Theoretical Results on Fractionally Integrated Exponential Generalized Autoregressive Conditional Heteroskedastic Processes

Sílvia R.C. Lopes[1,2], and Taiane S. Prass[2,*]

Mathematics Institute
Federal University of Rio Grande do Sul
Porto Alegre - RS, Brazil

Abstract

Here we present a theoretical study on the main properties of Fractionally Integrated Exponential Generalized Autoregressive Conditional Heteroskedastic (FIEGARCH) processes. We analyze the conditions for the existence, the invertibility, the stationarity and the ergodicity of these processes. We prove that, if $\{X_t\}_{t\in\mathbb{Z}}$ is a FIEGARCH(p,d,q) process then, under mild conditions, $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is an ARFIMA$(q,d,0)$ with correlated innovations, that is, an autoregressive fractionally integrated moving average process. The convergence order for the polynomial coefficients that describes the volatility is presented and results related to the spectral representation and to the covariance structure of both processes $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ and $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ are discussed. Expressions for the kurtosis and the asymmetry measures for any stationary FIEGARCH(p,d,q) process are also derived. The h-step ahead forecast for the processes $\{X_t\}_{t\in\mathbb{Z}}$, $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ and $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ are given with their respective mean square error of forecast. The work also presents a Monte Carlo simulation study showing how to generate, estimate and forecast based on six different FIEGARCH models. The forecasting performance of six models belonging to the class of autoregressive conditional heteroskedastic models (namely, ARCH-type models) and radial basis models is compared through an empirical application to Brazilian stock market exchange index.

Keywords: Long-Range Dependence, Volatility, Stationarity, Ergodicity, FIEGARCH Processes.
2000 MSC: 60G10, 62G05, 62G35, 62M10, 62M15, 62M20

1. Introduction

Financial time series present an important characteristic known as volatility which can be defined/measured in different ways but is not directly observable. A common approach, but not unique, is to define the volatility as the conditional standard deviation (or the conditional variance) of the process and use heteroskedastic models to describe it. The two main classes of econometric models used for representing the dynamic evolution of volatilities are the Autoregressive Conditional Heteroskedastic models, also called ARCH-type (introduced by [1]) and Stochastic Volatility (SV) models (see, [2] and references therein).

Both ARCH-type and SV models assume that the stochastic process $\{X_t\}_{t\in\mathbb{Z}}$ is written as

$$X_t = \sigma_t Z_t, \quad \text{for all } t \in \mathbb{Z}, \tag{1}$$

where $\{Z_t\}_{t\in\mathbb{Z}}$ is a sequence of independent identically distributed (i.i.d.) random variables, with zero mean and variance equal to one, and $\sigma_t := \mathrm{Var}(X_t|\mathcal{F}_{t-1})$, where $\mathcal{F}_{t-1}$ denotes the sigma field generated

*Corresponding author. E-mail: taianeprass@gmail.com
[1]E-mail: silvia.lopes@ufrgs.br
[2]Mathematics Institute, Federal University of Rio Grande do Sul, 91509-900, Porto Alegre - RS, Brazil. Fax: 55-51-3308-7301

by the past information until time $t-1$. More specifically, for ARCH-type models, $\mathcal{F}_t := \sigma(\{X_s\}_{s\leq t})$ or $\mathcal{F}_t := \sigma(\{Z_s\}_{s\leq t})$, while for SV models $\mathcal{F}_t := \sigma(\{Z_s, \eta_s\}_{s\leq t})$, where $\{\eta_t\}_{t\in\mathbb{Z}}$ is a sequence of latent variables, independent of $\{Z_t\}_{t\in\mathbb{Z}}$, but which is not directly observable. Since the volatility of SV models is specified as a latent variable, parameter estimation is much more challenging then for ARCH-type ones.

By ARCH-type models we mean not only the ARCH(p) model proposed by [1], where

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2, \quad \text{for all } t \in \mathbb{Z},$$

(which characterizes the volatility as a function of powers of past observed values, consequently, the volatility can be observed one-step ahead), but also the several generalizations that were lately proposed to properly model the dynamics of the volatility.

Among the generalizations of the ARCH model are the Generalized ARCH process (see [3]), denoted by GARCH(p,q), and the Exponential GARCH process (see [4]), denoted by EGARCH(p,q), for which σ_t^2 is, respectively defined through

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \quad \text{and} \quad \ln(\sigma_t^2) = \omega + \frac{\alpha(\mathcal{B})}{\beta(\mathcal{B})} g(Z_{t-1}), \quad \text{for all } t \in \mathbb{Z},$$

where $\mathcal{B}$ is the backward shift operator defined by $\mathcal{B}^k(X_t) = X_{t-k}$, for all $k \in \mathbb{N}$, $g(\cdot)$ is defined by

$$g(Z_t) = \theta Z_t + \gamma\left[|Z_t| - \mathbb{E}(|Z_t|)\right], \quad \text{for any } \theta, \gamma \in \mathbb{R}, \tag{2}$$

$\alpha(\cdot)$ and $\beta(\cdot)$ are, respectively, defined by

$$\alpha(z) = \sum_{i=0}^{p} (-\alpha_i) z^i = 1 - \sum_{i=1}^{p} \alpha_i z^i \quad \text{and} \quad \beta(z) = \sum_{j=0}^{q} (-\beta_j) z^j = 1 - \sum_{j=1}^{q} \beta_j z^j, \tag{3}$$

$\alpha_0 = \beta_0 = -1$ and $\beta(z) \neq 0$, if $|z| \leq 1$. It is also assumed that $\alpha(\cdot)$ and $\beta(\cdot)$ have no common roots so that the operator $\alpha(\mathcal{B})[\beta(\mathcal{B})]^{-1}$ is well defined.

The main advantages of EGARCH models over ARCH/GARCH ones are the fact that the volatility is specified in terms of the logarithm function, assuring its positivity, and their capability of modeling the volatility's asymmetry[3]. ARCH and GARCH models do not account for asymmetry since they consider only the squared returns in the volatility's definition. On the other hand, it is easy to see that the function $g(\cdot)$, given in the EGARCH definition, can be rewritten as

$$g(Z_t) = \begin{cases} (\theta+\gamma)Z_t - \gamma\mathbb{E}(|Z_t|), & \text{if } Z_t \geq 0; \\ (\theta-\gamma)Z_t - \gamma\mathbb{E}(|Z_t|), & \text{if } Z_t < 0. \end{cases}$$

This expression clearly shows the asymmetry in response to positive and negative returns. Also, it is easy to see that $g(\cdot)$ is non-linear if $\theta \neq 0$ and the asymmetry is due to the values of $\theta \pm \gamma$. While the parameter θ, also known in the literature as *leverage parameter*, shows the return's sign effect, the parameter γ denotes the return's magnitude effect. Therefore, the model is able to capture the fact that a negative return usually results in higher volatility than a positive one.

ARCH, GARCH and EGARCH are all short-range dependence models. Among the ARCH-type models that capture the effects of long-range dependence characteristic in the conditional variance are the Fractionally Integrated GARCH processes (see [5]), denoted by FIGARCH(p,d,q), and the Fractionally Integrated EGARCH processes (see [6]), denoted by FIEGARCH (p,d,q). These models generalize, respectively, the

[3]By asymmetry we mean that the volatility reacts in an asymmetrical form to the returns, that is, volatility tends to rise in response to "bad" news and to fall in response to "good" news.

GARCH and EGARCH models, which are obtained, respectively, from (4) and (5), by letting $d = 0$. For a FIGARCH(p, d, q), σ_t^2 is given by

$$\sigma_t^2 = \omega[\beta(\mathcal{B})]^{-1} + \big(1 - [\beta(\mathcal{B})]^{-1}\phi(\mathcal{B})(1-\mathcal{B})^d\big)X_t^2, \quad \text{for all } t \in \mathbb{Z}, \tag{4}$$

while for a FIEGARCH(p, d, q), σ_t^2 is defined through the relation,

$$\ln(\sigma_t^2) = \omega + \frac{\alpha(\mathcal{B})}{\beta(\mathcal{B})}(1-\mathcal{B})^{-d} g(Z_{t-1}), \quad \text{for all } t \in \mathbb{Z}, \tag{5}$$

with $\phi(\mathcal{B}) = \sum_{k=0}^{p} \phi_k \mathcal{B}^k$, where $\phi_0 = 0$, $\alpha(\cdot)$ and $\beta(\cdot)$ given in (3), $g(\cdot)$ given in (2) and $(1-\mathcal{B})^d$ defined by its Maclaurin series expansion, namely,

$$(1-\mathcal{B})^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} := \sum_{k=0}^{\infty} \delta_{d,k}\, \mathcal{B}^k := \delta_d(\mathcal{B}), \tag{6}$$

where $\Gamma(\cdot)$ is the gamma function, and the coefficients $\delta_{d,k}$ are such that $\delta_{d,0} = 1$ and $\delta_{d,k} = \delta_{d,k-1}\big(\frac{k-1-d}{k}\big)$, for all $k \geq 1$.

Remark 1. Expression (6) is valid only for non-integer values of d. When $d \in \mathbb{N}$, $(1-\mathcal{B})^d$ is merely the difference operator $1 - \mathcal{B}$ iterated d times. Also, one observe that, upon replacing d by $-d$, the operator $(1-\mathcal{B})^{-d}$ has the same binomial expansion as the polynomial given in (6), that is

$$(1-\mathcal{B})^{-d} = \sum_{j=0}^{\infty} \delta_{-d,j}\mathcal{B}^j := \sum_{k=0}^{\infty} \pi_{d,k}\mathcal{B}^k, \tag{7}$$

where $\pi_{d,j} = \delta_{-d,j}$, for all $j \in \mathbb{N}$.

It is easy to see that, by letting $\lambda(\cdot)$ be the polynomial defined by

$$\lambda(z) = \frac{\alpha(z)}{\beta(z)}(1-z)^{-d} := \sum_{k=0}^{\infty} \lambda_{d,k} z^k, \qquad \text{for all } |z| < 1, \tag{8}$$

where $\alpha(\cdot)$ and $\beta(\cdot)$ are defined in (3), one can rewrite (5), equivalently (see Remark 2), as

$$\ln(\sigma_t^2) = \omega + \sum_{k=0}^{\infty} \lambda_{d,k} g(Z_{t-1-k}) = \omega + \lambda(\mathcal{B})g(Z_{t-1}), \quad \text{for all } \ t \in \mathbb{Z}. \tag{9}$$

Remark 2. Since it is assumed that $\beta(\cdot)$ has no roots in the closed disk $\{z : |z| \leq 1\}$, and also $\alpha(\cdot)$ and $\beta(\cdot)$ have no common roots, the function $\lambda(z)$ is analytic in the open disk $\{z : |z| < 1\}$ (if $d \leq 0$, in the closed disk $\{z : |z| \leq 1\}$). Therefore, the power series representation for $\lambda(\cdot)$ is unique.

