

Kraicova, Lucie; Barunik, Jozef

**Working Paper**

## Estimation of long memory in volatility using wavelets

FinMaP-Working Paper, No. 33

**Provided in Cooperation with:**

Collaborative EU Project FinMaP - Financial Distortions and Macroeconomic Performance, Kiel University et al.

*Suggested Citation:* Kraicova, Lucie; Barunik, Jozef (2015) : Estimation of long memory in volatility using wavelets, FinMaP-Working Paper, No. 33, Kiel University, FinMaP - Financial Distortions and Macroeconomic Performance, Kiel

This Version is available at:

<https://hdl.handle.net/10419/108901>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# FinMaP-Working Paper No.33



This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 612955



FINMAP –

**FINANCIAL DISTORTIONS AND MACROECONOMIC  
PERFORMANCE: EXPECTATIONS, CONSTRAINTS AND  
INTERACTION OF AGENTS**

**DATE: 03/23/2015**

**TITLE**

## Estimation of Long Memory in Volatility Using Wavelets

by: Lucie Kraicova and Jozef Barunik

## 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 behaviour 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.

*Keywords: volatility, long memory, FIEGARCH, wavelets, Whittle, Monte Carlo.*

## AUTHORS

### 1. Lucie Kraicova

Institute of Information Theory and Automation,  
Academy of Sciences of the Czech Republic  
Pod Vodarenskou Vezi 4  
CZ-18208 Prague

Institute of Economic Studies  
Charles University  
Opletalova 21  
CZ-110 00 Prague

Email: kraicova.L@seznam. cz

### 2. Jozef Barunik

Institute of Information Theory and Automation,  
Academy of Sciences of the Czech Republic  
Pod Vodarenskou Vezi 4  
CZ-18208 Prague

Institute of Economic Studies  
Charles University  
Opletalova 21  
CZ-110 00 Prague

Email: Barunik@utia.cas.cz

# Estimation of Long Memory in Volatility Using Wavelets\*

Lucie Kraicova<sup>†</sup>      Jozef Baruník<sup>‡</sup>

March 9, 2015

## 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.

**Keywords:** volatility, long memory, FIEGARCH, wavelets, Whittle, Monte Carlo

---

\*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). Please note that the Online Appendix to this manuscript is available for download at <http://ies.fsv.cuni.cz/en/staff/kraicova>

<sup>†</sup>Institute of Economic Studies, in Prague, Opletalova 21, 110 00, Prague, CR and Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodarenskou Vezi 4, 182 00, Prague, Czech Republic.

<sup>‡</sup>Corresponding author: Institute of Economic Studies, in Prague, Opletalova 21, 110 00, Prague, CR and Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodarenskou Vezi 4, 182 00, Prague, Czech Republic. Email: [barunik@utia.cas.cz](mailto:barunik@utia.cas.cz).

# 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 led by the desire to 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 & Mikkelsen, 1996) to the original (G)ARCH modeling framework of 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 estimation of parameters 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. The main advantage of applying wavelet-based Whittle estimator in volatility modeling is that wavelets are time localized and can better approximate spectral density in case of non-stationarities found in volatility process.

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 & Nielsen (2005) and related works), literature assessing performance of their parametric counterparts is not extensive. Though, results of the studies completed so far suggest that the performance of WWE in parametric setting is an interesting and important research topic. 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 & 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 estima-

tion strategy. Heni & Mohamed (2011) apply this strategy on a FIGARCH-GARMA model, further application can be seen in Gonzaga & Hauser (2011).

Literature focusing on WWE studies various models, but estimation of FIEGARCH has not been fully explored yet with exception of Perez & 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 & 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). In an experiment setup mirroring that of Perez & 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. A 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, while in terms of efficiency the latter is usually better. 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.

The rest of the text is structured as follows: section 2 introduces the FIEGARCH model and the individual estimators; in section 3 the setup of the Monte Carlo experiment is described and results are discussed. Due to the extend of the results, we relegate all supporting results to the online appendix available for download at <http://ies.fsv.cuni.cz/en/staff/kraicova>. In section 4 we present the extended experiment; in section 5 we compare our results with related literature; section 6 concludes.

## 2 Estimation Frameworks for FIEGARCH( $q, d, p$ )

### 2.1 FIEGARCH( $q, d, p$ ) Process

Despite the extremely wide spectrum of processes generating financial returns time series, there are some stylized features which many of them have in common. They have been detected over years of financial market analysis and have shaped the means of financial time series modeling. One of the main features, the time-variant dependence in volatility led to development of conditional volatility models by Engle (1982). Over time, performance of these models (ARCH family models) in practical applications has demonstrated the importance of conditional volatility in time series analysis and feasibility of direct volatility estimation and forecasting. Although several alternative concepts based on explicitly modeled volatility have been developed since Engle (1982), generalized ARCH models are still among those best performing in practical applications. This makes it relevant to study the performance of new parameter estimators in their context. In our study we focus on one of the generalizations of the ARCH model, the FIEGARCH( $q, d, p$ ), where the log-returns  $\{\epsilon_t\}_{t=1}^T$  are modeled conditionally on their past realizations as:

$$\epsilon_t = z_t h_t^{1/2} \quad (1)$$

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

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

where  $z_t$  is an  $\mathcal{N}(0, 1)$  independent identically distributed (*i.i.d.*) unobservable innovations process,  $\epsilon_t$  is observable discrete-time real valued process with conditional log-variance process dependent on the past innovations  $E_{t-1}(\epsilon_t^2) = h_t$ , 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  $\Gamma(\cdot)$  being gamma function.

Any FIEGARCH( $q, d, p$ ) process is then fully determined by the number of parameters, their values and distribution of the standardized innovations  $z_t$ . Concerning the last factor, the three most frequent assumptions in the literature are the standard normal distribution providing a convenient estimation environment, student- $t$  distribution assuming thicker tails, and Generalized Error Distribution (GED) with parameter  $v$  determining the tail thickness. Normal distribution is nested as a special case of GED for  $v = 2$ .

The model captures important stylized features of the real financial time series data; short-term temporal variation in financial returns volatility (volatility clustering), long-term temporal variation in financial returns volatility (long memory), negative relationship between past returns and volatility (leverage effect) and fat-tailed sample distribution of returns. We provide plots of a FIEGARCH process with three different levels of long memory for illustration in Section ?? of an online appendix.

While correct model specification is important for capturing all the empirical features

of the data, feasibility of estimation of its parameters is crucial. In general, estimation of the FIEGARCH model can be carried out by various methods. Below, those considered in this work (the benchmark estimators MLE and FWE, and the newly introduced WWE) are described together with practical aspects of their application.

## 2.2 (Quasi) Maximum Likelihood Estimator

The Maximum Likelihood Estimation is often considered the gold-standard for parameter estimation, hence it serves as a natural benchmark estimation framework used in this study. For a general zero mean, stationary Gaussian process  $\{x_t\}_{t=1}^T$ , the maximum likelihood estimator (MLE) is defined as

$$\hat{\zeta}_{MLE} = \underset{\zeta \in \Theta}{\operatorname{argmin}} \mathcal{L}_{MLE}(\zeta), \quad (4)$$

where  $\mathcal{L}_{MLE}(\zeta)$  is the negative log-likelihood function

$$\mathcal{L}_{MLE}(\zeta) = \frac{T}{2} \ln(2\pi) + \frac{1}{2} \ln |\Sigma_T| + \frac{1}{2} (x'_t \Sigma_T^{-1} x_t), \quad (5)$$

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

Despite the favorable properties of the MLE, there are some issues limiting its practical applicability. The usual problem is that we have to deal with the inversion of the covariance matrix of the process and with its determinant. Although it may not be a problem when the matrix is diagonal or sufficiently sparse, in cases of dense covariance matrices (characteristic for long memory processes) it may be extremely time demanding, or even unfeasible in case of large datasets. Moreover, as discussed in (Beran (1994), chapter 5), solution may be even unstable in the presence of long memory, when the covariance matrix is close to singularity. Next, empirical data often does not have zero mean, hence the mean has to be estimated and deducted. The efficiency and bias of the estimator of the mean contributes to the efficiency and bias of the MLE. In case of long-memory processes it can cause significant deterioration of the MLE (Cheung & 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 & Frederiksen, 2005).

Since the assumption of a specific distribution is usually too restrictive for practical applications, it is important to study the estimator in situations when it is constructed for some process but applied to a different process. In the context of GARCH processes with non-normal error distribution, Quasi-Maximum Likelihood Estimator (QMLE) has been studied by Bollerslev & Wooldridge (1992), who show that the estimator remains consistent, but loses efficiency. The efficiency loss, as argued in Engle & Gonzalez-Rivera (1991), is rather small for symmetric  $t$ -distributed processes, but can be significant under asymmetric distribution. As discussed in Bollerslev & Wooldridge (1992), standard test statistics become biased and to ensure valid inference, their robustified counterparts, such as those proposed by the authors,



should be used. Unlike the case of GARCH model, the asymptotic theory for EGARCH and FIEGARCH QMLE is not yet available. Though, the (Q)MLE is the basic tool for their estimation suggested by their authors in Nelson (1991) and Bollerslev & Mikkelsen (1996) respectively and the properties of the estimator are assessed via empirical studies (Baillie et al., 1996; Bollerslev & Mikkelsen, 1996; Perez & Zaffaroni, 2008).

