counterpart to local WWE) the other applies an alternative specification of the WWE (using maximal overlap discrete wavelet transform, MODWT). We show that the partial decomposition can improve the estimates in case of short samples, and make the WWE superior to the FWE (and to the MLE for negative memory), while in case of large samples, full decomposition is more appropriate. Yet, the second solution (MODWT-WWE) is argued to be better. Compared to the former method, it ensures the number of coefficients at every level equal to the sample size and does not lead to any decline in the share of spectrum used in the estimation (information loss). The only cost to bear is a somewhat longer estimation time. As our results suggest, using the MODWT instead of the DWT improves the WWE performance in all scenarios.

In addition, we study the properties of estimators under the presence of jumps in the processes. The accuracy of individual parameters estimates using MLE is significantly impaired, even if we apply a simple data correction; the FWE and the WWE are superior. Yet, based on the forecasting performance, MLE should be preferred in all scenarios at least in case of small samples, where it can be computed in reasonable time; FWE and WWE can be recommended only as faster alternatives.

It can be concluded that after optimization of the estimation setup, the WWE may become a very attractive alternative to the traditional estimation methods. It provides a robust alternative to time-localized irregularities in data. In small samples, due to more precise approximation of spectral density, wavelet-based Whittle estimation delivers better parameter estimates.

Acknowledgments: We would like to express our gratitude to Ana Perez, who provided us with the code for MLE and FWE estimation of FIEGARCH processes, and we gratefully acknowledge financial support from the the Czech Science Foundation under project No. 13-32263S. The research leading to these results has received funding from the European Unions Seventh Framework Programme (FP7/2007-2013) under grant agreement No. FP7-SSH- 612955 (FinMaP).

A Appendix: Tables and Figures

Table 1: Energy decomposition.

Coefficients	d	ω	α	β	θ	γ
(a) Coefficient sets						
A	0.25	0	0.5	0.5	−0.3	0.5
B	0.45	0	0.5	0.5	−0.3	0.5
C	−0.25	0	0.5	0.5	−0.3	0.5
D	0.25	0	0.9	0.9	−0.3	0.5
E	0.45	0	0.9	0.9	−0.3	0.5
F	−0.25	0	0.9	0.9	−0.3	0.5
G	0.25	0	0.9	0.9	−0.9	0.9
H	0.45	0	0.9	0.9	−0.9	0.9

	A	B	C	D	E	F	G	H
(b) Integrals over frequencies respective to levels for the coefficient sets from Table 1								
Level 1	1.1117	1.1220	1.0897	1.1505	1.1622	1.1207	1.1261	1.1399
Level 2	0.5473	0.5219	0.6274	0.4776	0.4691	0.5306	0.6187	0.6058
Level 3	0.3956	0.3693	0.4330	0.3246	0.3056	0.3959	1.1354	1.3453
Level 4	0.3029	0.3341	0.2425	0.5559	0.7712	0.3528	2.9558	4.8197
Level 5	0.2035	0.2828	0.1175	1.0905	2.1758	0.3003	6.0839	13.2127
Level 6	0.1279	0.2297	0.0550	1.4685	3.9342	0.1965	8.2136	23.4144
Level 7	0.0793	0.1883	0.0259	1.3523	4.7975	0.0961	7.6026	28.4723
Level 8	0.0495	0.1584	0.0123	1.0274	4.8302	0.0408	5.8268	28.7771
Level 9	0.0313	0.1368	0.0059	0.7327	4.5720	0.0169	4.1967	27.3822
Level 10	0.0201	0.1206	0.0029	0.5141	4.2610	0.0071	2.9728	25.6404
Level 11	0.0130	0.1080	0.0014	0.3597	3.9600	0.0030	2.0977	23.9192
Level 12	0.0086	0.0979	0.0007	0.2518	3.6811	0.0013	1.4793	22.2986

Table 1 (continued)

	A	B	C	D	E	F	G	H
(c) Sample variances of DWT Wavelet Coefficients for the coefficient sets from Table 1								
Level 1	4.4468	4.4880	4.3588	4.6020	4.6488	4.4828	4.5044	4.5596
Level 2	4.3784	4.1752	5.0192	3.8208	3.7528	4.2448	4.9496	4.8464
Level 3	6.3296	5.9088	6.9280	5.1936	4.8896	6.3344	18.1664	21.5248
Level 4	9.6928	10.6912	7.7600	17.7888	24.6784	11.2896	94.5856	154.2304
Level 5	13.0240	18.0992	7.5200	69.7920	139.2512	19.2192	389.3696	845.6128
Level 6	16.3712	29.4016	7.0400	187.9680	503.5776	25.1520	1051.3408	2997.0432
Level 7	20.3008	48.2048	6.6304	346.1888	1228.1600	24.6016	1946.2656	7288.9088
Level 8	25.3440	81.1008	6.2976	526.0288	2473.0624	20.8896	2983.3216	14733.8752
Level 9	32.0512	140.0832	6.0416	750.2848	4681.7280	17.3056	4297.4208	28039.3728
Level 10	41.1648	246.9888	5.9392	1052.8768	8726.5280	14.5408	6088.2944	52511.5392
Level 11	53.2480	442.3680	5.7344	1473.3312	16220.1600	12.2880	8592.1792	97973.0432
Level 12	70.4512	801.9968	5.7344	2062.7456	30155.5712	10.6496	12118.4256	182670.1312

Table 2: Monte Carlo No jumps: d = 0.25/0.45; MLE, FWE, MODWT(D4), 2: Corrected using Donoho and Johnstone threshold.