FIEGARCH models have not only the capability of modeling clusters of volatility (as ARCH and GARCH models do) and capturing its asymmetry (as EGARCH models do) but they also take into account the characteristic of long-range dependence in the volatility (as FIGARCH models do). Besides non-stationarity (in the weak sense), FIGARCH(p, d, q) models have the drawback that we must have $d \geq 0$ and the polynomial coefficients in their definition must satisfy some restrictions so that the conditional variance will be positive. FIEGARCH(p, d, q) models do not have this problem since the variance is defined in terms of the logarithm function and, as we shall prove here, they are weakly stationary when $d < 0.5$.

Although, in practice, often a simple FIEGARCH(p, d, q) model with $p, q \in \{0, 1\}$ suffices to fully describe financial time series (for instance, [7] and [8], consider FIEGARCH$(0, d, 1)$ models and [9] considers FIEGARCH$(1, d, 1)$ models), there are evidences that for some financial time series higher values of p and q are in fact necessary (see, for instance, [10],[11] and [12]). In view of that, the theoretical study carried on here considers FIEGARCH(p,d,q), with any $p,q \geq 0$. One of the contributions of this paper is to extend, for any p and q, the results already known in the literature for $d=0$ and/or $p,q \in \{0,1\}$. More specifically, we show under which conditions the results regarding existence, stationarity and ergodicity in [4] (case $d=0$) also apply when $d \neq 0$. Moreover, we provide the expressions for the asymmetry and kurtosis measures of FIEGARCH(p,d,q) processes, for all $p,q \geq 0$, extending the results in [8] (these authors considered only the case $p=0$ and $q=1$ and only the kurtosis measure was derived).

Another contribution of this work is to derive conditions under which $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is well defined when $\{X_t\}_{t\in\mathbb{Z}}$ is a FIEGARCH process. We also show that, under mild conditions, it has an ARFIMA representation. To the best of our knowledge, this result is absent in the literature. This representation is very useful in model identification and parameter estimation since the literature regarding ARFIMA models is well developed (see [13] and references therein). Since (1) implies that $\ln(X_t^2) = \ln(\sigma_t^2) + \ln(Z_t^2)$, for all $t \in \mathbb{Z}$, to derive the properties of $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, we first analyze the properties of the non-observable process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$. Studying the characteristics of $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ includes deriving the properties of the polynomial $\lambda(\cdot)$, given in (8). The conditions for the existence of a power series representation for $\lambda(\cdot)$ and the behavior of the coefficients in this representation are fundamental not only for simulation purposes but also to draw conclusions on the autocorrelation and spectral density functions decay of the processes $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ and $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$. We also provide a recurrence formula to calculate the coefficients of the series expansion of $\lambda(\cdot)$, for any $p,q \geq 0$. This recurrence formula allows to easily simulate FIEGARCH processes.

The fact that $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is an ARFIMA(q,d,p) process and the result that any FIEGARCH process is a martingale difference with respect to the natural filtration $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$, where $\mathcal{F}_t := \sigma(\{Z_s\}_{s\leq t})$, are applied to obtain the h-step ahead predictor for the processes $\{X_t\}_{t\in\mathbb{Z}}$ and $\{X_t^2\}_{t\in\mathbb{Z}}$. We also derive h-step ahead predictors for both processes $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ and $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, considering different approaches. First, a h-step ahead predictor is obtained by considering the conditional expectation method then another h-step ahead predictor is derived by considering a second order Taylor expansion for the logarithm function. The relation between these predictors is also derived. To the best of our knowledge, for FIEGARCH processes, the predictor based on the second order Taylor's expansion for the logarithm function was never proposed in the literature.

This work also contributes to the literature of FIEGARCH models by presenting a simulation study including generation, estimation and forecasting features of FIEGARCH models. One of the goals in this study is to analyze the finite sample performance of the quasi-likelihood estimation procedure. We shall notice that, despite the fact that the quasi-likelihood is one of the most applied methods in non-linear process estimation, asymptotic results for FIEGARCH processes are still an open question (see [14])[4]. Another goal of this study is to analyze the forecasting performance of the fitted models.

The paper is organized as follows: Section 2 presents the theoretical properties of FIEGARCH(p,d,q) process. In particular, a recurrence formula to obtain the coefficients $\{\lambda_{d,k}\}_{k\in\mathbb{Z}}$ in the power series expansion of the polynomial $\lambda(\cdot)$ is provided and their asymptotic behavior is derived. The autocovariance and spectral density functions of the processes $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ and $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ are presented and analyzed. The asymmetry and kurtosis measures of any stationary FIEGARCH process are also presented. Section 3 presents the theoretical results regarding forecasting. Section 4 presents a Monte Carlo simulation study including the generation of FIEGARCH time series, estimation of the model parameters and the forecasting procedure based on the fitted model. Section 5 presents the analysis of an observed time series and the comparison of the forecasting performance for different ARCH-type and radial basis models. Section 6 concludes the paper.

2. FIEGARCH Process

In this section we discuss the properties of the *Fractionally Integrated Exponential Generalized Autoregressive Conditional Heteroskedastic* process (FIEGARCH), defined by (1) and (5). As mentioned before,

[4]The asymptotic properties for the quasi-likelihood method are well established for ARCH/GARCH models (see, for instance, [15], [16], [17], [18] and [19]) and also for EGARCH models (see, for instance, [20]).

this class of processes, introduced by [6], describes not only the volatility varying in time and the volatility clusters (known as ARCH/GARCH effects) but also the volatility long-range dependence and its asymmetry.

For the sake of completeness, results previously established in the literature are also stated and the references of their proofs are provided. Due to the extent of the work, all proofs are presented in the appendix. Hereafter, given $a \in \mathbb{R} \cup \{-\infty, +\infty\}$, $f(x) = O(g(x))$ means that $|f(x)| \leq c|g(x)|$, for some $c > 0$, as $x \to a$; $f(x) = o(g(x))$ means that $f(x)/g(x) \to 0$, as $x \to a$; $f(x) \sim g(x)$ means that $f(x)/g(x) \to 1$, as $x \to a$. We also say that $f(x) \approx g(x)$, as $x \to \infty$, if for any $\varepsilon > 0$, there exists $x_0 \in \mathbb{R}$ such that $|f(x) - g(x)| < \varepsilon$, for all $x \geq x_0$. Also, given any set T, T^* corresponds to the set $T\backslash\{0\}$ and $\mathbb{I}_A(\cdot)$ is the indicator function defined as $\mathbb{I}_A(z) = 1$, if $z \in A$, and 0, otherwise.

For practical purpose, it is important to observe that slightly different definitions of FIEGARCH processes are found in the literature. Usually it is easy to show that, under certain conditions, the different definitions for the conditional variance are equivalent to (5). For instance, [21] defines σ_t^2 through the equation

$$\beta(\mathcal{B})(1-\mathcal{B})^d \ln(\sigma_t^2) = a + \sum_{i=0}^{p} [\psi_i |Z_{t-1-i}| + \gamma_i Z_{t-1-i}],$$

with $\beta(\cdot)$ given in (3). This is the definition considered, for instance, in the software S-Plus (see [21]) and it is equivalent to (5) whenever $d > 0$, $a = -\gamma \mathbb{E}(|Z_t|)\alpha(1)$, $\psi_0 = \gamma$, $\gamma_0 = \theta$, $\psi_i = -\gamma\alpha_i$ and $\gamma_i = -\theta\alpha_i$ for all $1 \leq i \leq p$. This equivalence is mentioned in [14] and a detailed proof is provided in [10]. In [8] only the case $p = 0$ and $q = 1$ is considered and $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is defined as

$$(1-\phi\mathcal{B})(1-\mathcal{B})^d \ln(\sigma_t^2) = \omega^* + \alpha\big[|Z_{t-1}| - \sqrt{2/\pi}\big] + \gamma^* Z_{t-1}, \quad \text{for all } t \in \mathbb{Z}, \tag{10}$$

where $\{Z_t\}_{t\in\mathbb{Z}}$ is a Gaussian white noise process with variance equal to one. This is the definition considered, for instance, in the G@RCH package version 4.0 of [22]. Notice that, by setting $\phi = \beta_1$, $\alpha = \gamma$ and $\gamma^* = \theta$, (10) is equivalent to (5) if and only if the equality $\omega^* = (1-\beta)(1-\mathcal{B})^d\omega$ holds. Example 1 presents two FIEGARCH time series obtained by considering the definition obtained through (1) and (5).

Remark 3. Henceforth, GED shall stands for the so-called Generalized Error Distribution. Whenever we consider $Z_0 \sim \text{GED}(\nu)$, and ν is the tail-thickness parameter, we assume that the random variable was normalized to have mean zero and variance equal to one. In this case, the probability density function of Z_0 is given by

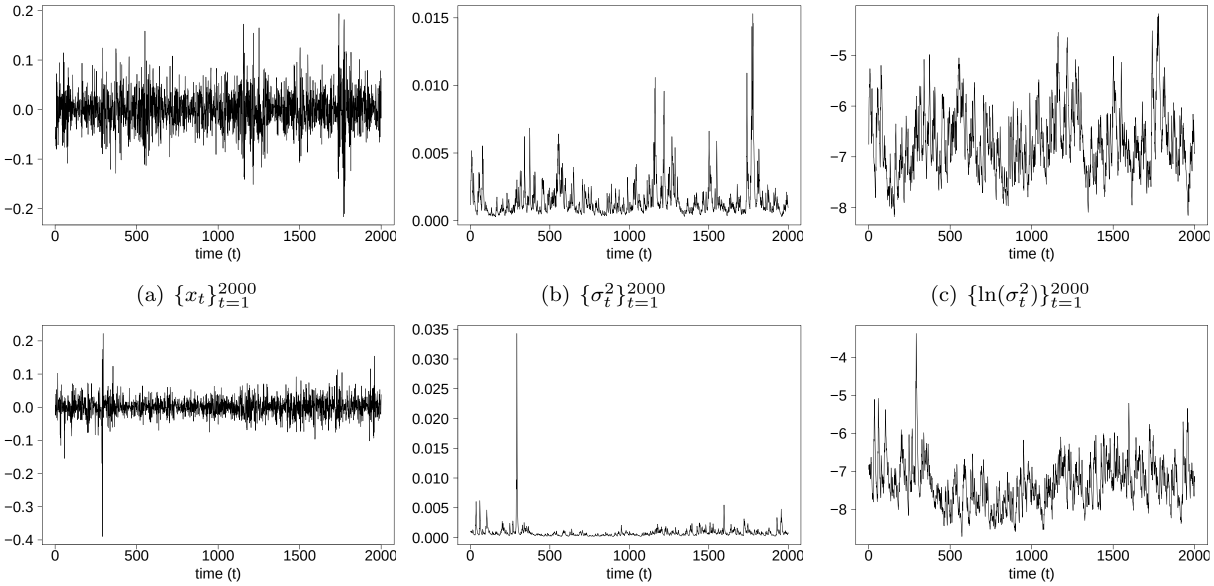
$$p_Z(z|\nu) = \frac{\nu \exp\big\{-\frac{1}{2}|z\lambda_\nu^{-1}|^\nu\big\}}{\lambda_\nu 2^{1+1/\nu}\Gamma(1/\nu)} \quad \text{where } \lambda_\nu = \left[2^{-2/\nu}\frac{\Gamma(1/\nu)}{\Gamma(3/\nu)}\right]^{1/2}, \qquad \text{for all } z \in \mathbb{R}.$$

Example 1. Figure 1 presents samples from FIEGARCH$(0, d, 1)$ processes, with $n = 2000$ observations, considering two different underlying distributions for Z_0. In both cases we set $d = 0.3578$, $\theta = -0.1661$, $\gamma = 0.2792$, $\omega = -7.2247$ and $\beta_1 = 0.6860$. These are the parameter values of the FIEGARCH model fitted to the Bovespa index log-returns in Section 5. Figures 1 (a) - (c) consider $Z_0 \sim \mathcal{N}(0, 1)$ and show, respectively, the time series $\{x_t\}_{t=1}^n$, the conditional variance $\{\sigma_t^2\}_{t=1}^n$ and the logarithm of the conditional variance $\{\ln(\sigma_t^2)\}_{t=1}^n$. Figures 1 (d) - (e) show the same time series as in Figures 1 (a) - (c) when $Z_0 \sim \text{GED}(1.5)$.