To derive (Q)MLE of the FIEGARCH process, we just rewrite the general formula for the negative log-likelihood function (Eq. 5). Since we assume log-returns  $\epsilon_t$  to follow a Gaussian, zero-mean process of independent variables,  $\Sigma_T$  will be diagonal with conditional variances  $h_t$  as its elements, and determinant will reduce to a sum of its diagonal terms. The reduced-form negative log-likelihood function can be written as:

$$\mathcal{L}_{(Q)MLE}(\zeta) = \sum_{t=1}^T \left( \ln h_t(\zeta) + \frac{\epsilon_t^2}{h_t(\zeta)} \right), \quad (6)$$

where  $\zeta$  is a vector of parameters in the equation for conditional log-variance in (Eq. 1). Then the (Q)MLE estimator is defined as  $\hat{\zeta}_{(Q)MLE} = \operatorname{argmin}_{\zeta \in \Theta} \mathcal{L}_{(Q)MLE}(\zeta)$ , where  $\Theta$  is the parameter space.

Despite the simple formula, some practical issues arise because of the need to model the conditional volatility process during the optimization. As discussed in Bollerslev & Mikkelsen (1996), since in the context of GARCH models the forecasted volatility depends on the past forecasts, the log-likelihood function becomes dependent on initial conditions, the basis for conditional variance recursion. Since, under the general specification of the model,  $h_t$  depends on  $\{h_{t-i}, i = 1, 2, \dots, \infty\}$ , a truncation is needed. In case of a short memory process, the dependence on the past declines exponentially, which enables to use a relatively small number of initializing values. In contrast, when long memory is present, large number of pre-sample “observations” is necessary to prevent a significant loss of information about long-run dependencies in the process. Following the approach of Bollerslev & Mikkelsen (1996), in our Monte Carlo experiment, the pre-sample conditional volatilities are set to the sample volatility, conditional mean is assumed to be known and equal to zero (later on also the non-zero case is assumed) and for both the simulation and estimation the truncation is done at lag 1000. Using these initializing values, for each set of parameters considered by the optimization algorithm the whole vector of conditional volatilities is estimated, together with the implied squared returns needed for further iterations. Given the complexity of this procedure, the method is significantly time consuming.

### 2.3 Fourier-based Whittle Estimator

The Fourier-based Whittle estimator (FWE) is a spectral-based counterpart of the MLE, where the problematic terms in the log-likelihood function (with the possibly dense covariance matrix) the  $|\Sigma_T|$  and  $x_t' \Sigma_T^{-1} x_t$ , are replaced by their asymptotic frequency domain representations. The link between time domain and frequency domain is defined by means of Fourier transform. The approximation is based on a study of eigenvectors and respective

eigenvalues of the covariance matrix leading to a conclusion that the matrix can be diagonalized by means of Fourier transform. Orthogonality of the Fourier transform projection matrix then allows to achieve the approximation by means of multiplications by identity matrices, simple rearrangements and approximation of integrals by Riemann sums, see Beran (1994). The reduced-form approximated Whittle negative log-likelihood function for estimation of parameters under Gaussianity assumption is:

$$\mathcal{L}_W(\zeta) = \frac{1}{T} \sum_{j=1}^{m^*} \left( \ln f(\lambda_j, \zeta) + \frac{I(\lambda_j)}{f(\lambda_j, \zeta)} \right), \quad (7)$$

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  $m^* = \max \{m \in \mathbb{Z}; m \leq (T-1)/2\}$ , i.e.  $\lambda_j < 1/2$ , and its link to the variance-covariance matrix of the process  $x_t$  is:

$$\text{cov}(x_t, x_s) = \int_{-1/2}^{1/2} f(\lambda, \zeta) e^{i2\pi\lambda(s-t)} d\lambda = 2 \int_0^{1/2} f(\lambda, \zeta) e^{i2\pi\lambda(s-t)} d\lambda; \quad (8)$$

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

$$I(\lambda_j) = (2\pi T)^{-1} \left| \sum_{t=1}^T x_t e^{i2\pi\lambda_j t} \right|^2, \quad (9)$$

and the respective Fourier-based Whittle estimator is defined as (for a detailed FWE treatment see e.g. Beran (1994)):

$$\hat{\zeta}_W = \underset{\zeta \in \Theta}{\operatorname{argmin}} \mathcal{L}_W(\zeta). \quad (10)$$

It can be shown that the FWE has the same asymptotic distribution as the exact MLE, hence is asymptotically efficient for Gaussian processes (Fox & 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' knowledge, the only one out of the three estimators considered in this work, for which an asymptotic theory is currently available. The theory is derived in Zaffaroni (2009) for a whole class of

exponential volatility processes; both strong consistency and asymptotic normality are established, 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, as discussed and derived in Perez & Zaffaroni (2008) and Zaffaroni (2009), it is necessary to rewrite the model in a signal plus noise form:

$$x_t = \ln(\epsilon_t^2) = \ln(z_t^2) + \omega + \sum_{s=0}^{\infty} \Phi_s g(z_{t-s-1}) \quad (11)$$

$$g(z_t) = \theta z_t + \gamma[|z_t| - E(|z_t|)] \quad (12)$$

$$\Phi(L) = (1 - L)^{-d} [1 + \alpha_{[2]}(L)][\beta(L)]^{-1}. \quad (13)$$

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  $\epsilon_t$ . For the detailed derivation of the transformed process and for the formula for its spectral density see Appendix A.

## 2.4 Wavelet Whittle Estimator

Although FWE seems to be a good alternative to MLE for the FIEGARCH model estimation in the case of application on FIEGARCH underlying processes (Perez & Zaffaroni, 2008), its use on real data may be, in some cases, problematic. This is because the FWE performance depends on the accuracy of the spectral density estimation using periodogram, which may be impaired by various time-localized patterns in the data diverging from the underlying FIEGARCH process. 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), we use discrete wavelet transform (DWT). While DFT is projection of the time series on periodic functions with infinite support, DWT is a projection on a finite-support function, which may be advantageous particularly for some datasets.

### 2.4.1 Discrete Wavelet Transform

To provide an introduction to the Wavelet Whittle Estimation, we briefly describe the wavelet transform that determines its properties and makes it different from the FWE. The core of any wavelet transform is a wavelet system, whose construction, together with means of the projection applied, determine the characteristics of the transformed data. For any  $s \in \mathbb{R}$ , a basic wavelet system can be defined as a set  $\{\{\varphi_{j_0,k}(s)\}, \{\psi_{j,k}(s)\}; k \in \mathbb{Z}, j = j_0, j_0 - 1, j_0 - 2, \dots\}$  creating an orthonormal basis in  $L^2(\mathbb{R})$ ; which means that any function  $f \in L^2(\mathbb{R})$  can be

expressed as

$$f(s) = \sum_k \alpha_{j_0,k} \varphi_{j_0,k}(s) + \sum_{j=-\infty}^{j_0} \sum_s \beta_{j,k} \psi_{j,k}(s), \quad (14)$$

where  $\alpha_k = \int f(s) \overline{\varphi_{j_0,k}(s)} ds$ , and  $\beta_{j,k} = \int f(s) \overline{\psi_{j,k}(s)} ds$ , where the elements  $\alpha_k$ ,  $\beta_{j,k}$ ,  $\varphi(s)$  and  $\psi(s)$  are called scaling coefficients, detail (wavelet) coefficients, scaling function (father wavelet) and wavelet function (mother wavelet) respectively, and the translated and dilated transformations of the mother wavelet are called daughter wavelets. With increasing  $j$ , these daughter wavelets get wider, with  $j \leq 0$  they are narrower than the mother wavelet.

The basic conditions for  $\psi(s)$  to be a valid mother wavelet are that  $\int \psi(s) ds = 0$  and  $\int \psi^2(s) ds = 1$ , while the usual requirement is also the “admissibility” condition  $\int \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega < \infty$ , where  $\hat{\psi}$  is the Fourier transform of  $\psi$ . This condition ensures that we can reconstruct the original time series from its transform. For complete conditions on  $\varphi(s)$ ,  $\psi(s)$  to be valid father and mother wavelets in the context various subsets of  $L^2(\mathbb{R})$  and for other details concerning construction of wavelet systems see Hardle et al. (1998). In addition, we provide examples of wavelets in Section ?? and ?? of Online Appendix.