PAR	True	Method	No jumps; N = 2048			No jumps; N = 16,384			PAR	No jumps; N = 2048		
			Mean	Bias	RMSE	Mean	Bias	RMSE		Mean	Bias	RMSE
$\hat{d}$	0.250	WWE MODWT	0.165	−0.085	0.225	0.253	0.003	0.042	0.450	0.362	−0.088	0.170
		WWE MODWT 2	0.168	−0.082	0.227	–	–	–		–	–	–
		FWE	0.212	−0.038	0.147	0.251	0.001	0.036		0.415	−0.035	0.087
		FWE 2	0.213	−0.037	0.146	–	–	–		–	–	–
		MLE	0.220	−0.030	0.085	–	–	–		0.433	−0.017	0.043
		MLE 2	0.228	−0.022	0.086	–	–	–		–	–	–
$\hat{\omega}$	−7.000	MLE	−7.076	−0.076	0.174	–	–	–		−7.458	−0.458	0.739
		MLE 2	−7.083	−0.083	0.182	–	–	–		–	–	–
		OTHER	−7.002	−0.002	0.197	−7.003	−0.003	0.074		−6.999	0.001	0.696
		OTHER 2	−7.015	−0.015	0.198	–	–	–		–	–	–
$\hat{\alpha}_2$	0.500	WWE MODWT	0.434	−0.066	0.349	0.328	−0.172	0.229		0.324	−0.176	0.395
		WWE MODWT 2	0.426	−0.074	0.358	–	–	–		–	–	–
		FWE	0.527	0.027	0.343	0.512	0.012	0.168		0.475	−0.025	0.348
		FWE 2	0.521	0.021	0.333	–	–	–		–	–	–
		MLE	0.503	0.003	0.121	–	–	–		0.487	−0.013	0.128
		MLE 2	0.464	−0.036	0.136	–	–	–		–	–	–
$\hat{\beta}_1$	0.500	WWE MODWT	0.559	0.059	0.249	0.523	0.023	0.078		0.610	0.110	0.178
		WWE MODWT 2	0.561	0.061	0.253	–	–	–		–	–	–
		FWE	0.520	0.020	0.199	0.499	−0.001	0.065		0.554	0.054	0.135
		FWE 2	0.517	0.017	0.214	–	–	–		–	–	–
		MLE	0.529	0.029	0.101	–	–	–		0.527	0.027	0.063
		MLE 2	0.537	0.037	0.109	–	–	–		–	–	–
$\hat{\theta}$	−0.300	WWE MODWT	−0.283	0.017	0.180	−0.337	−0.037	0.078		−0.314	−0.014	0.146
		WWE MODWT 2	−0.261	0.039	0.182	–	–	–		–	–	–
		FWE	−0.244	0.056	0.182	−0.279	0.021	0.077		−0.242	0.058	0.158
		FWE 2	−0.222	0.078	0.189	–	–	–		–	–	–
		MLE	−0.301	−0.001	0.026	–	–	–		−0.301	−0.001	0.024
		MLE 2	−0.282	0.018	0.031	–	–	–		–	–	–
$\hat{\gamma}$	0.500	WWE MODWT	0.481	−0.019	0.196	0.489	−0.011	0.085		0.504	0.004	0.218
		WWE MODWT 2	0.472	−0.028	0.193	–	–	–		–	–	–
		FWE	0.509	0.009	0.175	0.504	0.004	0.083		0.526	0.026	0.202
		FWE 2	0.497	−0.003	0.174	–	–	–		–	–	–
		MLE	0.499	−0.001	0.045	–	–	–		0.507	0.007	0.044
		MLE 2	0.491	−0.009	0.048	–	–	–		–	–	–

Table 3: Monte Carlo jumps: d = 0.25/0.45; Poisson lambda = 0.028; N(0; 0.2); FWE, MODWT (D4), 2: Corrected using Donoho and Johnstone threshold.

PAR	True	Method	Jump [0.028, N(0, 0.2)]; N=2048			Jump [0.028, N(0, 0.2)]; N=16,384			PAR	Jump [0.028, N(0, 0.2)] N=2048		
			Mean	Bias	RMSE	Mean	Bias	RMSE		Mean	Bias	RMSE
$\hat{d}$	0.250	WWE MODWT	0.145	−0.105	0.214	–	–	–	0.450	–	–	–
		WWE MODWT 2	0.154	−0.096	0.225	0.235	−0.015	0.042		0.347	−0.103	0.179
		FWE	0.195	−0.055	0.142	–	–	–		–	–	–
		FWE 2	0.206	−0.044	0.143	0.231	−0.019	0.038		0.403	−0.047	0.091
		MLE	0.018	−0.232	0.353	–	–	–		–	–	–
		MLE 2	0.099	−0.151	0.251	–	–	–		0.187	−0.263	0.314
$\hat{\omega}$	−7.000	MLE	−5.662	1.338	1.450	–	–	–		–	–	–
		MLE 2	−6.282	0.718	0.801	–	–	–		−5.529	1.471	1.662
		OTHER	−6.887	0.113	0.221	–	–	–		–	–	–
		OTHER 2	−6.942	0.058	0.203	−6.941	0.059	0.096		−6.946	0.054	0.677
$\hat{\alpha}_2$	0.500	WWE MODWT	0.475	−0.025	0.437	–	–	–		–	–	–
		WWE MODWT 2	0.492	−0.008	0.402	0.557	0.057	0.243		0.390	−0.110	0.447
		FWE	0.561	0.061	0.454	–	–	–		–	–	–
		FWE 2	0.582	0.082	0.400	0.731	0.231	0.297		0.535	0.035	0.428
		MLE	0.667	0.167	0.385	–	–	–		–	–	–
		MLE 2	0.652	0.152	0.287	–	–	–		0.605	0.105	0.308
$\hat{\beta}_1$	0.500	WWE MODWT	0.592	0.092	0.290	–	–	–		–	–	–
		WWE MODWT 2	0.578	0.078	0.264	0.535	0.035	0.087		0.636	0.136	0.203
		FWE	0.546	0.046	0.266	–	–	–		–	–	–
		FWE 2	0.529	0.029	0.231	0.519	0.019	0.068		0.579	0.079	0.164
		MLE	0.406	−0.094	0.452	–	–	–		–	–	–
		MLE 2	0.503	0.003	0.250	–	–	–		0.619	0.119	0.240
$\hat{\theta}$	−0.300	WWE MODWT	−0.491	−0.191	0.272	–	–	–		–	–	–
		WWE MODWT 2	−0.385	−0.085	0.189	−0.398	−0.098	0.108		−0.384	−0.084	0.174
		FWE	−0.455	−0.155	0.246	–	–	–		–	–	–
		FWE 2	−0.348	−0.048	0.175	−0.356	−0.056	0.065		−0.324	−0.024	0.153
		MLE	−0.214	0.086	0.130	–	–	–		–	–	–
		MLE 2	−0.211	0.089	0.104	–	–	–		−0.176	0.124	0.137
$\hat{\gamma}$	0.500	WWE MODWT	0.203	−0.297	0.365	–	–	–		–	–	–
		WWE MODWT 2	0.322	−0.178	0.271	0.287	−0.213	0.231		0.276	−0.224	0.322
		FWE	0.257	−0.243	0.315	–	–	–		–	–	–
		FWE 2	0.365	−0.135	0.231	0.313	−0.187	0.202		0.317	−0.183	0.291
		MLE	0.287	−0.213	0.256	–	–	–		–	–	–
		MLE 2	0.340	−0.160	0.180	–	–	–		0.347	−0.153	0.179