Proposition 1 below presents the properties of the stochastic process $\{g(Z_t)\}_{t\in\mathbb{Z}}$. Although the proof is straightforward, these properties are extremely important to prove the results stated in the sequel.

Proposition 1 (Prass, 2008). *Let $\{Z_t\}_{t\in\mathbb{Z}}$ be a sequence of i.i.d. random variables, with $\mathbb{E}(|Z_0|) < \infty$. Let $\{g(Z_t)\}_{t\in\mathbb{Z}}$ be defined by (2) and assume that θ and γ are not both equal to zero. Then $\{g(Z_t)\}_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic process. If $\mathbb{E}(Z_0^2) < \infty$, then $\{g(Z_t)\}_{t\in\mathbb{Z}}$ is also weakly stationary with zero mean (therefore a white noise process) and variance σ_g^2 given by*

$$\sigma_g^2 = \theta^2 + \gamma^2 - [\gamma\mathbb{E}(|Z_0|)]^2 + 2\,\theta\,\gamma\,\mathbb{E}(Z_0|Z_0|). \tag{11}$$



(a) $\{x_t\}_{t=1}^{2000}$ (b) $\{\sigma_t^2\}_{t=1}^{2000}$ (c) $\{\ln(\sigma_t^2)\}_{t=1}^{2000}$

(d) $\{x_t\}_{t=1}^{2000}$ (e) $\{\sigma_t^2\}_{t=1}^{2000}$ (f) $\{\ln(\sigma_t^2)\}_{t=1}^{2000}$

Figure 1: Samples from FIEGARCH$(0, d, 1)$ processes, with $n = 2000$ observations, considering $Z_0 \sim \mathcal{N}(0, 1)$ (first row) and $Z_0 \sim \text{GED}(1.5)$ (second row). (a) and (d) show the time series $\{x_t\}_{t=1}^{n}$; (b) and (e) show the conditional variance of $\{x_t\}_{t=1}^{n}$; (c) and (f) show the logarithm of the conditional variance.

Remark 4. We observe that the theory presented in this work can be easily adapted to a more general framework than (5) by considering

$$\ln(\sigma_t^2) = \omega_t + \sum_{k=0}^{\infty} \lambda_k g(Z_{t-1-k}) := \omega_t + \lambda(\mathcal{B}) g(Z_{t-1}), \qquad \text{for all } t \in \mathbb{Z}, \tag{12}$$

where $\{\omega_t\}_{t\in\mathbb{Z}}$ and $\{\lambda_k\}_{k\in\mathbb{N}}$ are real, nonstochastic, scalar sequences for which the process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is well defined, $\{Z_t\}_{t\in\mathbb{Z}}$ is a white noise process with variance not necessarily equal to one and $g(\cdot)$ is any measurable function (notice that (9) is a particular case of this parametrization). In particular, Theorems 1 and 2 below, which are stated and proved in [4], assume that $\ln(\sigma_t^2)$ is given by (12) (the notation was adapted to reflect the one used in this work), with $\{Z_t\}_{t\in\mathbb{Z}}$ and $g(\cdot)$ as in (1) and (2), respectively. However, although (12) is more general than (5), in practice the applicability of the model is somewhat limited given that the parameter estimation is far more complicated when compared to the model (5).

Theorem 1 below provides a criterion for the stationarity and the ergodicity of EGARCH (FIEGARCH) processes. As pointed out by [4], the stationarity and ergodicity criterion in Theorem 1 is exactly the same as for a general linear process with finite variance innovations. Obviously, different definitions of $\lambda(\cdot)$ in (12) will lead to different conditions for the criterion in Theorem 1 to hold. In [4] it is stated that, in many applications, an ARMA process provides a parsimonious parametrization for $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$. In this case, $\lambda(\cdot)$ is defined as $\lambda(z) = \alpha(z)[\beta(z)]^{-1}$, $|z| \leq 1$, where $\alpha(\cdot)$ and $\beta(\cdot)$ are the polynomials given in (3), leading to an EGARCH(p, q) process. For this model, the criterion in Theorem 1 holds whenever the roots of $\beta(\cdot)$ are outside of the closed disk $\{z : |z| \leq 1\}$. We shall later discuss the condition for the criterion in Theorem 1 to hold when $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is defined by (5), leading to a FIEGARCH(p, d, q) process.

Theorem 1 (Nelson, 1991). *Define $\{\sigma_t^2\}_{t\in\mathbb{Z}}$, $\{X_t\}_{t\in\mathbb{Z}}$ and $\{Z_t\}_{t\in\mathbb{Z}}$ by*

$$X_t = \sigma_t Z_t; \qquad Z_t \sim i.i.d., \quad \mathbb{E}(Z_t) = 0 \quad and \quad \mathrm{Var}(Z_t) = 1; \tag{13}$$

$$\ln(\sigma_t^2) = \omega_t + \sum_{k=1}^{\infty} \lambda_k g(Z_{t-k}), \quad \lambda_1 = 1; \qquad g(Z_t) = \theta Z_t + \gamma\left[|Z_t| - \mathbb{E}(|Z_t|)\right]; \tag{14}$$

where $\{\omega_t\}_{t\in\mathbb{Z}}$ and $\{\lambda_k\}_{k\in\mathbb{N}^*}$ are real, nonstochastic, scalar sequences, and assume that θ and γ do not both equal zero. Then $\{e^{-\omega_t}\sigma_t^2\}_{t\in\mathbb{Z}}$, $\{e^{-\omega_t/2}X_t\}_{t\in\mathbb{Z}}$ and $\{\ln(\sigma_t^2) - \omega_t\}_{t\in\mathbb{Z}}$ are strictly stationary and ergodic and $\{\ln(\sigma_t^2) - \omega_t\}_{t\in\mathbb{Z}}$ is covariance stationary if and only if $\sum_{k=1}^{\infty}\lambda_k^2 < \infty$. If $\sum_{k=1}^{\infty}\lambda_k^2 = \infty$, then $|\ln(\sigma_t^2) - \omega_t| = \infty$ almost surely. If $\sum_{k=1}^{\infty}\lambda_k^2 < \infty$, then for $k > 0$, $\mathrm{Cov}\big(Z_{t-k}, \ln(\sigma_t^2)\big) = \lambda_k[\theta + \gamma\mathbb{E}(Z_t|Z_t|)]$, and $\mathrm{Cov}\big(\ln(\sigma_t^2), \ln(\sigma_{t-k}^2)\big) = \mathrm{Var}\big(g(Z_t)\big)\sum_{j=1}^{\infty}\lambda_j\lambda_{j+k}$.

Theorem 2 shows the existence of the r-th moment for the random variables X_t and σ_t^2, defined by (13)-(14), when $\sum_{k=1}^{\infty}\lambda_k^2 < \infty$ and the distribution of Z_0 is the Generalized Error Distribution (GED). In the sequel we shall present the condition under which $\sum_{k=0}^{\infty}\lambda_{d,k}^2 < \infty$ (see Theorem 3), with $\{\lambda_{d,k}\}_{k\in\mathbb{Z}}$ defined by (8), so the result of Theorem 2 also applies to FIEGARCH(p,d,q) processes.

Theorem 2 (Nelson, 1991). *Define $\{\sigma_t^2, X_t\}_{t\in\mathbb{Z}}$ by (13)-(14), and assume that θ and γ do not both equal zero. Let $\{Z_t\}_{t\in\mathbb{Z}}$ be i.i.d. GED with mean zero, variance one, and tail-thickness parameter $\nu > 1$, and let $\sum_{k=1}^{\infty}\lambda_k^2 < \infty$. Then $\{e^{-\omega_t}\sigma_t^2\}_{t\in\mathbb{Z}}$ and $\{e^{-\omega_t/2}X_t\}_{t\in\mathbb{Z}}$ possess finite, time-invariant moments of arbitrary order. Further, if $0 < r < \infty$, conditioning information at time 0 drops out of the forecast r-th moments of $e^{-\omega_t}\sigma_t^2$ and $e^{-\omega_t/2}X_t$, as $t \to \infty$:*

$$\begin{aligned}
&\operatorname*{plim}_{t\to\infty}\big[\mathbb{E}(e^{-r\omega_t}\sigma_t^{2r}\big|Z_0, Z_{-1}, Z_{-2}, \cdots) - \mathbb{E}(e^{-r\omega_t}\sigma_t^{2r})\big] = 0 \quad and\\
&\operatorname*{plim}_{t\to\infty}\big[\mathbb{E}(e^{-r\omega_t/2}|X_t|^r\big|Z_0, Z_{-1}, Z_{-2}, \cdots) - \mathbb{E}(e^{-r\omega_t/2}|X_t|^r)\big] = 0,
\end{aligned}$$

where plim *denotes the limit in probability.*

Theorem 3 below gives the convergence order of the coefficients $\lambda_{d,k}$, defined by (8), as k goes to infinity. This theorem is important for two reasons. First, it provides an approximation for $\lambda_{d,k}$, as $k \to \infty$, and this result plays an important role when choosing the truncation point in the series representation for simulation purposes. Second, and most important, the asymptotic representation provided in this theorem plays the key role to establish the necessary condition for square summability of $\{\lambda_{d,k}\}_{k\in\mathbb{N}}$. More specifically, from Theorem 3 one concludes that $\{\lambda_{d,k}\}_{k\in\mathbb{N}} \in \ell^2$ if and only if $d < 0.5$ and $\{\lambda_{d,k}\}_{k\in\mathbb{N}} \in \ell^1$ whenever $d < 0$.

Theorem 3. *Let $\lambda(\cdot)$ be the polynomial defined by (8). Then, for all $k \in \mathbb{N}$, the coefficients $\lambda_{d,k}$ satisfy*

$$\lambda_{d,k} \sim \frac{1}{\Gamma(d)\,k^{1-d}}\frac{\alpha(1)}{\beta(1)}, \qquad as\ k \to \infty. \tag{15}$$

Consequently, $\lambda_{d,k} = O(k^{d-1})$, as k goes to infinity.

Proposition 2 presents a recurrence formula for calculating the coefficients $\lambda_{d,k}$, for all $k \in \mathbb{N}$. This formula is used to generate the FIEGARCH time series in the simulation study presented in Section 4.