Next, any method that decomposes original data using the wavelet system and expresses them in terms of coefficients  $\{\alpha_k, \beta_{j,k}\}$  and functions  $\{\varphi(s), \psi(s)\}$  defined above, is a wavelet transform. In case of  $j \in \mathbb{Z}$ , as applied in our work, we speak about a discrete wavelet transform (DWT), while for  $j \in \mathbb{R}$  the transform is continuous (CWT). By tradition, the default choice of scales is  $\{2^{1-j}; j \in \mathbb{Z}\}$ , thus the standard DWT can be defined in terms of the wavelet expansion (Eq. 14) with scaling defined as  $s(j) = 2^{1-j}$  (i.e. “scale  $j$ ” refers to the scaling  $2^{1-j}$ , scale 1 refers to  $2^0 = 1$ ). The DWT coefficients are obtained for scales  $j_0 = J$ ,  $j_0 - 1 = J - 1$ ,  $j_0 - 2 = J - 2$ , ...,  $j_0 - (j_0 - 1) = 1$  using two-channel filter banks and down-sampling, so that at each level of decomposition  $j$  of a series of length  $M$  we get  $M/2^j$  DWT coefficients, see e.g. Jensen (2000). These coefficients can be in turn used for decomposition of the variance  $\sigma^2$  of the process  $x_t$ :

$$\sigma^2 = E(x_t^2) - [E(x_t)]^2 = \frac{\|x_t\|^2}{T} - [E(x_t)]^2 = \frac{\sum_{j=1}^J \|W_j\|^2 + \|V_J\|^2}{T} - [E(x_t)]^2, \quad (15)$$

where  $W_j$ ;  $j = 1, \dots, J$  and  $V_J$  are vectors of wavelet and scaling coefficients respectively and the  $[E(x_t)]^2$  can be estimated using the squared sample mean  $\bar{x}_t^2$ , or using the true squared mean whenever known. Alternatively, we can use the coefficients for estimating the spectral density  $f(\lambda, \zeta)$  of  $x_t$  using the relationship:

$$\frac{\|W_j\|^2}{T} = \frac{\sigma_{W,j}^2}{2^j} \approx 2 \int_{1/2^{j+1}}^{1/2^j} f(\lambda, \zeta) d\lambda \quad (16)$$

$$\frac{\|V_J\|^2}{T} - \bar{x}_t^2 = \frac{\sigma_{V,J}^2}{2^J} - \left(1 - \frac{1}{2^J}\right) \bar{x}_t^2 \approx 2 \int_0^{1/2^{J+1}} f(\lambda, \zeta) d\lambda, \quad (17)$$

where  $\sigma_{W,j}^2$  and  $\sigma_{V,J}^2$  are the sample variances of the wavelet and scaling coefficients respec-

tively for  $j = 1, 2, \dots, J$ .

#### 2.4.2 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 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 & Walden (2000), the reduced-form Wavelet-Whittle objective function can be defined as:

$$\mathcal{L}_{WW}(\zeta) = \ln |\Lambda_T| + (W'_{j,k} \Lambda_T^{-1} W_{j,k}) \quad (18)$$

$$= \sum_{j=1}^J \left[ N_j \ln \left( 2 \int_{1/2^{j+1}}^{1/2^j} 2^j f(\lambda, \zeta) d\lambda \right) + \sum_{k=1}^{N_j} \frac{W_{j,k}^2}{2 \int_{1/2^{j+1}}^{1/2^j} 2^j f(\lambda, \zeta) d\lambda} \right], \quad (19)$$

where  $W_{j,k}$  are the wavelet (detail) coefficients, and  $\Lambda_T$  is a diagonal matrix with elements  $\{C_1, C_1, \dots, C_1, C_2, \dots, C_J\}$ , 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

$$\hat{\zeta}_{WW} = \underset{\zeta \in \Theta}{\operatorname{argmin}} \mathcal{L}_{WW}(\zeta), \quad (20)$$

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. At this time, 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.

### 2.4.3 Full vs. Partial Decomposition

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 & Oppenheim (1992) and studied by Moulines et al. (2008), who derive the asymptotic theory for LWWE with general upper and lower bound for levels of decomposition  $\{j \in \langle L, U \rangle; 1 \leq L < U \leq J\}$ , 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 make the point, let's rewrite the WWE objective function as:

$$\mathcal{L}_{WW}(\zeta) = \sum_{j=1}^J N_j \left( \ln \sigma_{W,j,DWT}^2(\zeta) + \frac{\hat{\sigma}_{W,j,DWT}^2}{\sigma_{W,j,DWT}^2(\zeta)} \right), \quad (21)$$

where  $\sigma_{W,j,DWT}^2(\zeta)$  is the theoretical variance of  $j^{th}$  level DWT coefficients and  $\hat{\sigma}_{W,j,DWT}^2$  is its sample counterpart,  $\zeta$  is the vector of parameters in  $\sigma_{W,j,DWT}^2(\zeta)$  and  $\{W_{j,DWT}; j = 1, \dots, J\}$  are vectors of DWT coefficients used to calculate  $\hat{\sigma}_{W,j,DWT}^2$ . Using the definition of wavelet variance  $v_j^2 = 2 \int_{1/2^{j+1}}^{1/2^j} f(\lambda, \zeta) d\lambda = \frac{\sigma_{W,j,DWT}^2}{2^j}$ ;  $j = 1, 2, \dots, J$  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

$$\mathcal{L}_{WW}^*(\zeta) = \sum_{j=1}^J \frac{N_j}{N^*} \left( \ln \sigma_{W,j,DWT}^2(\zeta) + \frac{\hat{v}_{W,j,DWT}^2}{v_{W,j,DWT}^2(\zeta)} \right), \quad (22)$$

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

The quality of our estimate of  $\zeta$  depends on the the quality of our estimates of  $\sigma_{W,j,DWT}^2(\zeta)$  using sample variance of DWT coefficients, or equivalently, on the quality of our estimates of  $v_{W,j,DWT}^2(\zeta)$  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, Figure 3 and Figure 4 in Appendix C, 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 3, 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 et al. (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 & Walden (2000), the convergence of the wavelet variance estimator is relatively fast, so that 128 ( $2^7$ ) coefficients should already ensure a reasonable accuracy<sup>1</sup>. 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 6 lags ensuring 128 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 & 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

---

<sup>1</sup>Accuracy of the wavelet variance estimate, not the parameters in approximate MLE.

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 & Walden (2000)

$$\hat{\sigma}_{W,j,DWT}^2 = \frac{\sum_{k=1}^{N_j} W_{j,k,DWT}^2}{N_j} = 2^{j+1} \int_{1/2^{j+1}}^{1/2^j} f(\lambda, \zeta) d\lambda \quad (23)$$

$$\hat{\sigma}_{W,j,MODWT}^2 = \frac{\sum_{k=1}^T W_{j,k,MODWT}^2}{T} = 2 \int_{1/2^{j+1}}^{1/2^j} f(\lambda, \zeta) d\lambda, \quad (24)$$

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

$$\hat{\sigma}_{W,j,DWT}^2 = 2^j \hat{\sigma}_{W,j,MODWT}^2. \quad (25)$$

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

$$\mathcal{L}_{WW,MODWT}^* = \sum_{j=1}^J \frac{N_j}{N^*} \left( \ln \sigma_{W,j}^2(\zeta) + \frac{2^j \hat{\sigma}_{W,j,MODWT}^2}{\sigma_{W,j}^2(\zeta)} \right), \quad (26)$$

and the MODWT-based WWE estimator as:

$$\hat{\zeta}_{WW,MODWT} = \underset{\zeta \in \Theta}{\operatorname{argmin}} \mathcal{L}_{WW,MODWT}^*. \quad (27)$$

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. Results for this alternative estimator are presented later in the text.

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 Table 1 and Figure 3 in Appendix C, 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 et al. (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, this would be beyond the scope of our current research. Therefore, the analysis we present reduces to an extension of the set of Monte Carlo experiments to cover both the full and the partial decomposition, to demonstrate the relevancy of the problems mentioned above and to provide a motivation for further research in this area.

#### 2.4.4 FIEGARCH WWE

After defining the general form of the estimator and discussing its properties, let's focus on the FIEGARCH 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. Percival & Walden (2000) in chapter 4 discuss some general practical considerations, including the wavelet choice, handling of boundary coefficients, choice of the decomposition level and application of the DWT on series with different length than  $2^j$ . The filters chosen for the Monte Carlo experiment in our work are the same as those chosen in Percival & 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 & Walden (2000), Jensen (1999), Jensen (2000) or Johnstone & Silverman (1997)). In Section ?? of Online Appendix, the quality of the DWT-based decorrelation is assessed based on the dependencies among the resulting wavelet coefficients. We provide 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. Based on the results, the approximation of the spectral density can be applied. 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.

### 2.4.5 Preliminary Results: FWE vs. WWE

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. For this purpose, a set of figures in Appendix B is provided, showing the rationale for WWE application. 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 2, 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 Section ?? of Online Appendix.