Table 4: Forecasting: N = 2048; No jumps; MLE, FWE, MODWT(D4), DWT(D4), 2: Corrected using Donoho and Johnstone threshold.

Cat	Method	Main stats			MAD quantiles			
		Mean err	MAD	RMSE	0.50	0.90	0.95	0.99
In	WWE MODWT	6.1308e−05	0.00039032	0.0052969	–	–	–	–
	WWE MODWT 2	2.2383e−05	0.00040778	0.0034541	–	–	–	–
	WWE DWT	8.3135e−05	0.00044932	0.011577	–	–	–	–
	WWE DWT 2	2.363e−05	0.00043981	0.0050438	–	–	–	–
	FWE	5.9078e−05	0.00037064	0.00087854	–	–	–	–
	FWE 2	1.4242e−05	0.00038604	0.0011961	–	–	–	–
	MLE	6.0381e−06	9.3694e−05	0.00019804	–	–	–	–
	MLE 2	−3.3242e−05	0.00011734	0.00028776	–	–	–	–

Table 4 (continued)

Cat	Method	Main stats			MAD quantiles			
		Mean err	**MAD**	**RMSE**	**0.50**	**0.90**	**0.95**	**0.99**
Out	WWE MODWT	105.3851	105.3856	3277.1207	0.00015361	0.0010525	0.0019431	0.0060853
	WWE MODWT 2	−7.6112e−05	0.00058276	0.0020436	0.00016482	0.0010512	0.0020531	0.0072711
	WWE DWT	0.00013817	0.00066928	0.0031579	0.00017219	0.0012156	0.0020541	0.0065611
	WWE DWT 2	0.00087082	0.0015663	0.027558	0.00017181	0.0012497	0.0022271	0.0072191
	FWE	2.9498e−05	0.00050763	0.0015745	0.00014566	0.0010839	0.0018926	0.005531
	FWE 2	0.00038499	0.0010395	0.014191	0.00015425	0.0011351	0.001989	0.0081017
	MLE	−1.5041e−06	0.00012211	0.00035147	4.0442e−05	0.00024483	0.00043587	0.001595
	MLE 2	−0.00010455	0.00024125	0.0015605	4.3599e−05	0.00030553	0.00063378	0.0038043
		Error quantiles A				**Error quantiles B**		
		0.01	**0.05**	**0.10**		**0.90**	**0.95**	**0.99**
Out	WWE MODWT	−0.0033827	−0.0010722	−0.0004865		0.00063385	0.0010345	0.0048495
	WWE MODWT 2	−0.0045312	−0.0013419	−0.00062698		0.00053531	0.00089684	0.0051791
	WWE DWT	−0.0040691	−0.001223	−0.00059588		0.00065501	0.0012118	0.004838
	WWE DWT 2	−0.003994	−0.0013952	−0.00071166		0.00059679	0.0010616	0.0050457
	FWE	−0.0035752	−0.0010712	−0.00053169		0.00061822	0.001086	0.004636
	FWE 2	−0.0042148	−0.00129	−0.00063194		0.0005657	0.0010577	0.0048622
	MLE	−0.00079412	−0.00024382	−0.00013312		0.00010297	0.00026481	0.0013195
	MLE 2	−0.0019587	−0.00046351	−0.00021734		7.5622e−05	0.00018216	0.00077541

Not corrected: Mean of N-next true values to be forecasted: 0.0016845; Total valid = 967, i.e. 96.7% of M fails-MLE = 0%, fails-FWE = 0.8%, fails-MODWT = 2.1%, fails-DWT = 1.5%, Corrected: Mean of N-next true values to be forecasted: 0.0016852; Total valid = 959, i.e. 95.9% of M fails-MLE = 0%, fails-FWE = 1.3%, fails-MODWT = 1.7%, fails-DWT = 2.7%.

Table 5: Forecasting: N = 16,384, No Jumps; MLE, FWE, MODWT(D4), DWT(D4), 2: Corrected using Donoho and Johnstone threshold.