Proposition 2. *Let $\lambda(\cdot)$ be the polynomial defined by (8). The coefficients $\lambda_{d,k}$, for all $k \in \mathbb{N}$, are given by*

$$\lambda_{d,0} = 1 \quad and \quad \lambda_{d,k} = -\alpha_k^* + \sum_{i=0}^{k-1}\lambda_i\sum_{j=0}^{k-i}\beta_j^*\delta_{d,k-i-j}, \quad for\ all\ \ k \geq 1, \tag{16}$$

where the coefficients $\delta_{d,k}$, for all $k \in \mathbb{N}$, are given in (6) and

$$\alpha_m^* := \begin{cases} \alpha_m, & if \quad 0 \leq m \leq p; \\ 0, & if \quad\ \ m > p \end{cases} \qquad and \quad \beta_m^* := \begin{cases} \beta_m, & if \quad 0 \leq m \leq q; \\ 0, & if \quad\ \ m > q. \end{cases} \tag{17}$$

The applicability of Theorem 1 to long-range dependence models was briefly mentioned (without going into details) in [4]. Corollary 1 below is a direct application of Theorem 3 and provides a simple condition for the criterion in Theorem 1 to hold when $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is defined by (5), which leads to a long-range dependence model whenever $d > 0$.

Corollary 1. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a* FIEGARCH(p,d,q) *process defined by* (1) *and* (5)*. If $d < 0.5$, $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is stationary (weakly and strictly), ergodic and the random variable $\ln(\sigma_t^2)$ is almost surely finite, for all $t \in \mathbb{Z}$. Moreover, $\{X_t\}_{t\in\mathbb{Z}}$ and $\{\sigma_t^2\}_{t\in\mathbb{Z}}$ are strictly stationary and ergodic processes.*

The square summability of $\{\lambda_{d,k}\}_{k\in\mathbb{N}}$ implies that the process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is stationary (weakly and strictly), ergodic and the random variable $\ln(\sigma_t^2)$ is almost surely finite, for all $t \in \mathbb{Z}$ (see Theorem 1). Now, since $\{g(Z_t)\}_{t\in\mathbb{Z}}$ is a white noise process (see Proposition 1), it follows immediately that $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is an ARFIMA(q,d,p) process (for details on ARFIMA processes see, for instance, [13] and [23]). This result is very useful, not only for forecasting purposes (see Section 3) but also, to conclude the following properties

P1: if $d < 0.5$, the autocorrelation function of the process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is such that

$$\rho_{\ln(\sigma_t^2)}(h) \sim ch^{2d-1}, \quad \text{as} \quad h \to \infty,$$

where $c \neq 0$, and the spectral density function of the process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is such that

$$f_{\ln(\sigma_t^2)}(\lambda) = \frac{\sigma_g^2}{2\pi}\frac{|\alpha(e^{-i\lambda})|^2}{|\beta(e^{-i\lambda})|^2}|1-e^{-i\lambda}|^{-2d} \sim \frac{\sigma_g^2}{2\pi}\left[\frac{\alpha(1)}{\beta(1)}\right]^2\lambda^{-2d}, \quad \text{as } \lambda \to 0,$$

where $\sigma_g^2 = \mathrm{Var}\big(g(Z_t)\big)$ is given in (11);

P2: if $d \in (-1, 0.5)$ and $\alpha(z) \neq 0$, for $|z| \leq 1$, the process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is invertible, that is,

$$\lim_{m\to\infty} \mathbb{E}\bigg(\bigg|\sum_{k=0}^{m}\tilde{\lambda}_{d,k}\big[\ln(\sigma_{t-k}^2) - \omega - g(Z_{t-1})\big]\bigg|^r\bigg) = 0, \quad \text{for all} \quad 0 < r \leq 2,$$

where $\displaystyle\sum_{k=0}^{\infty}\tilde{\lambda}_{d,k}z^k = \tilde{\lambda}(z) := \lambda^{-1}(z) = \frac{\beta(z)}{\alpha(z)}(1-z)^d, \quad |z| < 1.$

Remark 5. The proof of **P1** can be found in [23], theorem 13.2.2. Regarding **P2**, in the literature one usually finds that an ARFIMA(p,d,q) process is invertible for $|d| < 0.5$ (see, for instance, [23], theorem 13.2.2). However, [24] already proved that this range can be extended to $d \in (-1, 0.5)$, for an ARFIMA$(0,d,0)$ and, more recently, [25] show that this result actually holds for any ARFIMA(p,d,q).

Corollary 1 shows that $\{X_t\}_{t\in\mathbb{Z}}$ and $\{\sigma_t^2\}_{t\in\mathbb{Z}}$ are strictly statiqonary and ergodic processes. However, as mentioned in [4], this does not imply weak stationarity when the random variable Z_t, for $t \in \mathbb{Z}$, is such that either its mean or its variance is not finite. Theorem 2 considers the GED function and proves the existence of the moment of order $r > 0$, for the random variables X_t and σ_t^2, for all $t \in \mathbb{Z}$, when the process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is defined in terms of a square summable sequence of coefficients. Corollary 2 below applies the result of Theorem 3 to state a simple condition so that Theorem 2 holds for FIEGARCH(p,d,q) processes.

Corollary 2. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a* FIEGARCH(p,d,q) *process defined by* (1) *and* (5)*. Assume that θ and γ are not both equal to zero and that $\{Z_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d.* GED(ν) *random variables, with $\nu > 1$, zero mean and variance equal to one. If $d < 0.5$, then $\mathbb{E}(X_t^r) < \infty$ and $\mathbb{E}(\sigma_t^{2r}) < \infty$, for all $t \in \mathbb{Z}$ and $r > 0$.*

As a consequence of Corollary 2, if $d < 0.5$ and $\{Z_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. GED(ν) random variables, with $v > 1$, zero mean and variance equal to one, then $\mathbb{E}(X_t^4) < \infty$ (consequently, $\mathbb{E}(X_t^3) < \infty$) and both, the asymmetry (A_X) and kurtosis (K_X) measures of $\{X_t\}_{t\in\mathbb{Z}}$ exist. Now, since $\mathbb{E}(X_t^r) = \mathbb{E}(\sigma_t^r)\mathbb{E}(Z_t^r)$, for all $r > 0$ (it follows from the independence of σ_t and Z_t), $\mathbb{E}(X_t) = 0$ and $\mathbb{E}(Z_t^2) = 1$, the measures A_X and K_X can be rewritten as

$$A_X := \frac{\mathbb{E}(X_t^3)}{\big[\mathbb{E}(X_t^2)\big]^{3/2}} = \frac{\mathbb{E}(\sigma_t^3)\mathbb{E}(Z_t^3)}{\big[\mathbb{E}(\sigma_t^2)\big]^{3/2}} \quad \text{and} \quad K_X := \frac{\mathbb{E}(X_t^4)}{\big[\mathbb{E}(X_t^2)\big]^{2}} = \frac{\mathbb{E}(\sigma_t^4)\mathbb{E}(Z_t^4)}{\big[\mathbb{E}(\sigma_t^2)\big]^{2}}, \quad \text{for all } t \in \mathbb{Z}. \tag{18}$$

An expression for K_X (as a function of the FIEGARCH model parameters) was already given in [8] by assuming that $\{Z_t\}_{t\in\mathbb{Z}}$ is a Gaussian white noise process with variance equal to one, $d > 0$, $p = 0$, $q = 1$ and by defining $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ through expression (10). In accordance with that definition it can be shown that K_X can be written as

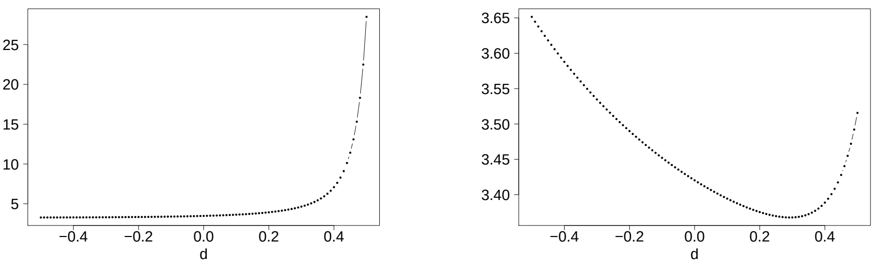
$$K_X = 3\,\frac{\prod_{j=1}^{\infty}\mathbb{E}\big(\exp\{2\lambda_j g(Z_{t-j})\}\big)}{\Big[\prod_{j=1}^{\infty}\mathbb{E}\big(\exp\{\lambda_j g(Z_{t-j})\}\big)\Big]^2} \quad \text{with} \quad \begin{cases} g(Z_t) = \theta Z_t + \gamma[|Z_t| - \sqrt{2/\pi}], & t \in \mathbb{Z} \\ \lambda_j = \displaystyle\sum_{i=0}^{j-1} \frac{\Gamma(i+d)}{\Gamma(i+1)\Gamma(d)}\beta^{j-i-1}, & j \in \mathbb{N}^* \text{ and } d > 0. \end{cases}$$

In Proposition 3 below we consider stationary FIEGARCH(p, d, q) processes (therefore, $d < 0.5$) with $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ defined by (5) and show that a similar expression holds for any $p, q \geq 0$. We do not impose that $\{Z_t\}_{t\in\mathbb{Z}}$ is a Gaussian white noise process since Corollary 1 shows that the asymmetry and kurtosis measures exist for a larger class of FIEGARCH models.

Proposition 3. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a stationary* FIEGARCH(p, d, q) *process defined by* (1) *and* (5). *Thus, if $\mathbb{E}(X_0^4) < \infty$, the asymmetry and kurtosis measures of $\{X_t\}_{t\in\mathbb{Z}}$ are given, respectively, by*

$$A_X = \mathbb{E}(Z_0^3)\,\frac{\prod_{k=0}^{\infty}\mathbb{E}\Big(\exp\Big\{\tfrac{3}{2}\lambda_{d,k} g(Z_0)\Big\}\Big)}{\big[\prod_{k=0}^{\infty}\mathbb{E}\big(\exp\{\lambda_{d,k} g(Z_0)\}\big)\big]^{3/2}} \quad \textit{and} \quad K_X = \mathbb{E}(Z_0^4)\,\frac{\prod_{k=0}^{\infty}\mathbb{E}(\exp\{2\lambda_{d,k} g(Z_0)\})}{\big[\prod_{k=0}^{\infty}\mathbb{E}\big(\exp\{\lambda_{d,k} g(Z_0)\}\big)\big]^{2}},$$

where $\lambda_{d,k}$ are given in (8) *and $g(\cdot)$ is defined by* (2).



(a) $\beta_1 = 0.6860$ (b) $\beta_1 = -0.6860$

Figure 2: (a) Kurtosis measure of a FIEGARCH$(0, d, 1)$ process as a function of the parameter d with $\theta = -0.1661$, $\gamma = 0.2792$, $\omega = -7.2247$ and $\beta_1 = 0.6860$; (b) Kurtosis measure of a FIEGARCH$(0, d, 1)$ process with the same parameters as in (a) but with β_1 replaced by $\beta_1 = -0.6860$.