## 3 Monte Carlo Experiments

In order to study how the WWE performs compared to the two benchmark estimators (MLE and FWE), an extensive Monte Carlo experiment has been carried out. 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. The experiment setup mirrors that of Perez & Zaffaroni (2008), which ensures consistency with that work and enables interpretation of the new results as an extension to those already published. Since in this benchmark study no wavelet-based methods are used, choices concerning the WWE application and extension of simulations to longer data sets have been made with respect to other relevant literature (Jensen (1999), Percival & Walden (2000)), as discussed earlier. Technical details of the experiment and tables with results are provided in Section ?? of Online Appendix.

### 3.1 Part I: Maximal Decomposition

At first, the experiment has been performed using MLE, FWE and DWT-based WWE with maximal level of decomposition. The maximal level of decomposition means that for sample

of length  $2^j$ , using Haar, D4 and LA8 wavelets, we have levels  $j$ ,  $j - 2$  and  $j - 3$  respectively. This is due to the truncation of boundary coefficients explained in the previous section.

In general, the WWE works fairly well in all setups. Especially biases are low, and in most cases decline with increasing sample size. The exception is parameter  $\alpha$  for all filters and parameter  $\theta$  for Haar filter, where the convergence is problematic. Yet, even in these situations the bias remains low for all filters and samples up to  $2^{11}$ , and RMSE (although relatively high) is declining with sample size as for all the other parameter estimates across all setups. Focusing on the differences between individual filters, the strength of long memory, sample size and parameter concerned seem to be important. In all setups, Haar performs the best in estimating the long-memory parameter  $d$ . Other parameters ( $\alpha$ ,  $\beta$ ,  $\gamma$  and in case of small samples also the  $\theta$ ), under  $d = 0.25$ , are better estimated using filters with larger support. In case of  $d = 0.45$  and  $d = -0.25$  the relative performance slightly improves for filters with smaller support. 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  $2^{14}$ ).

Concerning the relative performance of the WWE, FWE and MLE, the WWE works in general comparably to the FWE. In many cases it outperforms the FWE in terms of bias, while in terms of RMSE the FWE is better. Yet, the absolute differences are usually small. As expected, estimates using MLE are in most cases the best. This remains true even in cases with strong long memory, since the long memory is in the variances, not in the (log-return) process itself and the problem with mean estimation under long memory does not apply. The Whittle estimators outperform the exact MLE in some cases, but usually it is in situations with negative memory in the data, which is, based on the current literature on financial returns analysis, not of a great interest for most practical applications.

### 3.2 Part II: Partial Decomposition

The additional Monte Carlo experiments have been designed to mirror the setup used in the case of full decomposition, with the only difference in the number of levels used for the estimation. For all sample lengths ( $2^M$ ,  $M = 9, 10, \dots, 14$ ) experiments for levels  $J$ ,  $J = 4, 5, \dots, M$  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. For results including mean estimates, respective levels of bias and RMSE see tables in Section ?? of Online Appendix.

As the results suggest, for small samples, estimation under the restriction to first four levels of decomposition leads to better estimates of  $d$  and worse estimates of  $\alpha$  in terms of both bias and RMSE, while for longer samples the opposite holds. Other coefficients are estimated sometimes with lower bias, sometimes with lower RMSE than in the case

of full decomposition depending on the sample size, strength of the long memory and also on the filter applied. With increasing sample size the performance of the estimator under partial decomposition deteriorates relatively to that using full decomposition. 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 small samples with  $d = 0.25$ , where Haar seems to be better. In general, it can be said that this partial decomposition setup offers just a different set of bias and RMSE for all parameters than the full decomposition. The choice would depend on the weights assigned to the bias and RMSE and the importance we attach to individual parameters, which could be based on a bias and RMSE of one day forecasts added to the Monte Carlo experiment. While there is a relatively high probability that based on the more detailed analysis the level 4 setup may be preferred in case of short samples, for long samples the full decomposition is likely to be more appropriate.

Moving on to the truncation at level 5, significant overall improvement in the short-sample estimates is apparent for both  $d = 0.25$  and  $d = 0.45$ . Not only are they better compared to the level 4 setup, but also compared to the full decomposition. Relative performance with respect to FWE and MLE also changes, WWE works in most cases better than the FWE for all filter specifications. Focusing on the relative performance of the filters considered, the results suggest to use D4 for  $2^{10} - 2^{13}$  and switching to LA8 for  $2^9$  and  $2^j$ ;  $j > 13$  in case of  $d = 0.25$ ; under  $d = 0.45$  LA8 performs the best for all sample sizes as in the case of preceding partial decomposition setup.

Next, under truncation at level 6 the estimator seems to work comparably to the case of truncation at level 5. In most cases it offers an alternative of somewhat lower RMSE at the cost of slightly higher bias, for some parameters even the bias improves. Though, due to the significantly worse estimates of long memory, even to some extent counterbalanced by better estimates of other parameters, the truncation at level 5 may be preferred. The relative performance could be assessed based on the bias and RMSE of one day forecasts added to the Monte Carlo experiment as already proposed in the case of level 4 truncation. Regarding the relative performance of the filters considered, in case of  $d = 0.25$  D4 performs the best for almost all sample sizes, while it is outperformed by LA8 when the parameter  $d$  becomes larger. Compared to the full decomposition, in case of small samples the estimator works better in most cases in terms of both bias and RMSE. In case of longer samples, the estimates of the long memory parameter deteriorate relatively to their full decomposition counterparts, while short-term dynamics parameters are still estimated in most cases with lower bias and in case of  $d = 0.45$  also with lower RMSE under the truncation.

We conclude that the results well demonstrate the effects mentioned when discussing the partial decomposition in 2.4.3. 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. Yet, the convergence with sample size and with the level of decomposition is not easy to interpret. To see the interesting convergence patterns determined by the synergy of various effects of the truncation on the estimates see 3D plots in Appendix D providing a graphical decomposition of the convergence into the convergence with sample size and convergence with increasing level of decomposition; graphs for the estimates of  $d$  and  $\alpha$  under  $d = 0.25$ ;  $D4$ ,  $LA8$  and  $d = 0.45$ ;  $LA8$  are available. To make the figures comprehensive and well interpretable, additional Monte Carlo experiments have been performed enabling to present the whole spectrum of possible truncations from that leading to estimation at level 4 to full decomposition. As can be seen, 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. To understand the specific patterns, next to the derivation of the asymptotic theory, it would be interesting to compare the results with their MODWT-based counterparts, which would enable to separate the effect of deteriorating DWT variance estimates and would potentially lead to better interpretable convergence patterns.

## 4 Monte Carlo Extension: Jumps and Forecasting

As has been concluded in the previous section, on simulated pure FIEGARCH processes the best estimator in terms of both bias and RMSE (in case of individual parameters estimation) seems to be the MLE, followed by FWE and somewhat less “accurate” WWE. But, as discussed in the sequel, deprecating WWE based on these results only might be premature. Next, we assume a more realistic scenario, where the simulated process is augmented by specific time-localized irregularities - jumps - in the log-return process. Since the evaluation based on individual parameters estimation only may not be the best practice when forecasting is the main concern, let’s analyze also the relative forecasting performance. As a motivation for this step additional plots have been prepared, which can be found in Appendix E (online) and some of them also in ???. They show the bias of the mean estimated spectral densities using FWE and DWT-based WWE under various setups. Except for small samples, where the performance of the FWE is significantly worse than that of WWEs in terms of bias, both the estimators perform very well and in most cases differences are almost negligible. This suggests that the forecasting RMSE should play the major role. Then, based on the results for individual coefficients, FWE can be expected to dominate the DWT-based WWE, at least in case of larger samples and, of course, data generated by a pure FIEGARCH process. But this is just an ex ante guess, the need for a Monte Carlo experiment extension is apparent. Practical issues of this kind of evaluation are discussed later in this section.

## 4.1 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 summary see e.g. Mancini & Calvori (2012). So even if the FIEGARCH process could well approximate the true underlying volatility, it is important to study the jump process and use the additional information for forecasts improvement.

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:

$$\epsilon_t = z_t h_t^{1/2} + J_t(\lambda), \quad (28)$$

where the process  $h_t$  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 one 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 & 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.

## 4.2 Forecasting

Next extension, as mentioned above, is the evaluation of the in-sample and out-of-sample forecasting performance. For each simulation the fitted values and a one day ahead forecast per each estimator are calculated. The out-of-sample forecasts are directly stored for further analysis, the in-sample forecasts are transformed to mean error, mean absolute deviation and mean squared error statistics. These statistics are stored and used for overall statistics calculation. When we get the data from all 1000 simulations, we compute the mean error, mean absolute deviation and root mean squared error for both the in-sample and out-of-sample forecasts.

Although the idea of forecasting evaluation seems to be simple, there are some issues we had to deal with. The most important one is the dependency of the forecasting results on the fitting algorithm. This algorithm is technically the same as the one used in MLE and it is in fact possible to manage the maximal error of the forecasts in case of divergence of

the fitted time series as well as to ensure robustness to extreme log-returns observations in the data input. We have chosen a basic algorithm that only ensures that the operation does not break and that in case of non-positive and/or infinite fitted conditional variance the algorithm returns sample variance instead and continues computing. But then we have situations, especially in case of jumps, where a few forecasts can have a huge finite error, but other are quite accurate. Then, average error measures are not the best practice, since then the estimator with slightly lower maximal error would be considered better even though in most cases it could be much worse than the alternatives. In case of out-of-sample forecasts (one forecast per simulation) we solve this problem by using error and absolute deviation quantiles.