Cat	Method	Main stats			MAD quantiles			
		Mean err	**MAD**	**RMSE**	**0.50**	**0.90**	**0.95**	**0.99**
In	WWE MODWT	7.522e−06	0.00015889	0.00032423	–	–	–	–
	WWE DWT	8.5378e−06	0.00017736	0.00038026	–	–	–	–
	FWE	6.3006e−06	0.00014367	0.00032474	–	–	–	–
	MLE	–	–	–	–	–	–	–
Out	WWE MODWT	9.3951e−06	0.00018394	0.00054618	7.1556e−05	0.00040438	0.0007369	0.0015565
	WWE DWT	2.1579e−05	0.0002107	0.00084705	7.1483e−05	0.00041876	0.00074376	0.0018066
	FWE	−3.3569e−06	0.00017794	0.00067805	5.6684e−05	0.00035132	0.0005922	0.001811
	MLE	–	–	–	–	–	–	–
		Error quantiles A				**Error quantiles B**		
		0.01	**0.05**	**0.10**		**0.90**	**0.95**	**0.99**
Out	WWE MODWT	−0.0011566	−0.00040454	−0.00021569		0.00021003	0.0004025	0.0013515
	WWE DWT	−0.0011033	−0.00038209	−0.00018247		0.00023587	0.00049025	0.0016387
	FWE	−0.00087002	−0.00034408	−0.00018571		0.00018593	0.00035741	0.0010787
	MLE	–	–	–		–	–	–

Not corrected: Mean of N-next true values to be forecasted: 0.0015516; Total valid = 1000, i.e. 100% of M fails-MLE = 0%, fails-FWE = 0%, fails-MODWT = 0%, fails-DWT = 0%.

Table 6: Forecasting: N = 2048; d = 0.45; No jumps; MLE, FWE, MODWT(D4), DWT(D4), 2: Corrected using Donoho and Johnstone threshold.

Cat	Method	Main stats			MAD quantiles			
		Mean err	MAD	RMSE	0.50	0.90	0.95	0.99
In	WWE MODWT	0.00026281	0.0010841	0.02023	–	–	–	–
	WWE DWT	0.00022639	0.0011206	0.013151	–	–	–	–
	FWE	0.00027127	0.0010458	0.005243	–	–	–	–
	MLE	5.7279e–05	0.00026995	0.0012167	–	–	–	–
Out	WWE MODWT	Inf	Inf	Inf	0.00016653	0.0024597	0.005308	0.040031
	WWE DWT	924.8354	924.8375	28648.7373	0.00017788	0.0024428	0.0049679	0.040403
	FWE	−0.00010684	0.0015807	0.0078118	0.00016022	0.0025471	0.0057388	0.031548
	MLE	0.0002289	0.0004843	0.0039972	4.2589e–05	0.00052307	0.0010187	0.0078509
		Error quantiles A				Error quantiles B		
		0.01	0.05	0.10		0.90	0.95	0.99
Out	WWE MODWT	−0.013427	−0.0024269	−0.00087296		0.0010521	0.0026019	0.013128
	WWE DWT	−0.013075	−0.0025811	−0.0010018		0.00089545	0.0023095	0.0165
	FWE	−0.012356	−0.002209	−0.00081063		0.0010042	0.002777	0.014773
	MLE	−0.0016025	−0.00044789	−0.0002568		0.00017179	0.00056713	0.0051968

Not corrected: Mean of N-next true values to be forecasted: 0.0064152; Total valid = 962, i.e. 96.2% of M fails-MLE = 0%, fails-FWE = 1%, fails-MODWT = 1.9%, fails-DWT = 2.1%.

Table 7: Forecasting: N = 2048; Jumps lambda = 0.028, N(0, 0.2); MLE, FWE, MODWT(D4), DWT(D4), 2: Corrected using Donoho and Johnstone threshold.

Cat	Method	Main stats			MAD quantiles			
		Mean err	MAD	RMSE	0.50	0.90	0.95	0.99
In	WWE MODWT	0.0024292	0.0027027	0.031109	–	–	–	–
	WWE MODWT 2	0.00049847	0.00088238	0.010266				
	WWE DWT	0.0022873	0.0025833	0.029094	–	–	–	–
	WWE DWT 2	0.00051788	0.00092081	0.013946				
	FWE	0.0024241	0.0026398	0.030474	–	–	–	–
	FWE 2	0.00046896	0.00080786	0.01562				
	MLE	0.00099962	0.0013136	0.0022127	–	–	–	–
	MLE 2	0.00021708	0.0005767	0.0011708				
Out	WWE MODWT	Inf	Inf	Inf	0.00027911	0.0019584	0.0043937	0.074726
	WWE MODWT 2	12837.4644	12837.4647	361761.8976	0.00020755	0.0012984	0.0021954	0.064956
	WWE DWT	0.010776	0.010967	0.14233	0.00032328	0.0020108	0.0048951	0.086811
	WWE DWT 2	Inf	Inf	Inf	0.00019516	0.0013594	0.002609	0.08235
	FWE	0.0098899	0.010026	0.15737	0.00025416	0.0019525	0.0047286	0.073048
	FWE 2	1.4046	1.4048	36.7106	0.00017622	0.0012063	0.0024169	0.057823
	MLE	0.0014788	0.0017573	0.0066777	0.00099906	0.001951	0.0027302	0.019721
	MLE 2	0.0022114	0.0025386	0.053463	0.00034636	0.00094187	0.0016393	0.0082677

Table 7 (continued)

		Error quantiles A			Error quantiles B		
		0.01	**0.05**	**0.10**	**0.90**	**0.95**	**0.99**
Out	WWE MODWT	−0.0014262	−0.00061569	−0.00023517	0.0018119	0.0043937	0.074726
	WWE MODWT 2	−0.002004	−0.0007648	−0.00033609	0.0010397	0.0018978	0.064956
	WWE DWT	−0.0014902	−0.00065937	−0.00028371	0.0018904	0.0048951	0.086811
	WWE DWT 2	−0.002635	−0.00076962	−0.00031466	0.00099973	0.0020434	0.08235
	FWE	−0.00097176	−0.00039409	−0.00021056	0.0019269	0.0047286	0.073048
	FWE 2	−0.0019851	−0.00057545	−0.00025878	0.00087435	0.0020268	0.057823
	MLE	−0.0033888	−0.0004285	0.00016064	0.0018323	0.0023536	0.019688
	MLE 2	−0.0030739	−0.00075814	−0.00029836	0.00066214	0.00099507	0.0059062