Example 2. Figure 2 shows the theoretical value of the kurtosis measure, as a function of the parameter d, for a FIEGARCH$(0, d, 1)$ process, with Gaussian noise and parameters $\theta = -0.1661$, $\gamma = 0.2792$, $\omega = -7.2247$ and (a) $\beta_1 = 0.6860$ (b) $\beta_1 = -0.6860$. The parameter values considered in Figure 2 (a) are the same ones (except for d) as considered in Figure 1 (a). For the specific model considered in Figure 1, $d = 0.3578$ and the theoretical value of the kurtosis measure is 5.6733. The sample kurtosis value of the simulated time series presented in Figure 1 (a) is 5.3197, which is very close to the theoretical one. It is easy to see that, while in Figure 2 (a) the kurtosis values increase exponentially as d increases, in Figure 2 (b) the kurtosis values decrease for $-0.5 \leq d \leq d_c$ and increase for $d_c < d \leq 0.5$, for some critical value $d_c \in [0.29, 0.3]$. To generate Figure 2 (b) an equally spaced grid of d values, with increment $\Delta d = 0.01$

was considered. In this specific case the critical value is $d_c = 0.29$. When $\Delta d = 0.001$ and 0.0001 (the corresponding graph can be obtained from the authors upon request), the corresponding critical values are both equal to 0.293.

Although $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is an ARFIMA process, in practice it cannot be directly observed and frequently, knowing its characteristics may not be sufficient for model identification and estimation purposes. On the other hand, $\{X_t\}_{t\in\mathbb{Z}}$ is an observable process and so is $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$. By noticing that, from (1), one can rewrite

$$\ln(X_t^2) = \ln(\sigma_t^2) + \ln(Z_t^2), \quad \text{for all} \quad t \in \mathbb{Z},$$

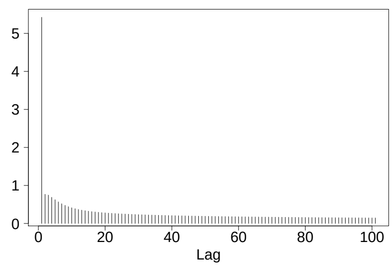
and now it is clear that the properties of $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ are useful to characterize the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$. This approach was already considered in the literature for parameter estimation purposes. For instance, [26] and [27] consider models such that X_t can be written as in (1), but σ_t can have a more general definition than (5). While [26] considers maximum likelihood and Whittle's method of estimation in the class of exponential volatility models, especially the EGARCH ones, [27] considers different semiparametric estimators of the memory parameter in general signal-plus-noise models. In both cases, to obtain an estimator by Whittle's method, the authors consider the spectral density function of $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$.

In what follows, we focus our attention to the case where X_t can be written as in (1), and σ_t is defined through the expression (5) and we present some properties of the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$. In particular, we show that, under mild conditions, this process also has an ARFIMA representation. To the best of our knowledge no formal proofs of these results are given in the literature of FIEGARCH(p,d,q) processes, especially the ARFIMA$(q,d,0)$ representation of $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$.

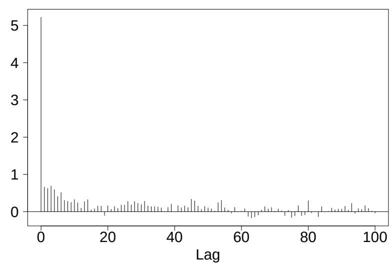
Theorem 4. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a* FIEGARCH(p,d,q) *process defined by* (1) *and* (5). *If* $\mathbb{E}([\ln(Z_0^2)]^2) < \infty$ *and* $d < 0.5$, *then the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is well defined and it is stationary (weakly and strictly) and ergodic. Moreover, the autocovariance function of $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is given by*

$$\gamma_{\ln(X^2)}(h) = \sigma_g^2 \sum_{k=0}^{\infty} \lambda_{d,k}\lambda_{d,k+|h|} + \mathrm{Var}\big(\ln(Z_t^2)\big)\mathbb{I}_{\{0\}}(h) + \lambda_{d,|h|-1}\mathcal{K}\mathbb{I}_{\mathbb{Z}^*}(h), \quad \textit{for all } h \in \mathbb{Z}, \tag{19}$$

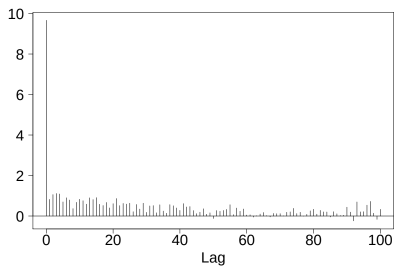
where σ_g^2 is given in (11) *and* $\mathcal{K} = \mathrm{Cov}\big(g(Z_0), \ln(Z_0^2)\big)$.



(a) $\gamma_{\ln(X^2)}(\cdot)$



(b) $\hat{\gamma}_{\ln(X^2)}(\cdot)$



(c) $\hat{\gamma}_{\ln(r^2)}(\cdot)$

Figure 3: (a) Theoretical autocovariance function of the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, where $\{X_t\}_{t\in\mathbb{Z}}$ is a FIEGARCH$(0,d,1)$ process and (b) sample autocovariance function of a time series $\{\ln(x_t^2)\}_{t=1}^{2000}$ derived from that FIEGARCH$(0,d,1)$ process; (c) sample autocovariance function of the time series $\{\ln(r_t^2)\}_{t=1}^{1717}$, where $\{r_t\}_{t=1}^{1717}$ is the Bovespa index log-returns time series.

Example 3. Figure 3 (a) presents the theoretical autocovariance function of the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, where $\{X_t\}_{t\in\mathbb{Z}}$ is a FIEGARCH$(0,d,1)$ process, with the same parameter values considered in Figures 1 and 4. Figure 3 (b) shows the sample autocovariance function of the time series $\{\ln(x_t^2)\}_{t=1}^{2000}$, where $\{x_t\}_{t=1}^{2000}$ is the simulated time series presented in Figure 1. Figure 3 (c) presents the sample autocovariance function of the time series $\{\ln(r_t^2)\}_{t=1}^{1717}$, where $\{r_t\}_{t=1}^{1717}$ is the Bovespa index log-returns time series (see

Section 5). By comparing the three graphs in Figure 3, one concludes that all three functions present a similar behavior in terms of function decay. Since the sample autocovariance function is an estimator of the theoretical autocovariance function, it is expected that their graphs will have the same behavior. The similarity between the decay in the graphs in Figures 3 (b) and (c) indicates that a FIEGARCH model could be an appropriate candidate for fitting the Bovespa index log-returns time series.

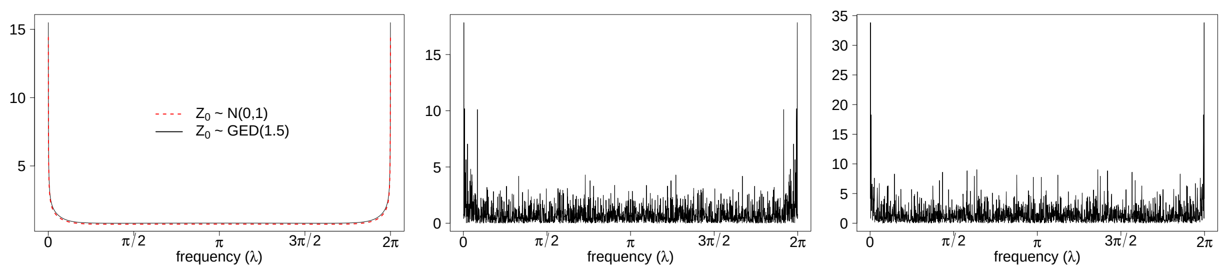
Example 4. Theorem 4 provides the expression for the autocorrelation function of $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$. The spectral density function of the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is given by (see [27])

$$\begin{aligned} f_{\ln(X_t^2)}(\lambda) &= f_{\ln(\sigma_t^2)}(\lambda) + \frac{\mathcal{K}}{\pi}\mathbb{R}e\big(e^{-\mathrm{i}\lambda}\Lambda(\lambda)\big) + f_{\ln(Z_t^2)}(\lambda) \\ &= \frac{\sigma_g^2}{2\pi}\frac{|\alpha(e^{-i\lambda})|^2}{|\beta(e^{-i\lambda})|^2}|1-e^{-i\lambda}|^{-2d} + \frac{\mathcal{K}}{\pi}\mathbb{R}e\big(e^{-\mathrm{i}\lambda}\Lambda(\lambda)\big) + \frac{1}{2\pi}\mathrm{Var}(\ln(Z_0^2)), \quad \text{for all } \lambda\in[0,\pi], \end{aligned} \tag{20}$$

where $\sigma_g^2 = \mathrm{Var}\big(g(Z_t)\big)$ is given in (11), $\mathcal{K} = \mathrm{Cov}\big(g(Z_0), \ln(Z_0^2)\big)$, $\mathbb{R}e(z)$ is the real part of z and $\Lambda(z) := \lambda(e^{-\mathrm{i}z})$. As an illustration, Figure 4 (a) shows the spectral density function of the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, where $\{X_t\}_{t\in\mathbb{Z}}$ is a FIEGARCH$(0,d,1)$ process with $d = 0.3578$, $\theta = -0.1661$, $\gamma = 0.2792$, $\omega = -7.2247$ and $\beta_1 = 0.6860$, assuming $Z_0 \sim \mathcal{N}(0,1)$ (dashed line) and $Z_0 \sim$ GED(1.5) (continuous line). The corresponding values of σ_g^2, $\mathcal{K}$ and Var(ln(Z_0^2)), used in the calculation of $f_{\ln(X_t^2)}(\cdot)$, are given in Table 1. Figure 4 (b) shows the periodogram function of the time series $\{\ln(x_t^2)\}_{t=1}^{2000}$, where $\{x_t\}_{t=1}^{2000}$ is the simulated time series presented in Figure 1 (a), with $Z_0 \sim \mathcal{N}(0,1)$. Figure 4 (c) shows the periodogram function of the time series $\{\ln(r_t^2)\}_{t=1}^{1717}$, where $\{r_t\}_{t=1}^{1717}$ is the Bovespa index log-returns time series.

Table 1: Theoretical values for the expectation and variance of functions of Z_0 and the corresponding values of σ_g^2 and $\mathcal{K}$ considering the Gaussian and the Generalized Error distribution functions. In both cases $\theta = -0.1661$ and $\gamma = 0.2792$.