### 4.3 Practical Aspects

Although we expected that the WWE could be easily adjusted to be robust to jumps (compared to FWE and MLE) and thus become a good alternative in case of FIEGARCH estimation on the real data, there are two technical details which make this hardly possible and the WWE theoretically rather than empirically evincible. First, the transformation needed for FWE and WWE derivation hides the jumps in the process (they add volatility, but are not detectable in the transformed data), as can be seen in Appendix F; second, for forecasting we need jump-free data as the input - else we get inaccurate estimates even in case of perfect coefficient estimates. Thus, the jump detection and data adjustment has to be done before the actual parameters estimation takes place. 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 & Johnstone (1994) and that is described in detail and successfully applied in Barunik & Vacha (2014). 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 thing to do. So we use this simple transform for our data with moderate jumps, but in case of data with extreme jumps, such as those in Appendix F, we propose to use a different method that would get rid of all the jumps, otherwise the estimation results would be poor. Thus, it is important to distinguish between the jump detection and model estimation as two separable tasks. This holds even in cases of large jumps which are detectable in the transformed data, since in real applications we do not know what kind of jump process are we dealing with and also even in this case it is easier to found the jumps in the data before the transformation. Then we can only study how are the individual estimators able to deal with the residual jumps, which are not detected and subtracted from the time series. And of course, the better the jump estimation method, the lesser the residuals impact.

#### 4.4 Results III: FIEGARCH-Jump

The main results of the Monte Carlo experiment are summarized in tables in Appendix G. In the first two tables MLE, FWE and MODWT-based WWE are compared in terms of individual parameters estimation performance, results for DWT-based WWE are not included due to the limited space. Concerning the comparison of these two estimators, 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 strictly preferred. This is only supported by the forecasting results presented in the next tables. 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, it remains the best followed by the two Whittles, which perform comparably, FWE in most cases works slightly better. Yet, the bias of the MLE is significant and we would prefer the use of FWE considering both the bias and the RMSE and 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 FWE 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. Unfortunately, as discussed earlier, WWE does not provide any significant estimation improvement or time savings. Moreover, the use of WWE causes about twice as many cases with extremely poor in-sample fit than the FWE. By its nature, MLE does not cause these poor fit situations at all. But of course, in practice, adjustment of the optimization algorithms as well as the forecasting algorithm could prevent these cases for all estimators. We did not apply any special adjustments just to keep the estimators comparison as “fair” as possible and we propose the question of the algorithms optimization as a topic for future research.

### 5 Comparison with Literature

As already emphasized, the Monte Carlo setup has been chosen to mirror that of Perez & Zaffaroni (2008) to keep consistency in the research and enable direct comparison of the results. Since the benchmark paper focuses on the relative performance of FWE and MLE under the same conditions as applied in our paper, it is interesting to check whether the results for these two estimators are in both works the same. In the case of MLE, the answer is yes,



up to small differences caused by the uniqueness of every simulated time series. For FWE, the differences are somewhat larger (although not extreme; the maximal deviation for both bias and RMSE is less than 0.1, mostly amounting to about 0.06 or less). Based on an analysis of the code underlying the benchmark paper, this may be caused by a different optimization setup, that utilizes an explicitly formulated analytical gradient of the objective function. This helps to estimate the individual coefficients more accurately. Then, it seems natural to run the Monte Carlo experiments for our research using this, virtually more efficient, algorithm (or search for even better one). Yet, this is a non-trivial task. This follows from the form of the WWE objective function and the respective derivations. To enable the re-estimation, more efficiently written code or an adjusted means of gradient calculation would be needed, which is beyond the scope of our current work. In general, the comparison above shows the importance of taking the estimation as a complex problem including many, both theoretical and practical, issues. Most importantly, when comparing several estimators, it is usually not feasible to separate the performance of the estimator from the performance of the optimization algorithm applied. Even using the same algorithm, as applied in this work, cannot generally solve this problem, since each of the estimators may be affected differently. Then, in case of empirical analysis, it is reasonable to speak about comparison of methods of estimation instead of comparison of the individual estimators. The focus is then on the whole sets of estimators and respective means of optimization. This highlights the importance to analyze all details of the methods and optimize the estimation setup as a whole, before any definite conclusions can be made.

Next, comparing the results in this paper with some other works on wavelet-based maximum likelihood estimation, no strange patterns that would contradict the earlier conclusions are found. A relatively good performance of the WWE comparable to that of FWE is observed, which is in compliance with studies using simulated smooth processes. The absolute performance of the WWE is somewhat worse than in the benchmark papers, which is expectable given the complexity of the FIEGARCH model implying more difficult parameters identification compared to the other models estimated in the related works, as well as given the asymmetry of the FIEGARCH process that makes the QMLE less accurate (other works focus on symmetric processes). As mentioned above, optimization of the estimation setup should improve the overall performance and lead to absolute results closer to those in the benchmark studies. Next, focusing on the filter choice, the relative performance of the Haar, D4 and LA8 filters seem to be in compliance with that in Percival & Walden (2000), as well as it supports the conclusion in Jensen (1999) that Haar can be dominated by longer filters. To sum it up, the current work seems to extend the current literature without any contradiction with earlier works. Given the lack of related studies, this seems to be good news. Though, to make any strong conclusions about the WWE performance in various applications, a lot of work has to be done in the future.

## 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 cases performs best in terms of efficiency. The Whittle estimators outperform the MLE in some cases, but usually it is in situations with negative memory, which is not of a great interest for most practical applications. 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, but further research in this area is needed.

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) that ensures some minimal number of coefficients at the highest level of decomposition, 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 a faster, but slightly less accurate alternatives. From these two FWE performs slightly better. Yet, we believe that after optimization of the estimation and

forecasting algorithms, the differences between the FWE and WWE disappear, or even the WWE becomes superior in some cases.

Finally, we discuss the effects of optimization algorithm choice on the experiment results. It is argued that in cases, when the identification of individual parameters in the objective function is problematic, as is the case of the Whittle estimators applied in this work, the performance of the estimator and of the optimization algorithm cannot be well separated. Based on a comparison of our results with those of Perez & Zaffaroni (2008), it is argued that application of more sophisticated optimization algorithms to both FWE and WWE should improve their absolute performance, and potentially also change the conclusions about their relative performance. Therefore, the question of algorithm choice is an important topic to address in the future.

It can be concluded that after optimization of the estimation setup, the WWE may become a very attractive alternative to the traditional estimation methods. Although it is not as useful in case of jumps in the data as we expected, the statement that, compared to FWE, it is more robust to time-localized irregularities is still valid. The only additional requirement is that the irregularities remain detectable even after the data transformation that is necessary for the FWE and WWE application. Although a lot of work has to be done before the WWE applicability and performance will be fully assessed, importance of the research results for volatility modeling is a sufficient motivation.

Due to the pioneering nature of this work and the complexity of the problem concerned, the results presented are not intended to be directly projected to changes in estimation methods used in practice. For practitioners, the presented conclusions should be interesting as a message that given sufficient demand for further research in this area, new, possibly highly efficient methods based on wavelet transform could be available in the future. Though, the target group are the academics. It is believed that the results provided are a good basis for future research.

## References

- Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1), 3–30.
- Barunik, J. & Vacha, L. (2014). Realized wavelet-based estimation of integrated variance and jumps in the presence of noise. *Quantitative Finance*, forthcoming.
- 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(3), 307–327.
- Bollerslev, T. & Mikkelsen, H. O. (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, 73(1), 151 – 184.

- Bollerslev, T. & Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, 11, 143–172.
- Cheung, Y.-W. & Diebold, F. X. (1994). On maximum likelihood estimation of the differencing parameter of fractionally-integrated noise with unknown mean. *Journal of Econometrics*, 62(2), 301 – 316.
- Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, 17(4), 1749–1766.
- Dahlhaus, R. (2006). Correction: Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, 34(2), pp. 1045–1047.
- Donoho, D. L. & Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3), 425–455.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4), 987–1007.
- Engle, R. F. & Gonzalez-Rivera, G. (1991). Semiparametric arch models. *Journal of Business & Economic Statistics*, 9(4), 345–359.
- Faÿ, G., Moulines, E., Roueff, F., & Taqqu, M. S. (2009). Estimators of long-memory: Fourier versus wavelets. *Journal of Econometrics*, 151(2), 159 – 177.
- Fox, R. & Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series. *The Annals of Statistics*, 14(2), 517–532.
- Frederiksen, P. H. & Nielsen, M. O. (2005). Finite sample comparison of parametric, semi-parametric, and wavelet estimators of fractional integration. *Econometric Reviews*, 24(4), 405–443.
- Gonzaga, A. & Hauser, M. (2011). A wavelet whittle estimator of generalized long-memory stochastic volatility. *Statistical Methods & Applications*, 20(1), 23–48.
- Hardle, W., Kerkycharian, G., Tsybakov, A. B., & Picard, D. (1998). *Wavelets, Approximation and Statistical Applications*. John Wiley & Sons, Incorporated.
- Heni, B. & Mohamed, B. (2011). A wavelet-based approach for modelling exchange rates. *Statistical Methods & Applications*, 20(2), 201–220.
- Jensen, M. J. (1999). An approximate wavelet mle of short- and long-memory parameters. *Studies in Nonlinear Dynamics Econometrics*, 3(4), 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(3), 361 – 387.