Corrected: Mean of N-next true values to be forecasted: 0.001324; Total valid = 801, i.e. 80.1% of M fails-MLE = 0%, fails-FWE = 7%, fails-MODWT = 13.4%, fails-DWT = 12.5% Not Corrected: Mean of N-next true values to be forecasted: 0.001324; Total valid = 45.1, i.e. % of M fails-MLE = 0%, fails-FWE = 35.2%, fails-MODWT = 43.4%, fails-DWT = 41.6%.

Table 8: Forecasting: N = 16,384, Jumps lambda = 0.028, N(0, 0.2); MLE, FWE, MODWT(D4), DWT(D4), 2: Corrected using Donoho and Johnstone threshold.

Cat	**Method**	**Main stats**			**MAD quantiles**			
		Mean err	**MAD**	**RMSE**	**0.50**	**0.90**	**0.95**	**0.99**
In	WWE MODWT	0.00079621	0.0010536	0.019562	–	–	–	–
	WWE DWT	0.00074677	0.0010165	0.018346	–	–	–	–
	FWE	0.00052705	0.00074759	0.0094745	–	–	–	–
	MLE	–	–	–	–	–	–	–
Out	WWE MODWT	Inf	Inf	Inf	0.0002394	0.0015	0.0037696	0.088609
	WWE DWT	Inf	Inf	Inf	0.00022172	0.0016482	0.0039952	0.043464
	FWE	0.010034	0.010249	0.23919	0.00017951	0.0014685	0.0037604	0.04528
	MLE	–	–	–	–	–	–	–
		Error quantiles A				**Error quantiles B**		
		0.01	**0.05**	**0.10**		**0.90**	**0.95**	**0.99**
Out	WWE MODWT	−0.0025009	−0.0006746	−0.00028513		0.0011776	0.0033089	0.088609
	WWE DWT	−0.0025573	−0.00057909	−0.00028814		0.0012299	0.0033547	0.043464
	FWE	−0.0018013	−0.00048645	−0.00022585		0.0011465	0.0030639	0.04528
	MLE	–	–	–		–	–	–

Corrected: Mean of N-next true values to be forecasted: 0.0016747; Total valid = 948, i.e. 94.8% of M fails-MLE = – %, fails-FWE = 0.4%, fails-MODWT = 3.1%, fails-DWT = 2.5%.

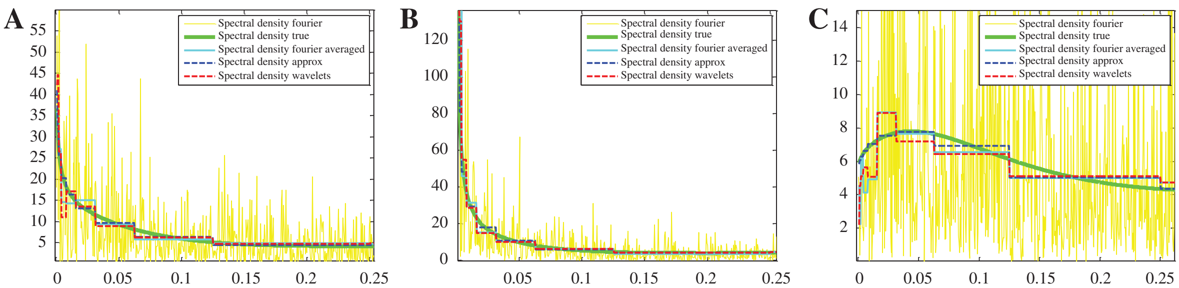

Figure 1: Spectral density estimation ($d=0.25/0.45/-0.25$), T = 2048 (2^{11}), level = 10, zoom. (A) $d=0.25$. (B) $d=0.45$. (C) $d=-0.25$.

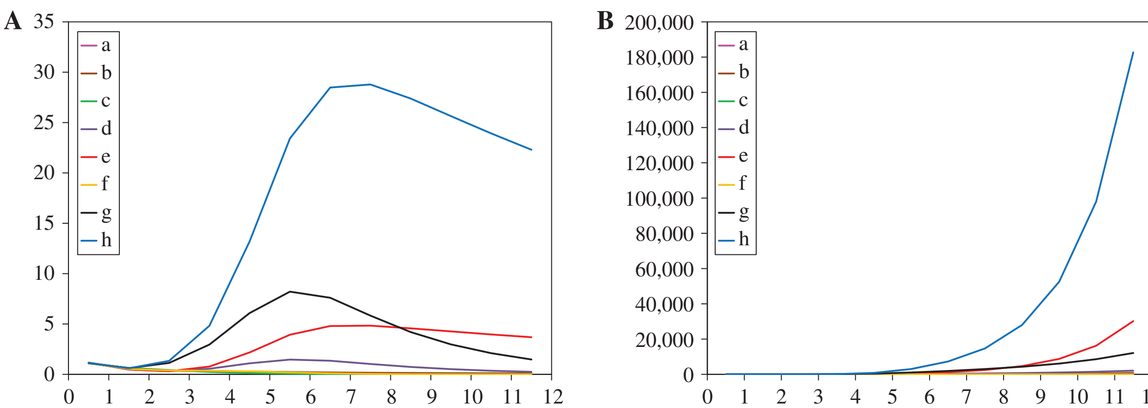

Figure 2: Energy decomposition: (A) Integrals of FIEGARCH spectral density over frequency intervals, and (B) true variances of wavelet coefficients respective to individual levels of decomposition, assuming various levels of long memory ($d=0.25$, $d=0.45$, $d=-0.25$) and the coefficient sets from Table 1.