Distribution	$\mathbb{E}(\lvert Z_0\rvert)$	$\mathbb{E}(\lvert Z_0\rvert\ln(Z_0^2))$	$\mathbb{E}(\ln(Z_0^2))$	$\mathrm{Var}(\ln(Z_0^2))$	σ_g^2	$\mathcal{K}$
$\mathcal{N}(0,1)$	0.7979	0.0925	-1.2704	4.9348	0.0559	0.3088
GED(1.5)	0.7674	0.0975	-1.4545	5.4469	0.0596	0.3389



(a) $f_{\ln(X_t^2)}(\cdot)$ (b) $I_{\ln(X_t^2)}(\cdot)$ (c) $I_{\ln(r_t^2)}(\cdot)$

Figure 4: (a) Theoretical spectral density function of the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, where $\{X_t\}_{t\in\mathbb{Z}}$ is a FIEGARCH$(0,d,1)$ process with $d = 0.3578$, $\theta = -0.1661$, $\gamma = 0.2792$, $\omega = -7.2247$ and $\beta_1 = 0.6860$, assuming $Z_0 \sim \mathcal{N}(0,1)$ (dashed line) and $Z_0 \sim$ GED(1.5) (continuous line); (b) the periodogram function related to a time series $\{x_t\}_{t=1}^{2000}$ derived from this FIEGARCH$(0,d,1)$ process with $Z_0 \sim \mathcal{N}(0,1)$; (c) the periodogram function related to the time series $\{\ln(r_t^2)\}_{t=1}^{1717}$, where $\{r_t\}_{t=1}^{1717}$ is the Bovespa index log-returns time series.

Figure 4 (a) shows that the graphs for the spectral density functions $f_{\ln(X_t^2)}(\cdot)$ of FIEGARCH$(0,d,1)$ processes with $Z_0 \sim \mathcal{N}(0,1)$ (dashed line) and $Z_0 \sim$ GED(1.5) (continuous line) are very close to each other. This fact indicates that the probability distribution of Z_0 may not be evident from the periodogram function (which is an estimate of $f_{\ln(X_t^2)}(\cdot)$). The small difference on the values of $f_{\ln(X_t^2)}(\cdot)$ for $Z_0 \sim \mathcal{N}(0,1)$ and $Z_0 \sim$ GED(1.5) is explained by the fact that the values of σ_g^2, $\mathcal{K}$ and $f_{\ln(Z_t^2)}(\lambda)$ are relatively close for $Z_0 \sim \mathcal{N}(0,1)$ and $Z_0 \sim \text{GED}(1.5)$ as shown in Table 1. The graphs in Figures 4 (b) and (c) present similar behavior, indicating that a FIEGARCH model could be adequate to fit the data. On the other hand, as mentioned before, given the similarity of the graphs in 4 (a), there is not enough evidence to conclude that the underlying probability distributions of $\{r_t\}_{t=1}^{1717}$(the observed time series) and $\{x_t\}_{t=1}^{2000}$ (the simulated time series) are the same. In fact, we apply the two-sample Kolmogorov-Smirnov test to verify the hypothesis that $I_{\ln(X_t^2)}(\cdot)$ and $I_{\ln(r_t^2)}(\cdot)$ have the same probability distribution. We also apply the test to the standardized versions of $I_{\ln(X_t^2)}(\cdot)$ and $I_{\ln(r_t^2)}(\cdot)$ (that is, we subtracted the sample mean and divided by the sample standard deviation). In the first case the test rejects the null hypothesis ($\alpha = 0.05$, test statistic $= 0.2208$, p-value $< 2.2 \times 10^{-16}$). In the second case (standardized version) the test did not reject the null hypothesis ($\alpha = 0.05$, test statistic $= 0.0285$, p-value $= 0.8472$).

To further investigate whether the correct probability distribution of Z_0 can be identified through the periodogram function of $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ when $\{X_t\}_{t\in\mathbb{Z}}$ is a FIEGARCH process, we consider the same time series $\{x_t\}_{t=1}^{2000}$ as in Figure 4 (b) (that is, the sample from a FIEGARCH$(0,d,1)$ process, with Gaussian innovations) and perform the Kolmogorov-Smirnov hypothesis test proposed in [23] (pages 339 - 342) and described in the following bullet points. The target spectral density function $f_{\ln(X_t^2)}(\cdot)$, given in (20), is obtained by assuming both that $Z_0 \sim \mathcal{N}(0,1)$, which corresponds to the correct choice since $\{x_t\}_{t=1}^{2000}$ was generated assuming that the innovation process has Gaussian distribution, and $Z_0 \sim \text{GED}(1.5)$. The test is performed as follows

- the null hypothesis of the test is that $\ln(X_t^2)$ has spectral density function $f_{\ln(X_t^2)}(\cdot)$;
- the testing procedure consists of plotting the Kolmogorov-Smirnov boundaries

$$y = \frac{x-1}{m-1} \pm k_\alpha (m-1)^{-1/2}, \quad 1 \le x \le m, \quad k_\alpha = \begin{cases} 1.36, & \text{if} \quad \alpha = 0.05; \\ 1.63, & \text{if} \quad \alpha = 0.01; \end{cases}$$

  and the function $C(x)$ defined as

$$C(x) = \begin{cases} 0, & \text{if} \quad x < 1; \\ Y_i, & \text{if} \quad i \le x < i+1, \quad \text{for } i \in \{1, \cdots, m\}; \\ 1, & \text{if} \quad x \ge m; \end{cases}$$

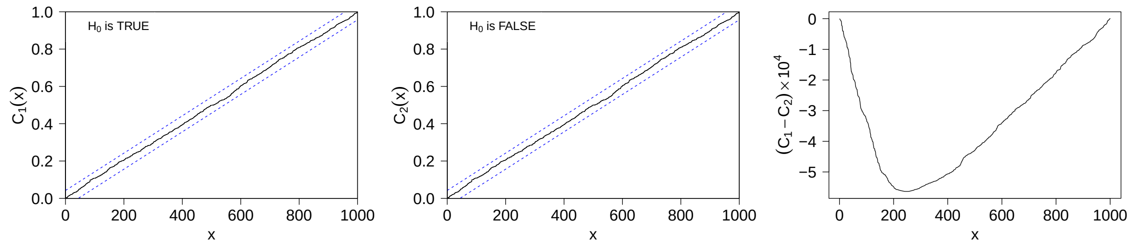
  with $Y_0 := 0$, $Y_m := 1$ and

$$Y_i := \left[\sum_{k=1}^{i} \frac{I_{\ln(X_t^2)}(\omega_k)}{f_{\ln(X_t^2)}(\omega_k)}\right]\left[\sum_{k=1}^{m} \frac{I_{\ln(X_t^2)}(\omega_k)}{f_{\ln(X_t^2)}(\omega_k)}\right]^{-1}, \qquad \text{with} \quad \omega_k = \frac{2k\pi}{n}, \ k \in \{1, \cdots, m\},$$

  where m is the integer part of $(n-1)/2$ and n is the time series sample size;
- the null hypothesis is rejected if $C(\cdot)$ exits the boundaries for some $1 \le x \le m$.

The results of the tests are given in Figure 5, where $C_1(\cdot)$ and $C_2(\cdot)$ denote the values of $C(\cdot)$ obtained, respectively, when assuming $Z_0 \sim \mathcal{N}(0,1)$ and $Z_0 \sim \text{GED}(1.5)$. From Figures 5 (a) and (b) one concludes that the Kolmogorov-Smirnov test does not reject the null hypothesis in both cases. This result was expected given the small difference between the values of $f_{\ln(X_t^2)}(\cdot)$, shown in Figure 4 (a). In fact, by comparing Figures 5 (a) and (b), one observes no visible difference between those graphs. Figure 5 (c) confirms that the difference is too small to be noticed since $|C_1(x) - C_2(x)| < 6 \times 10^{-4}$, for all $0 \le x \le 1000$. This shows that, for the FIEGARCH process considered in Example 4, the correct probability distribution of Z_0 cannot be identified through the periodogram function, given that the Kolmogorov-Smirnov hypothesis test failed to reject the null hypothesis when it was false.

To conclude this section we present the following theorem which shows that, under mild conditions, $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is an ARFIMA$(q,d,0)$ process with correlated innovations. This result is very useful in model identification and parameter estimation since the literature of ARFIMA models is well developed (see [13] and references therein).



(a) $C_1(x)$ (b) $C_2(x)$ (c) $C_1(x) - C_2(x)$

Figure 5: Function $C(x)$ and the Kolmogorov-Smirnov boundaries, with $\alpha = 0.05$ (dashed line), when $\{x_t\}_{t=1}^{2000}$ is a time series derived from a FIEGARCH$(0, d, 1)$ process with $Z_0 \sim \mathcal{N}(0, 1)$ and $f_{\ln(X_t^2)}(\cdot)$ is the theoretical spectral density function of a FIEGARCH$(0, d, 1)$ process with (a) $Z_0 \sim \mathcal{N}(0, 1)$ (therefore, the null hypothesis is true); (b) $Z_0 \sim \mathrm{GED}(1.5)$ (therefore, the null hypothesis is false). In all cases, $d = 0.3578$, $\theta = -0.1661$, $\gamma = 0.2792$, $\omega = -7.2247$ and $\beta_1 = 0.6860$. Panel (c), the difference between the values of $C(x)$ (multiplied by 10^4) assuming, respectively, $Z_0 \sim \mathcal{N}(0, 1)$ and $Z_0 \sim \mathrm{GED}(1.5)$.

Theorem 5. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a* FIEGARCH(p, d, q) *process defined by* (1) *and* (5). *Suppose $|d| < 0.5$ and $\mathbb{E}([\ln(Z_0^2)]^2) < \infty$. Then $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is an* ARFIMA$(q, d, 0)$ *process, with correlated innovations, given by*

$$\beta(\mathcal{B})(1 - \mathcal{B})^d(\ln(X_t^2) - \omega) = \varepsilon_t, \quad \textit{for all } t \in \mathbb{Z},$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a stochastic process with zero mean and autocovariance function $\gamma_\varepsilon(\cdot)$ given by

$$\gamma_\varepsilon(h) = \begin{cases} \sigma_g^2 \displaystyle\sum_{i=|h|}^{p} \alpha_i\alpha_{i-|h|} + \mathcal{K}\sum_{i=0}^{p}\alpha_i\phi_{i+|h|+1} + \mathcal{K}\sum_{i=|h|-1}^{p}\alpha_i\phi_{i-|h|+1} + \sigma_\ell^2\sum_{i=|h|}^{\infty}\phi_i\phi_{i-|h|}, & \textit{if} \quad 0 \le |h| \le p; \\ \mathcal{K}\alpha_p\phi_1 + \sigma_\ell^2 \displaystyle\sum_{i=p+1}^{\infty}\phi_i\phi_{i-(p+1)}, & \textit{if} \quad |h| = p+1; \\ \sigma_\ell^2 \displaystyle\sum_{i=|h|}^{\infty}\phi_i\phi_{i-|h|}, & \textit{if} \quad |h| > p+1, \end{cases} \tag{21}$$

with $\{\phi_k\}_{k\in\mathbb{N}}$ defined by

$$\phi(z) := \beta(z)(1 - z)^d = \sum_{k=0}^{\infty}\phi_k z^k, \quad \textit{for } |z| < 1, \tag{22}$$

σ_g^2 given in (11), *$\mathcal{K} = \mathrm{Cov}(g(Z_0), \ln(Z_0^2))$, $\{\alpha_i\}_{i=0}^p$ given in* (3) *and $\sigma_\ell^2 := \mathrm{Var}(\ln(Z_0^2))$.*

3. Forecasting

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a FIEGARCH(p, d, q) process defined by (1) and (5), and $\{x_t\}_{t=1}^n$ a time series obtained from this process. In this section, we prove that $\{X_t\}_{t\in\mathbb{Z}}$ is a martingale difference with respect to the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$, where $\mathcal{F}_t := \sigma(\{Z_s\}_{s\le t})$, and we provide the h-step ahead forecast for the process $\{X_t\}_{t\in\mathbb{Z}}$. Since the process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$, defined by (5), has an ARFIMA(q, d, p) representation, the h-step ahead forecasting for this process and its mean square error value can be easily obtained (for instance, see [13] and [29]). This fact is used to provide an h-step ahead forecast for $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ and the mean square error of forecasting. We also consider the fact that $\mathbb{E}(X_t^2) = \mathbb{E}(\sigma_t^2)$, for all $t \in \mathbb{Z}$, to provide an h-step ahead forecast for both processes, $\{X_t^2\}_{t\in\mathbb{Z}}$ and $\{\sigma_t^2\}_{t\in\mathbb{Z}}$, based on the predictions obtained from the process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$. The notation used in this section is introduced below.