- Johnstone, I. M. & Silverman, B. W. (1997). Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(2), 319–351.
- Mancini, C. & Calvori, F. (2012). *Jumps*, chapter 17, (pp. 403–445). John Wiley & Sons, Inc.
- Moulines, E., Roueff, F., & Taqqu, M. S. (2008). A wavelet whittle estimator of the memory parameter of a nonstationary gaussian time series. *The Annals of Statistics*, 36(4), 1925–1956.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2), 347–370.
- Nielsen, M. O. & Frederiksen, P. H. (2005). Finite sample comparison of parametric, semi-parametric, and wavelet estimators of fractional integration. *Econometric Reviews*, 24(4), 405–443.
- Percival, D. B. & Walden, A. T. (2000). *Wavelet Methods for Time Series Analysis (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press.
- Percival, D. P. (1995). On estimation of the wavelet variance. *Biometrika*, 82(3), 619–631.
- Perez, A. & Zaffaroni, P. (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.
- Whitcher, B. (2004). Wavelet-based estimation for seasonal long-memory processes. *Technometrics*, 46(2), 225–238.
- Whittle, P. (1962). Gaussian estimation in stationary time series. *Bulletin of the International Statistical Institute*, 39, 105–129.
- Wornell, G. W. & Oppenheim, A. (1992). Estimation of fractal signals from noisy measurements using wavelets. *Signal Processing, IEEE Transactions on*, 40(3), 611–623.
- Zaffaroni, P. (2009). Whittle estimation of egarch and other exponential volatility models. *Journal of Econometrics*, 151(2), 190–200.

## A FIEGARCH Transformation

As discussed in Perez & Zaffaroni (2008) and Zaffaroni (2009), to estimate the FIEGARCH model using FWE, it is necessary to rewrite the model in a signal plus noise form, for which

the spectral density can be derived. Let's begin with the original FIEGARCH(1,d,2) model:

$$\epsilon_t = z_t h_t^{1/2} \quad (29)$$

$$\ln(h_t) = \omega + \Phi(L)g(z_{t-1}) \quad (30)$$

$$g(z_t) = \theta z_t + \gamma[|z_t| - E(|z_t|)] \quad (31)$$

$$\Phi(L) = (1-L)^{-d}[1 + \alpha_{[2]}(L)][\beta(L)]^{-1} \quad (32)$$

where for FIEGARCH(1, d, 2)  $\alpha_{[2]}(L) = \alpha L$ ,  $\beta(L) = 1 - \beta L$ . Following Zaffaroni (2009), this can be rewritten as

$$\epsilon_t^2 = z_t^2 h_t \quad (33)$$

$$\ln(\epsilon_t^2) = \ln(z_t^2) + \ln(h_t) \quad (34)$$

$$\ln(h_t) = \omega + \Phi(L)g(z_{t-1}) \quad (35)$$

$$\Phi(L)g(z_{t-1}) = \sum_{s=0}^{\infty} \Phi_s g(z_{t-s-1}) \quad (36)$$

$$g(z_t) = \theta z_t + \gamma[|z_t| - E(|z_t|)], \quad (37)$$

which leads to

$$\ln(\epsilon_t^2) = \ln(z_t^2) + \omega + \sum_{s=0}^{\infty} \Phi_s g(z_{t-s-1}) \quad (38)$$

$$g(z_t) = \theta z_t + \gamma[|z_t| - E(|z_t|)] \quad (39)$$

$$\Phi(L) = (1-L)^{-d}[1 + \alpha L][1 - \beta L]^{-1}. \quad (40)$$

From Perez & Zaffaroni (2008), the spectral density of this process for  $z_t \sim N(0, 1)$ ,  $\omega = 0$  and simple Fourier frequencies  $\lambda_j = j/T$ ;  $\lambda_j \in \langle -1/2, 1/2 \rangle$  is

$$\begin{aligned} f(\lambda_j, v) &= \frac{A(\xi)}{2\pi} + \frac{B(v)}{2\pi} \left| \sum_{s=0}^{\infty} \Phi_s(\zeta) e^{2\pi i s \lambda_j} \right|^2 + \\ &+ \frac{C(v)}{2\pi} \left( e^{2\pi i \lambda_j} \sum_{s=0}^{\infty} \Phi_s(\zeta) e^{2\pi i s \lambda_j} \right) + \\ &+ \frac{C(v)}{2\pi} \left( e^{-2\pi i \lambda_j} \sum_{s=0}^{\infty} \Phi_s(\zeta) e^{-2\pi i s \lambda_j} \right) \end{aligned} \quad (41)$$

where  $A(\xi) = \text{var}(\ln(z_0^2))$ ,  $B(v) = \text{var}(g(z_0))$ ,  $C(v) = \text{cov}(\ln(z_0^2), g(z_0))$ ,  $v = (\xi', \zeta')'$ , and  $\xi$

is the vector of parameters in the distribution function of variable  $z$ . For  $z_t \sim N(0, 1)$ :

$$A(\xi) = \Psi\left(\frac{1}{2}\right) \quad (42)$$

$$B(v) = \theta_2 + \delta^2(1 - \mu_{|z|}^2) \quad (43)$$

$$C(v) = \delta\mu_z(\psi(1) - \psi(\frac{1}{2})) \quad (44)$$

$$\mu_{|z|} = E(|z|) = \sqrt{\frac{2}{\pi}}, \quad (45)$$

where  $\psi(x)$  and  $\Psi(x)$  are digamma and trigamma functions respectively. Evaluated at Fourier frequencies, this spectral density occurs in both terms of the FWE objective function. For a generalization to  $z_t$  following GED or Student-t distribution, see Perez & Zaffaroni (2008).

## B Spectral Density Estimation

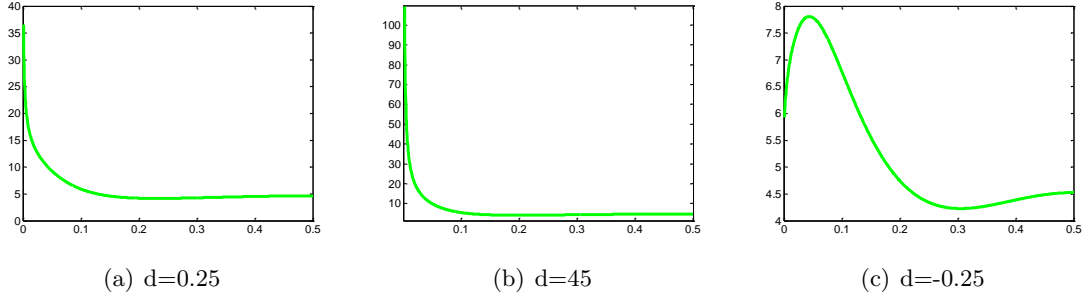
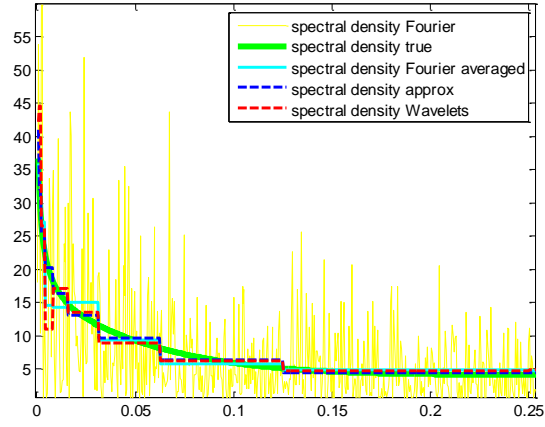
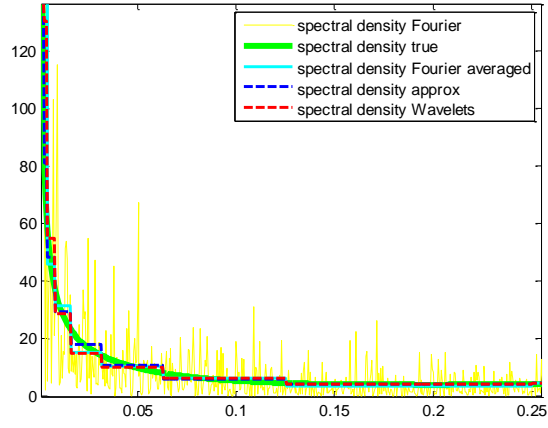