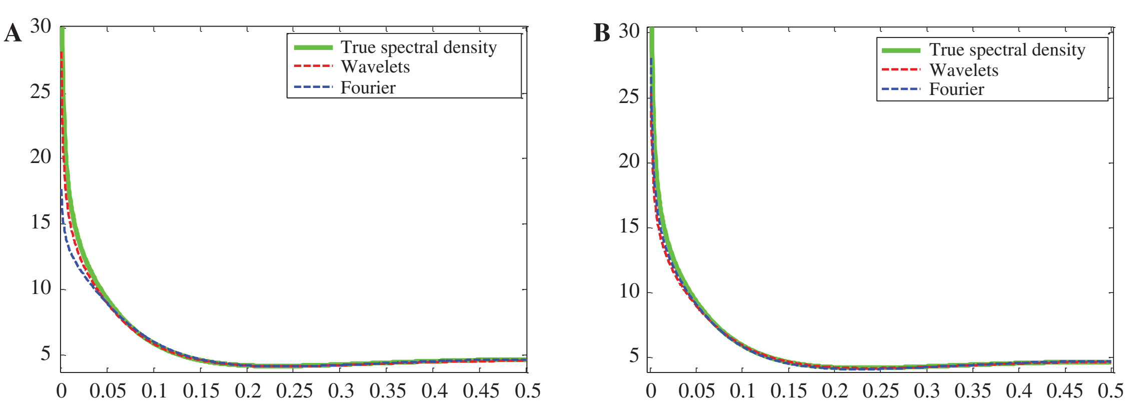

Figure 3: Spectral density estimation in small samples: Wavelets (Level 5) vs. Fourier. (A) T = 512 (2^9), level = 5(D4). (B) T = 2048 (2^{11}), level = 5(D4).

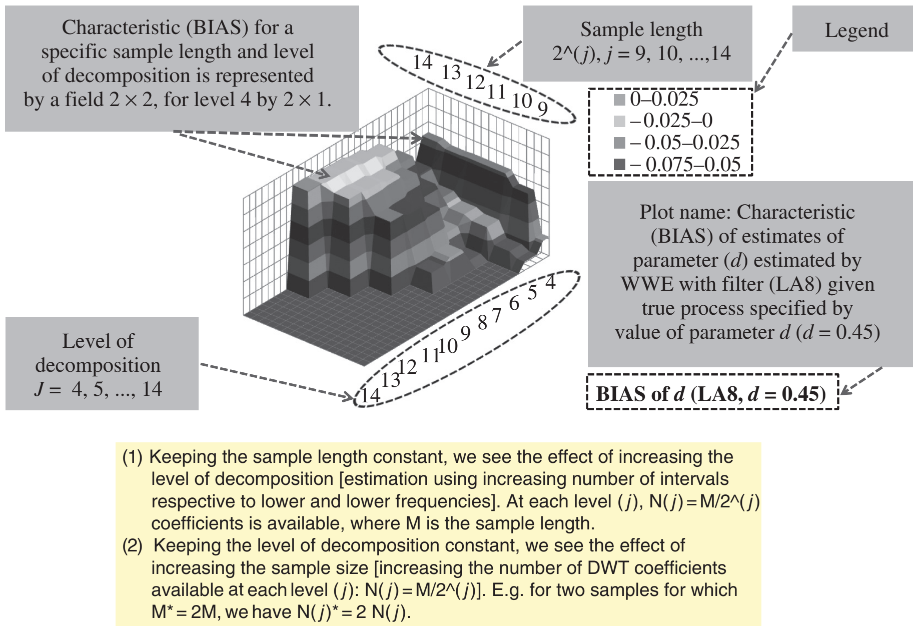

Figure 4: 3D plots guide.