Remark 6. Let Y_t, for $t \in \mathbb{Z}$, denote any random variable defined here. In the sequel we consider the following notation:

- we use the symbol "ˆ" to denote the h-step ahead forecast defined in terms of the conditional expectation, that is, $\hat{Y}_{t+h} = \mathbb{E}(Y_{t+h}|\mathcal{F}_t)$. Notice that this is the best linear predictor in terms of mean square error value. The symbols "˜" and "ˇ" are used to denote alternative estimators (e.g. $\tilde{Y}_{t+h}$ and $\check{Y}_{t+h}$);
- for simplicity of notation, for the h-step ahead forecast of $\ln(Y_{t+h})$, we write $\hat{\ln}(Y_{t+h})$ instead of $\widehat{\ln(Y_{t+h})}$ (analogously for "˜" and "ˇ");
- we follow the approach usually considered in the literature and denote the h-step ahead forecast Y^2_{t+h} as $\hat{Y}^2_{t+h}$ instead of $\widehat{Y^2_{t+h}}$. If necessary, to avoid confusion, we will denote the square of $\hat{Y}_{t+h}$ as $(\hat{Y}_{t+h})^2$ (analogously for "˜" and "ˇ").

The following lemma shows that a FIEGARCH(p, d, q) process is a martingale difference with respect to $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$. This result is useful in the proof of Lemma 2 that presents the h-step ahead forecast of X_{n+h}, for a fixed value of $n \in \mathbb{Z}$ and all $h \geq 1$, and the 1-step ahead forecast of X^2_{n+1}, given $\mathcal{F}_n$.

Lemma 1. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a* FIEGARCH(p, d, q) *process defined by* (1) *and* (5) *and $\mathcal{F}_t := \sigma(\{Z_s\}_{s\leq t})$. Then the process $\{X_t\}_{t\in\mathbb{Z}}$ is a martingale difference with respect to $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$.*

Lemma 2. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a stationary* FIEGARCH(p, d, q) *process defined by* (1) *and* (5)*. Then, for any fixed $n \in \mathbb{Z}$, the h-step ahead forecast of X_{n+h}, for all $h > 0$ and the 1-step ahead forecast of X^2_{n+1}, given $\mathcal{F}_n$, are, respectively, $\hat{X}_{n+h} = 0$ and $\hat{X}^2_{n+1} = \sigma^2_{n+1}$.*

To obtain the h-step ahead forecast for X^2_{n+h}, notice that σ_t and Z_t are independent and so are σ^2_t and Z^2_t, for all $t \in \mathbb{Z}$. Moreover, $\mathbb{E}(Z^2_{n+h}|\mathcal{F}_n) = \mathbb{E}(Z^2_{n+h}) = 1$, for all $h > 0$. It follows that

$$\hat{X}^2_{n+h} := \mathbb{E}(X^2_{n+h}|\mathcal{F}_n) = \mathbb{E}(\sigma^2_{n+h}|\mathcal{F}_n) := \hat{\sigma}^2_{n+h}, \quad \text{for all} \quad h > 0.$$

While for ARCH/GARCH models, $\mathbb{E}(\sigma^2_{n+h}|\mathcal{F}_t)$ can be easily calculated, for FIEGARCH processes, only the expression for the h-step ahead forecast for the process $\{\ln(\sigma^2_t)\}_{t\in\mathbb{Z}}$, for any $h > 1$, is easy to derive. The expressions for $\hat{\ln}(\sigma^2_{n+h}) := \mathbb{E}(\ln(\sigma^2_{n+h})|\mathcal{F}_t)$ and for the mean square error of forecast are given in Proposition 4. We shall use this result to discuss the properties of the predictor obtained by considering $\check{\sigma}^2_{n+h} := \exp\{\hat{\ln}(\sigma^2_{n+h})\}$, for all $h > 0$.

Proposition 4. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a* FIEGARCH(p, d, q) *process defined by* (1) *and* (5)*. Then the h-step ahead forecast $\hat{\ln}(\sigma^2_{n+h})$ of $\ln(\sigma^2_{n+h})$, given $\mathcal{F}_n = \sigma(\{Z_t\}_{t\leq n})$, $n \in \mathbb{N}$, is given by*

$$\hat{\ln}(\sigma^2_{n+h}) = \omega + \sum_{k=0}^{\infty} \lambda_{d,k+h-1}\, g(Z_{n-k}), \quad \textit{for all } h > 0. \tag{23}$$

Moreover, the mean square error of forecast is equal to zero, if $h = 1$, and it is given by

$$\mathbb{E}\big(\big[\ln(\sigma^2_{n+h}) - \hat{\ln}(\sigma^2_{n+h})\big]^2\big) = \sigma^2_g \sum_{k=0}^{h-2} \lambda^2_{d,k}, \quad \textit{if } h \geq 2, \tag{24}$$

where $\sigma^2_g := \mathbb{E}([g(Z_0)]^2)$ is given in (11)*.*

In practice, $\mathbb{E}(\sigma^2_{n+h}|\mathcal{F}_t)$ cannot be easily calculated for FIEGARCH models and thus, a common approach is to predict σ^2_{n+h} through the relation $\check{\sigma}^2_{n+h} := \exp\{\hat{\ln}(\sigma^2_{n+h})\}$, with $\hat{\ln}(\sigma^2_{n+h})$ defined by (23), for all $h > 0$. As a consequence, a h-step ahead forecast for X^2_{n+h} is defined as $\check{X}^2_{n+h} := \check{\sigma}^2_{n+h}$ and a naive estimator for $\ln(X^2_{n+h})$ is obtained by letting

$$\check{\ln}(X^2_{n+h}) := \ln(\check{X}^2_{n+h}) = \ln(\check{\sigma}^2_{n+h}) = \hat{\ln}(\sigma^2_{n+h}), \quad \text{for all} \quad h > 0. \tag{25}$$

From expressions (1) and (25), it is obvious that $\check{\ln}(X^2_{n+h})$ is a biased estimator for $\ln(X^2_{n+h})$, whenever $\mathbb{E}(\ln(Z^2_{n+h})) \neq 0$. Proposition 5 gives the mean square error of forecast for the h-step ahead forecast of $\ln(X^2_{n+h})$, defined through expression (25).

Proposition 5. *Let $\check{\ln}(X_{n+h}^2)$, for all $h>0$, be the h-step ahead forecast of $\ln(X_{n+h}^2)$, given the filtration $\mathcal{F}_n = \sigma(\{Z_s\}_{s\leq n})$, defined by expression (25). Then the mean square error of forecast is given by*

$$\mathbb{E}\big(\big[\ln(X_{n+h}^2) - \check{\ln}(X_{n+h}^2)\big]^2\big) = \sigma_g^2 \sum_{k=0}^{h-2} \lambda_{d,k}^2 + \mathbb{E}\big(\big[\ln(Z_{n+h}^2)\big]^2\big), \quad \textit{where } \sigma_g^2 := \mathbb{E}\big([g(Z_0)]^2\big).$$

Remark 7. If the values of X_t and σ_t are known only for $t \in \{1, \cdots, n\}$, by setting $g(Z_{n+h-1-k}) = 0$ whenever $n+h-1-k < 1$ (that is, $k > n+h-2$) or $n+h-1-k > n$ (that is, $k < h-1$), the h-step ahead forecast $\hat{\ln}(\hat{\sigma}_{n+h}^2)$ of $\ln(\sigma_{n+h}^2)$, is approximated by

$$\hat{\ln}(\sigma_{n+h}^2) \simeq \omega + \sum_{k=0}^{n-1} \lambda_{d,k+h-1}\, g(Z_{n-k}) = \omega + \sum_{k=h-1}^{n+h-2} \lambda_{d,k}\, g(Z_{n+h-1-k}), \quad \text{for all } h>0,$$

and, by definition, the same approximation follows for $\check{\ln}(X_{n+h}^2)$. It is easy to see that, in this case, the error of forecast for the processes $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is given by

$$\begin{aligned}\ln(\sigma_{n+h}^2) - \hat{\ln}(\sigma_{n+h}^2) &= \left[\omega + \sum_{k=0}^{\infty} \lambda_{d,k}\, g(Z_{n+h-1-k})\right] - \left[\omega + \sum_{k=h-1}^{n+h-2} \lambda_{d,k}\, g(Z_{n+h-1-k})\right] \\ &= \sum_{k=0}^{h-2} \lambda_{d,k}\, g(Z_{n+h-1-k}) + \sum_{k=n+h-1}^{\infty} \lambda_{d,k}\, g(Z_{n+h-1-k})\end{aligned}$$

and the mean square error of forecast values for the processes $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ and $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ are given, respectively, by

$$\begin{aligned}&\mathbb{E}\big(\big[\ln(\sigma_{n+h}^2) - \hat{\ln}(\sigma_{n+h}^2)\big]^2\big) = \sigma_g^2\left(\sum_{k=0}^{h-2} \lambda_{d,k}^2 + \sum_{k=n+h-1}^{\infty} \lambda_{d,k}^2\right) \quad \text{and} \\ &\mathbb{E}\big(\big[\ln(X_{n+h}^2) - \check{\ln}(X_{n+h}^2)\big]^2\big) = \sigma_g^2\left(\sum_{k=0}^{h-2} \lambda_{d,k}^2 + \sum_{k=n+h-1}^{\infty} \lambda_{d,k}^2\right) + \mathbb{E}\big(\big[\ln(Z_{n+h}^2)\big]^2, \quad \text{for all } h>0.\end{aligned}$$

From Jensen's inequality, one concludes that

$$\check{\sigma}_{n+h}^2 := \exp\{\hat{\ln}(\sigma_{n+h}^2)\} = \exp\{\mathbb{E}(\ln(\sigma_{n+h}^2)|\mathcal{F}_t)\} \leq \mathbb{E}(\sigma_{n+h}^2|\mathcal{F}_n) := \hat{\sigma}_{n+h}^2, \quad \text{for all } h>0,$$

so that $\mathbb{E}(\check{\sigma}_{n+h}^2 - \sigma_{n+h}^2) = \mathbb{E}(\mathbb{E}(\check{\sigma}_{n+h}^2 - \sigma_{t+h}^2|\mathcal{F}_n)) = \mathbb{E}(\check{\sigma}_{n+h}^2 - \hat{\sigma}_{n+h}^2) \leq 0$, for all $h>0$. In fact, from (15) and (23), we have

$$\check{\sigma}_{n+h}^2 := \exp\{\hat{\ln}(\sigma_{n+h}^2)\} = \exp\left\{\omega + \sum_{k=0}^{\infty} \lambda_{d,k+h-1}\, g(Z_{n-k})\right\} \xrightarrow{h\to\infty} e^{\omega} = \exp\{\mathbb{E}(\ln(\sigma_0^2))\}. \tag{26}$$