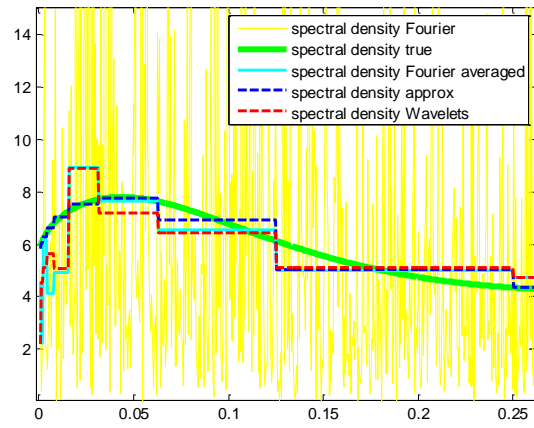
Figure 1: True spectral density



(a)  $d=0.25$



(b)  $d=0.45$



(c)  $d=-0.25$

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



## C Partial Decomposition

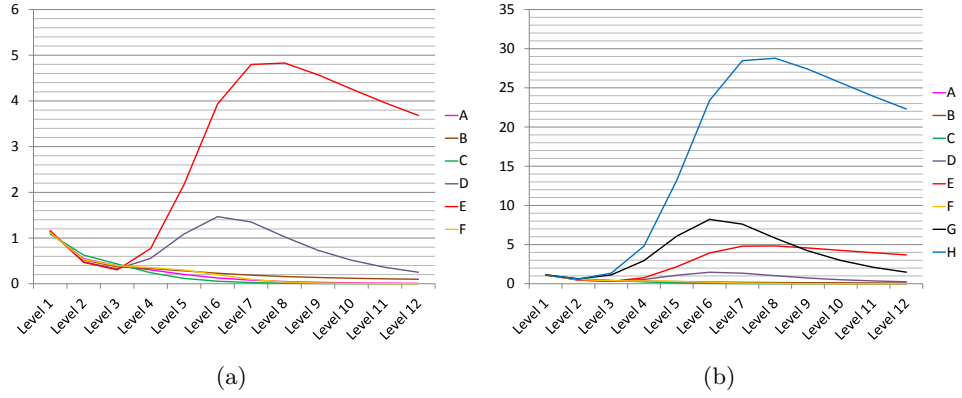


Figure 3: Energy decomposition: Integrals of FIEGARCH spectral density over frequency intervals respective to individual levels of decomposition, assuming various levels of long memory ( $d=0.25$ ,  $d=0.45$ ,  $d=-0.25$ ) and the coefficient sets from Table E.1 (a)

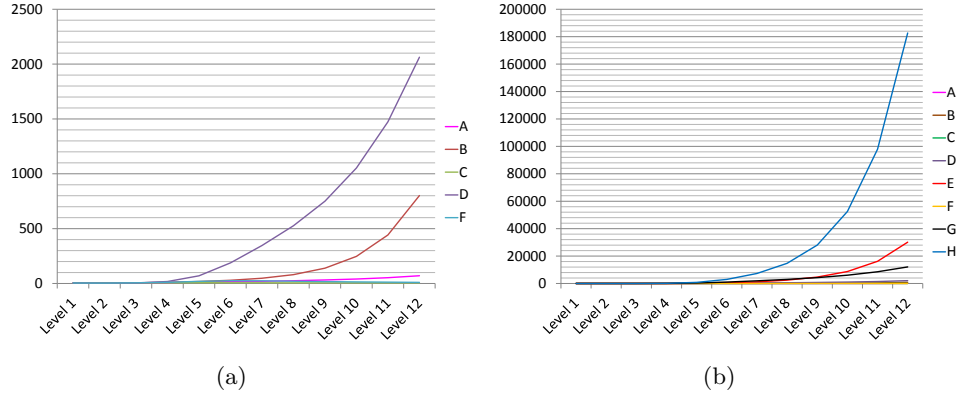


Figure 4: Energy decomposition: True variances of wavelet coefficients respective to individual levels of decomposition and various levels of long memory ( $d=0.25$ ,  $d=0.45$ ,  $d=-0.25$ ) and the coefficient sets from Table E.1 (a)

(a) Coefficient Sets

Coefficients:	d	$\omega$	$\alpha$	$\beta$	$\theta$	$\gamma$
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

(b) Integrals over frequencies respective to levels for the coefficient sets from Tab.E.1  
(a)

	A	B	C	D	E	F	G	H
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

(c) Sample Variances of DWT Wavelet Coefficients for the coefficient sets from Table E.1 (a)

	A	B	C	D	E	F	G	H
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 1: Energy decomposition

## D Convergence analysis

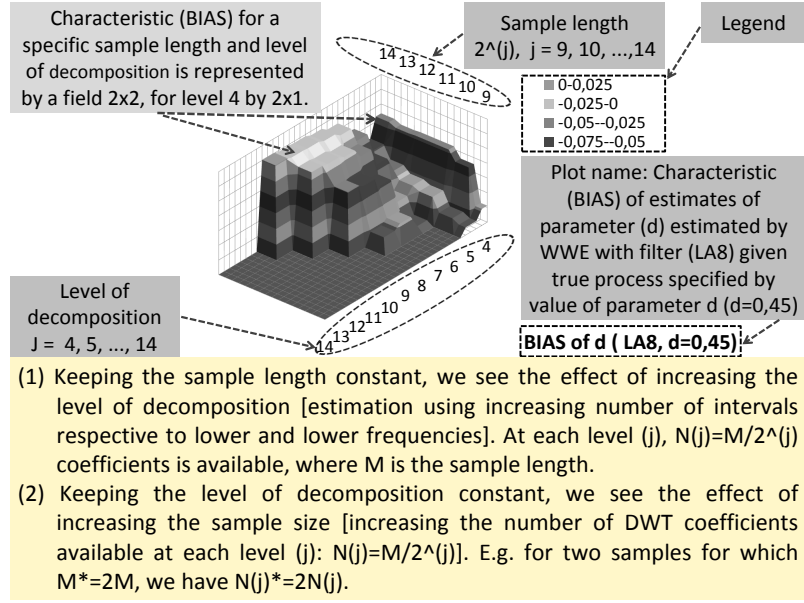


Figure 5: 3D Plots Guide

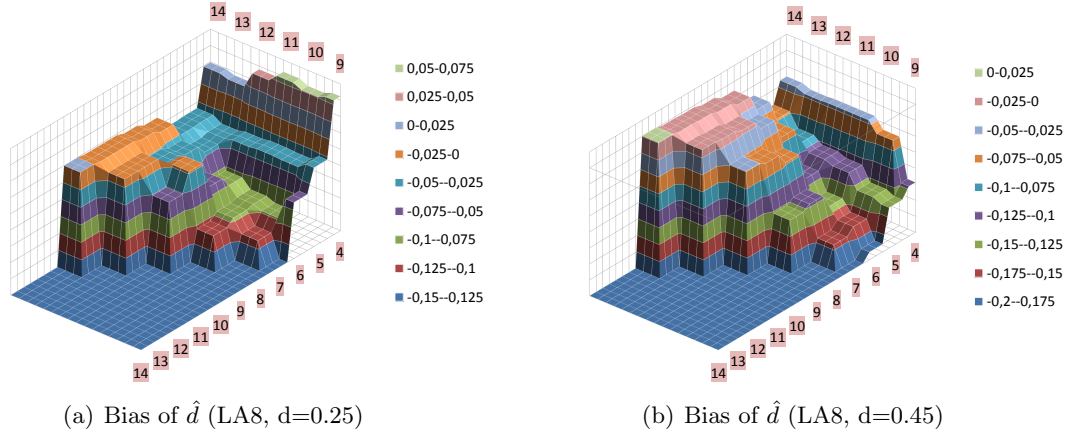
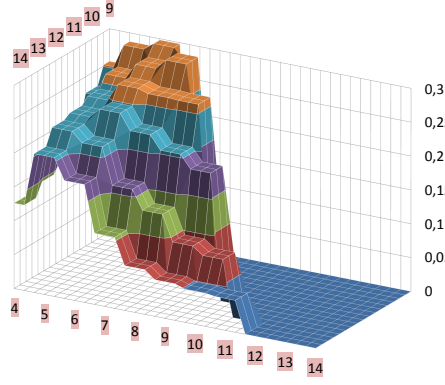
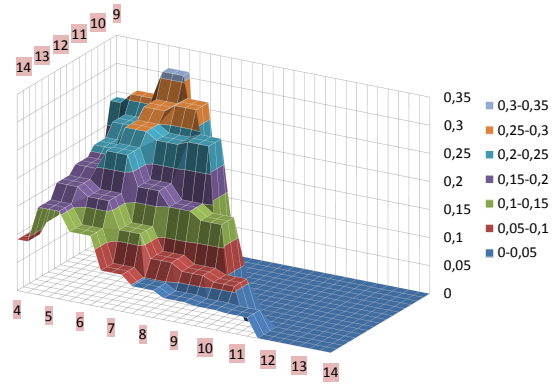