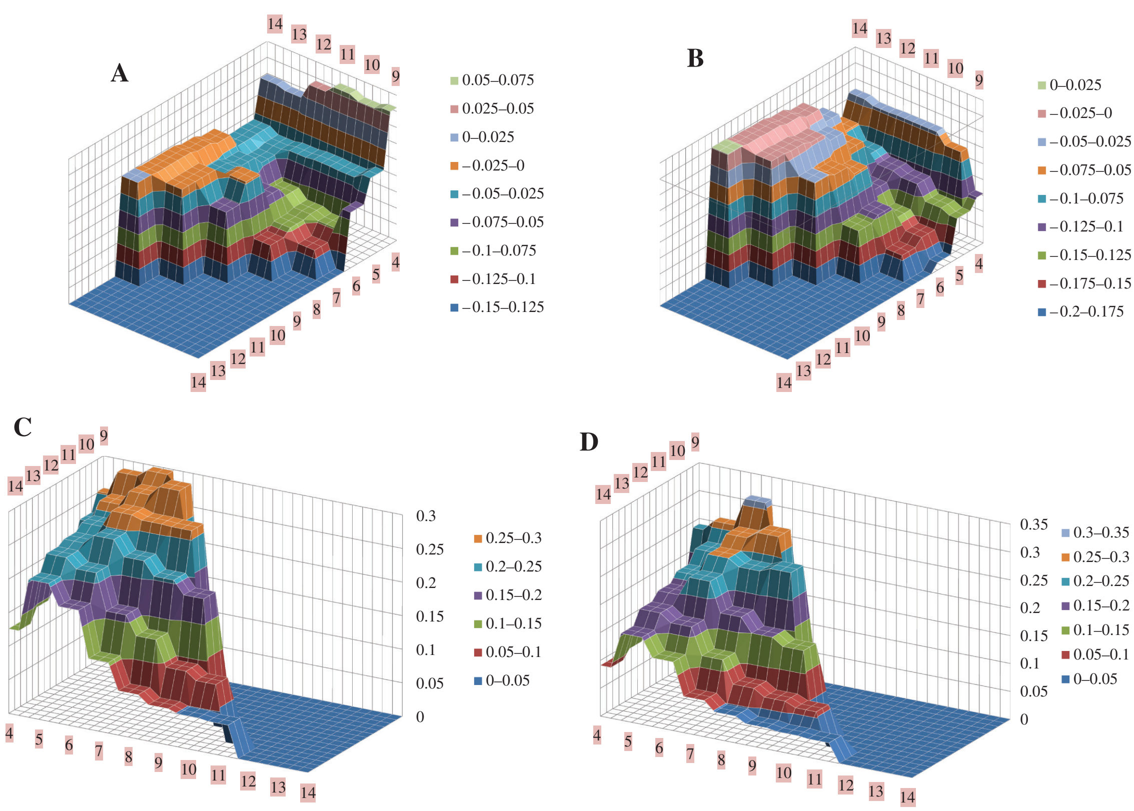

Figure 5: 3D plots: Partial decomposition: $\hat{d}$: Bias and RMSE. (A) Bias of $\hat{d}$ (LA8, d = 0.25). (B) Bias of $\hat{d}$ (LA8, d = 0.45. (C) RMSE of $\hat{d}$ (LA8, d = 0.25. (D) RMSE of $\hat{d}$ (LA8, d = 0.45).

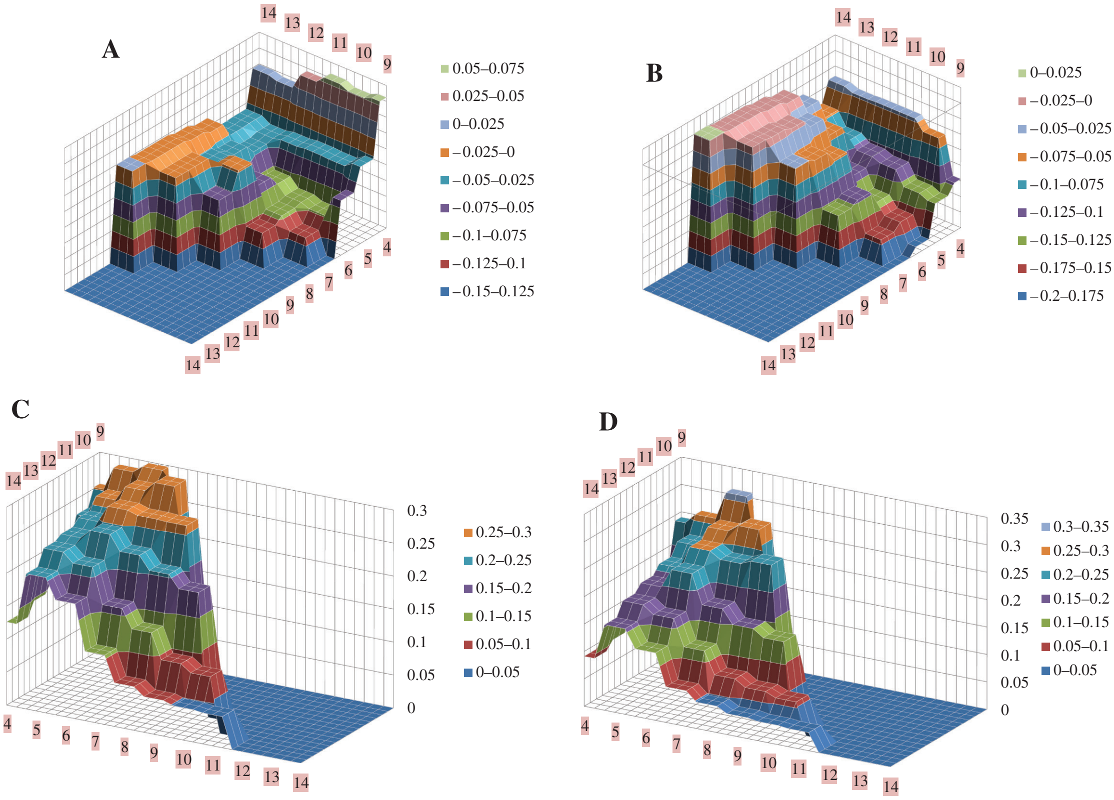

Figure 6: 3D plots: Partial decomposition: $\hat{\alpha}$: Bias and RMSE. (A) Bias of $\hat{\alpha}$ (LA8, $d=0.25$. (B) Bias of $\hat{\alpha}$ (LA8, $d=0.45$). (C) RMSE of $\hat{\alpha}$ (LA8, $d=0.25$). (D) RMSE of $\hat{\alpha}$ (LA8, $d=0.45$).

References

Barunik, J., and L. Vacha. 2015. “Realized Wavelet-Based Estimation of Integrated Variance and Jumps in the Presence of Noise.” *Quantitative Finance* 15: 1347–1364.

Barunik, J., and L. Vacha. 2016. “Do Co-Jumps Impact Correlations in Currency Markets?” Papers 1602.05489, arXiv.org.

Barunik, J., T. Krehlik, and L. Vacha. 2016. “Modeling and Forecasting Exchange Rate Volatility in Time-Frequency Domain.” *European Journal of Operational Reseach* 251: 329–340.

Beran, J. 1994. *Statistics for Long-Memory Processes*. Monographs on statistics and applied probability, 61, Chapman & Hall.

Bollerslev, T. 1986. “Generalized Autoregressive Conditional Heteroskedasticity.” *Journal of Econometrics* 31: 307–327.

Bollerslev, T. 2008. “Glossary to Arch (garch).” CREATES Research Papers 2008-49, School of Economics and Management, University of Aarhus, URL http://ideas.repec.org/p/aah/create/2008-49.html.

Bollerslev, T., and J. M. Wooldridge. 1992. “Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances.” *Econometric Reviews* 11: 143–172.

Bollerslev, T., and H. O. Mikkelsen. 1996. “Modeling and Pricing Long Memory in Stock Market Volatility.” *Journal of Econometrics* 73: 151–184.

Cheung, Y.-W., and F. X. Diebold. 1994. “On Maximum Likelihood Estimation of the Differencing Parameter of Fractionally-Integrated Noise with Unknown Mean.” *Journal of Econometrics* 62: 301–316.

Dahlhaus, R. 1989. “Efficient Parameter Estimation for Self-Similar Processes.” *The Annals of Statistics* 17: 1749–1766.

Dahlhaus, R. 2006. “Correction: Efficient Parameter Estimation for Self-Similar Processes.” *The Annals of Statistics* 34: 1045–1047.

Donoho, D. L., and I. M. Johnstone. 1994. “Ideal Spatial Adaptation by Wavelet Shrinkage.” *Biometrika* 81: 425–455.

Engle, R. F. 1982. “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation.” *Econometrica* 50: 987–1007.

Fan, J., and Y. Wang. 2007. “Multi-Scale Jump and Volatility Analysis for High-Frequency Financial Data.” *Journal of the American Statistical Association* 102: 1349–1362.

Fan, Y., and R. Gençay. 2010. “Unit Root Tests with Wavelets.” *Econometric Theory* 26: 1305–1331.

Faÿ, G., E. Moulines, F. Roueff, and M. S. Taqqu. 2009. “Estimators of Long-Memory: Fourier versus Wavelets.” *Journal of Econometrics* 151: 159–177.
Fox, R., and M. S. Taqqu. 1986. “Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series.” *The Annals of Statistics* 14: 517–532.
Frederiksen, P. H., and M. O. Nielsen. 2005. “Finite Sample Comparison of Parametric, Semiparametric, and Wavelet Estimators of Fractional Integration.” *Econometric Reviews* 24: 405–443.
Gençay, R., and N. Gradojevic. 2011. “Errors-in-Variables Estimation with Wavelets.” *Journal of Statistical Computation and Simulation* 81: 1545–1564.
Gencay, R., and D. Signori. 2015. “Multi-Scale Tests for Serial Correlation.” *Journal of Econometrics* 184: 62–80.
Gonzaga, A., and M. Hauser. 2011. “A Wavelet Whittle Estimator of Generalized Long-Memory Stochastic Volatility.” *Statistical Methods & Applications* 20: 23–48.
Heni, B., and B. Mohamed. 2011. “A Wavelet-Based Approach for Modelling Exchange Rates.” *Statistical Methods & Applications* 20: 201–220.
Jensen, M. J. 1999. “An Approximate Wavelet Mle of Short- and Long-Memory Parameters.” *Studies in Nonlinear Dynamics Econometrics* 3: 5.
Jensen, M. J. 2000. “An Alternative Maximum Likelihood Estimator of Long-Memory Processes using Compactly Supported Wavelets.” *Journal of Economic Dynamics and Control* 24: 361–387.
Johnstone, I. M., and B. W. Silverman. 1997. “Wavelet Threshold Estimators for Data with Correlated Noise.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 59: 319–351.
Mancini, C., and F. Calvori. 2012. “Jumps”, in *Handbook of Volatility Models and Their Applications*, edited by L. Bauwens, C. Hafner and S. Laurent, John Wiley & Sons, Inc., Hoboken, NJ, USA. doi: 10.1002/9781118272039.ch17.
Moulines, E., F. Roueff, and M. S. Taqqu. 2008. “A Wavelet Whittle Estimator of the Memory Parameter of a Nonstationary Gaussian Time Series.” *The Annals of Statistics* 36: 1925–1956.
Nielsen, M. O., and P. H. Frederiksen. 2005. “Finite Sample Comparison of Parametric, Semiparametric, and Wavelet Estimators of Fractional Integration.” *Econometric Reviews* 24: 405–443.
Percival, D. P. 1995. “On Estimation of the Wavelet Variance.” *Biometrika* 82: 619–631.
Percival, D. B., and A. T. Walden. 2000. *Wavelet Methods for Time Series Analysis (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press, URL http://www.worldcat.org/isbn/0521685087.
Perez, A., and P. Zaffaroni. 2008. “Finite-Sample Properties of Maximum Likelihood and Whittle Estimators in Egarch and Fiegarch Models.” *Quantitative and Qualitative Analysis in Social Sciences* 2: 78–97.
Tseng, M. C., and R. Gençay. 2014. “Estimation of Linear Model with One Time-Varying Parameter via Wavelets.” (unpublished manuscript).
Whitcher, B. 2004. “Wavelet-Based Estimation for Seasonal Long-Memory Processes.” *Technometrics* 46: 225–238.
Whittle, P. 1962. “Gaussian Estimation in Stationary Time Series.” *Bulletin of the International Statistical Institute* 39: 105–129.
Wornell, G. W., and A. Oppenheim. 1992. “Estimation of Fractal Signals from Noisy Measurements using Wavelets.” *Signal Processing, IEEE Transactions on* 40: 611–623.
Xue, Y., R. Gençay, and S. Fagan. 2014. “Jump Detection with Wavelets for High-Frequency Financial Time Series.” *Quantitative Finance* 14: 1427–1444.
Zaffaroni, P. 2009. “Whittle Estimation of Egarch and Other Exponential Volatility Models.” *Journal of Econometrics* 151: 190–200.

Supplemental Material: The online version of this article (DOI: 10.1515/snde-2016-0101) offers supplementary material, available to authorized users.