Figure 6: 3D Plots: Partial decomposition:  $\hat{d}$ : Bias

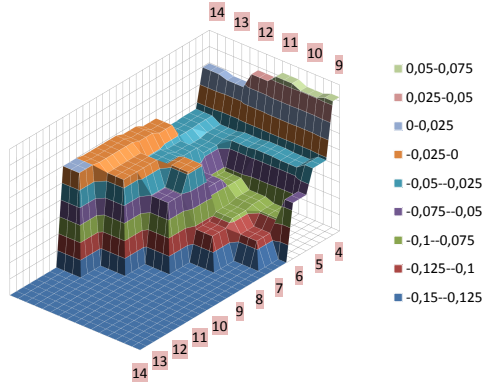


(a) RMSE of  $\hat{d}$  (LA8,  $d=0.25$ )

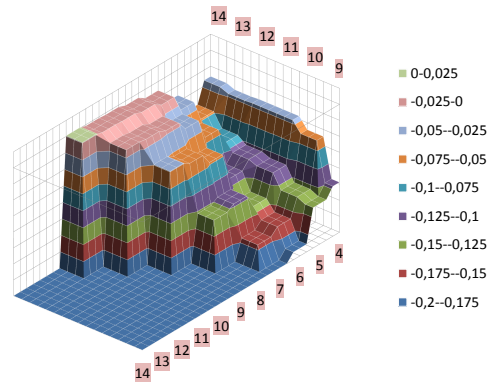


(b) RMSE of  $\hat{d}$  (LA8,  $d=0.45$ )

Figure 7: 3D Plots: Partial decomposition:  $\hat{d}$ : RMSE

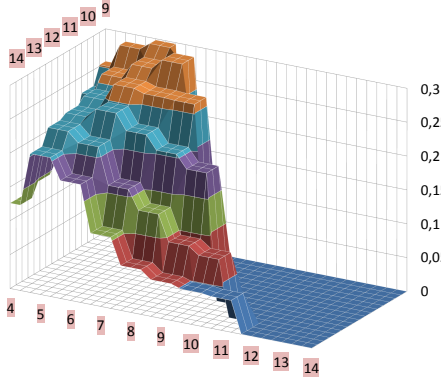


(a) Bias of  $\hat{\alpha}$  (LA8,  $d=0.25$ )

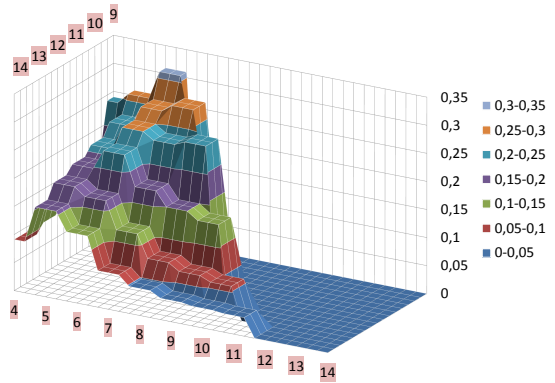


(b) Bias of  $\hat{\alpha}$  (LA8,  $d=0.45$ )

Figure 8: 3D Plots: Partial decomposition:  $\hat{\alpha}$ : Bias



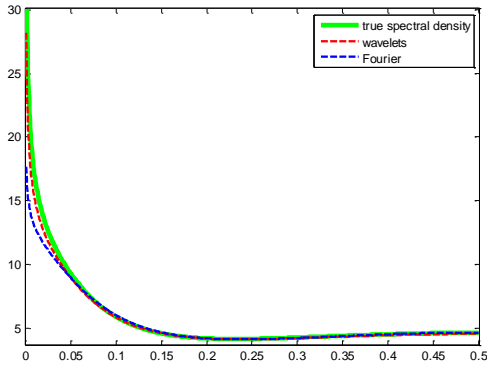
(a) RMSE of  $\hat{\alpha}$  (LA8,  $d=0.25$ )



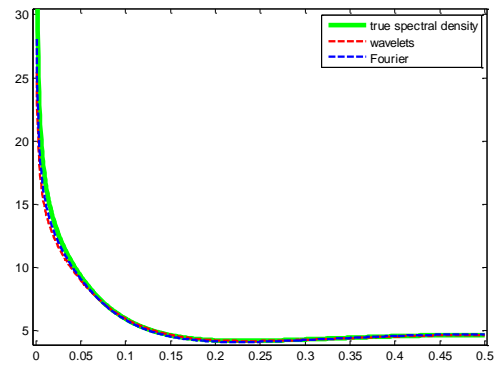
(b) RMSE of  $\hat{\alpha}$  (LA8,  $d=0.45$ )

Figure 9: 3D Plots: Partial decomposition:  $\hat{\alpha}$ : RMSE

## E Relative Performance



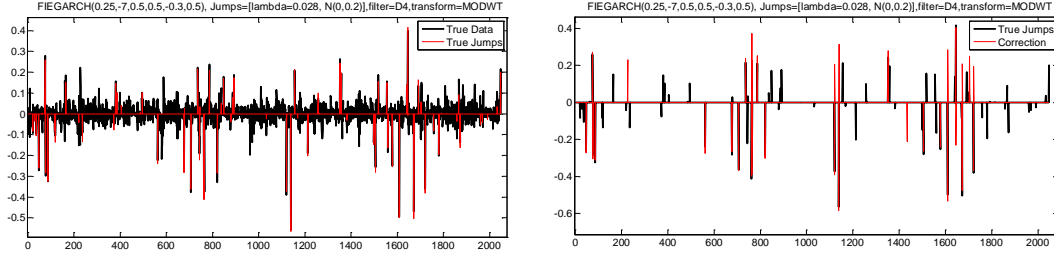
(a)  $T=512$  ( $2^9$ ), level=5(D4)



(b)  $T=2048$  ( $2^{11}$ ), level=5(D4)

Figure 10: Spectral density estimation: Wavelets (Level 5) vs Fourier

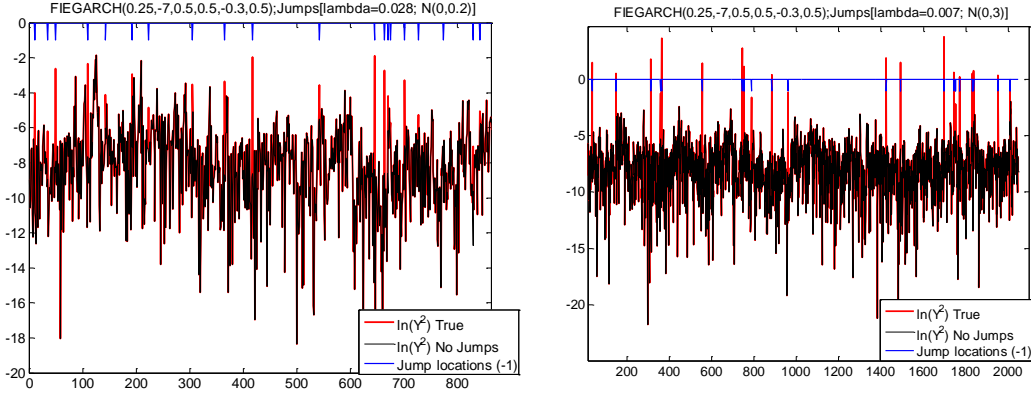
## F Jumps: Features



(a) Effects of jumps in the time series

(b) Jump detection

Figure 11: Jumps in the data



(a)  $\lambda=0.028$ ;  $N(0,0.2)$

(b) Stronger jumps:  $\lambda=0.007$ ;  $N(0,3)$

Figure 12: Jumps after transformation  $\ln(y^2)$

## G Jumps: Results

PAR	TRUE	METHOD	No Jumps;N=2048				No Jumps;N=16384				No Jumps;N=2048			
			MEAN	BIAS	RMSE		MEAN	BIAS	RMSE	PAR	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 2: Monte Carlo No Jumps:  $d=0.25/0.45$ ; MLE, 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=16384				Jump[0.028,N(0,0.2)];N=2048			
			MEAN	BIAS	RMSE		MEAN	BIAS	RMSE	PAR	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 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



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		-	-	-	-
	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
out	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
	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	
out	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 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	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 5: Forecasting: N=16384; 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	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 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	0.0024292	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					
	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
out	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
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 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	0.00079621	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									
out	WWE MODWT	0.01	0.05	0.10		0.90	0.95		0.99
	WWE DWT	-0.0025009	-0.0006746	-0.00028513		0.0011776	0.0033089		0.088609
	FWE	-0.0025573	-0.00057909	-0.00028814		0.0012299	0.0033547		0.043464
	MLE	-0.0018013	-0.00048645	-0.00022585		0.0011465	0.0030639		0.04528
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 %									

Table 8: Forecasting: N=16384; Jumps Lambda=0.028, N(0, 0.2); MLE, FWE, MODWT(D4), DWT(D4), 2: Corrected using Donoho and Johnstone threshold