Another h-step ahead predictor for σ_{n+h}^2 can be defined as follows. Consider an order 2 Taylor expansion of the exponential function and write

$$\begin{aligned}\sigma_{n+h}^2 &= \exp\big\{\mathbb{E}(\ln(\sigma_{n+h}^2)|\mathcal{F}_n)\big\} + \big[\ln(\sigma_{n+h}^2) - \mathbb{E}(\ln(\sigma_{n+h}^2)|\mathcal{F}_n)\big] \exp\big\{\mathbb{E}(\ln(\sigma_{n+h}^2)|\mathcal{F}_n)\big\} \\ &\quad + \frac{1}{2}\big[\ln(\sigma_{n+h}^2) - \mathbb{E}(\ln(\sigma_{n+h}^2)|\mathcal{F}_n)\big]^2 \exp\big\{\mathbb{E}(\ln(\sigma_{n+h}^2)|\mathcal{F}_n)\big\} + R_{n+h}, \quad \text{for all } h>0.\end{aligned} \tag{27}$$

From expression (27), a natural choice is to define a h-step ahead predictor for σ_{n+h}^2 as

$$\tilde{\sigma}_{n+h}^2 := \exp\big\{\mathbb{E}(\ln(\sigma_{n+h}^2)|\mathcal{F}_n)\big\} + \frac{1}{2}\mathbb{E}\big(\big[\ln(\sigma_{n+h}^2) - \mathbb{E}(\ln(\sigma_{n+h}^2)|\mathcal{F}_n)\big]^2\big) \exp\big\{\mathbb{E}(\ln(\sigma_{n+h}^2)|\mathcal{F}_n)\big\}, \tag{28}$$

for all $h > 0$.

From expressions (23), (24) and (28) one concludes that $\check{\sigma}^2_{n+h}$ and $\tilde{\sigma}^2_{n+h}$ are related through the equation

$$\tilde{\sigma}^2_{n+h} = \begin{cases} \exp\{\hat{\ln}(\sigma^2_{n+h})\} = \check{\sigma}^2_{n+h}, & \text{if} \quad h = 1; \\ \exp\{\hat{\ln}(\sigma^2_{n+h})\}\left(1 + \dfrac{1}{2}\sigma^2_g \displaystyle\sum_{k=0}^{h-2} \lambda^2_{d,k}\right) = \check{\sigma}^2_{n+h}\left(1 + \dfrac{1}{2}\sigma^2_g \displaystyle\sum_{k=0}^{h-2} \lambda^2_{d,k}\right), & \text{if} \quad h > 1. \end{cases} \tag{29}$$

Since σ_{t+1} is a $\mathcal{F}_t$-measurable random variable, for all $t \in \mathbb{Z}$, we have $\mathbb{E}(\tilde{\sigma}^2_{n+1} - \sigma^2_{n+1}) = \mathbb{E}(\check{\sigma}^2_{n+1} - \sigma^2_{n+1}) = 0$. From equation (27), we easily conclude that, for all $h > 1$,

$$\mathbb{E}(\tilde{\sigma}^2_{n+h} - \sigma^2_{n+h}) = -\mathbb{E}(R_{n+h}) \quad \text{and} \quad \mathbb{E}(\check{\sigma}^2_{n+h} - \sigma^2_{n+h}) = -\left(1 + \frac{1}{2}\sigma^2_g \sum_{k=0}^{h-2} \lambda^2_{d,k}\right)\mathbb{E}(\check{\sigma}^2_{n+h}) - \mathbb{E}(R_{n+h}).$$

Therefore, the relation between the bias for the estimators $\check{\sigma}^2_{n+h}$ and $\tilde{\sigma}^2_{n+h}$ is given by

$$\mathbb{E}(\tilde{\sigma}^2_{n+h} - \sigma^2_{n+h}) = \mathbb{E}(\check{\sigma}^2_{n+h} - \sigma^2_{n+h}) + \mathbb{E}(\check{\sigma}^2_{n+h})\left(1 + \frac{1}{2}\sigma^2_g \sum_{k=0}^{h-2} \lambda^2_{d,k}\right), \quad \text{for all } h > 1.$$

In Section 4 we analyze the performance of $\tilde{\sigma}^2_{n+h}$ through a Monte Carlo simulation study.

4. Simulation Study

In this section we present a Monte Carlo simulation study to analyze the performance of the quasi-likelihood estimator and also the forecasting of FIEGARCH(p, d, q) processes. Six different models are considered and, from now on, they shall be referred to as model Mi, for $i \in \{1, \cdots, 6\}$. For all models we assume that the distribution of Z_0 is the Generalized Error Distribution (GED) with tail-thickness parameter $\nu = 1.5$ (since $\nu < 2$ the tails are heavier than the Gaussian distribution tail). The set of parameters considered in this study is the same as in [11] and [12][5], except for models M5 and M6 (see Table 2). While model M5 considers $d = 0.49$, which is close to the non-stationary region ($d \geq 0.5$), model M6 considers $p = 1$ and $q = 0$. For comparison, we shall consider for model M6 the same parameter values as in model M3 (obviously, with the necessary adjustments regarding α_1 and β_1). We also present here the h-step ahead forecast, for $h \in \{1, \cdots, 50\}$, for the conditional variance of simulated FIEGARCH processes.

4.1. Data Generating Process (DGP)

To generate samples from FIEGARCH(p, d, q) processes we proceed as described in steps **DGP1 - DGP3** below. Notice that, while step 1 only needs to be repeated for each model, steps 2 and 3 must be repeated for each model and each replication. The parameters values considered in this simulation study are given in Table 2. For each model we consider $re = 1000$ replications, with sample size $N = 5050$.

DGP1: Apply the recurrence formula given in Proposition 2, to obtain the coefficients of the polynomial $\lambda(z) = \sum_{k=0}^{\infty} \lambda_{d,k} z^k$, defined by (8). For this simulation study the infinite sum (8) is truncated at $m = 50,000$. To select the truncation point m we consider Theorem 3 and the results presented in Table 3.

From Theorem 3, we have

$$\lambda_{d,k} \sim \frac{1}{\Gamma(d)k^{1-d}} \frac{\alpha(1)}{\beta(1)}, \quad \text{as } k \to \infty,$$

[5] [11] present a Monte Carlo simulation study on risk measures estimation in time series derived from a FIEGARCH process. [12] analyze a portfolio composed of stocks from the Brazilian market Bovespa. The authors consider the econometric approach to estimate the risk measure VaR and use FIEGARCH models to obtain the conditional variance of the time series.

Table 2: Parameters value for the models. By definition, M1:= FIEGARCH$(2, d, 1)$; M2 := FIEGARCH$(0, d, 4)$; M3 := FIEGARCH$(0, d, 1)$; M4 := FIEGARCH$(0, d, 1)$, M5 := FIEGARCH$(1, d, 1)$ and M6 := FIEGARCH$(1, d, 0)$.

Model	Parameter									
	d	θ	γ	ω	α_1	α_2	β_1	β_2	β_3	β_4
M1	0.4495	-0.1245	0.3662	-6.5769	-1.1190	-0.7619	-0.6195	-	-	-
M2	0.2391	-0.0456	0.3963	-6.6278	-	-	0.2289	0.1941	0.4737	-0.4441
M3	0.4312	-0.1095	0.3376	-6.6829	-	-	0.5454	-	-	-
M4	0.3578	-0.1661	0.2792	-7.2247	-	-	0.6860	-	-	-
M5	0.4900	-0.0215	0.3700	-5.8927	0.1409	-	-0.1611	-	-	-
M6	0.4312	-0.1095	0.3376	-6.6829	0.5454	-	-	-	-	-

and we conclude that $\lambda_{d,k} = o(k^d)$ and $\lambda_{d,k} = O(k^{d-1})$, as k goes to infinity. However, the convergence speed varies from model to model, as we show in Table 3. For simplicity, in this table, let $Q_1(\cdot)$ and $Q_2(\cdot)$ be defined as

$$Q_1(k) := \frac{\lambda_{d,k}}{k^d} \quad \text{and} \quad Q_2(k) := \lambda_{d,k}\left(\frac{1}{\Gamma(d)k^{1-d}}\frac{\alpha(1)}{\beta(1)}\right)^{-1}, \quad \text{for all } k > 0.$$

Table 3 presents the values of the coefficients $\lambda_{d,k}$, given in Proposition 2, for $k \in \{10;\ 100;\ 1000;\ 5000;\ 10000;\ 20000;\ 50000;\ 100000\}$, for each simulated model Mi, $i \in \{1, \cdots, 6\}$. Note that, for $k \geq 5000$, the coefficient values decrease slowly. We also report in Table 3 $Q_1(k)$ and $Q_2(k)$ values for the correspondent $\lambda_{d,k}$ value. Note that, for $k \in \{10{,}000;\ 50{,}000;\ 100{,}000\}$, the value $Q_1(k)$ is very close to zero, for all models. Also, notice that $Q_2(k)$ converges to 1 faster for model M1 than for the other models.

Table 3: Coefficients $\lambda_{d,k}$ and the quotients $Q_1(k)$ and $Q_2(k)$, for different values of k, for all models.

k	10	100	1000	5000	10000	25000	50000	100000
M1 := FIEGARCH$(2, d, 1)$								
$\lambda_{d,k}$	0.26537	0.07167	0.02015	0.00830	0.00567	0.00342	0.00234	0.00160
$Q_1(k)$	0.09426	0.00904	0.00090	0.00018	0.00009	0.00004	0.00002	0.00001
$Q_2(k)$	1.04410	1.00173	1.00017	1.00003	1.00002	1.00001	1.00000	1.00000
M2 := FIEGARCH$(0, d, 4)$								
$\lambda_{d,k}$	-0.09039	0.01450	0.00251	0.00074	0.00043	0.00022	0.00013	0.00008
$Q_1(k)$	-0.05212	0.00482	0.00048	0.00010	0.00005	0.00002	0.00001	0.00000
$Q_2(k)$	-1.08434	1.00292	1.00027	1.00005	1.00003	1.00001	1.00001	1.00000
M3 := FIEGARCH$(0, d, 1)$								
$\lambda_{d,k}$	0.31434	0.07844	0.02106	0.00843	0.00568	0.00337	0.00227	0.00153
$Q_1(k)$	0.11647	0.01077	0.00107	0.00021	0.00011	0.00004	0.00002	0.00001
$Q_2(k)$	1.08789	1.00576	1.00056	1.00011	1.00006	1.00002	1.00001	1.00001
M4 := FIEGARCH$(0, d, 1)$								
$\lambda_{d,k}$	0.36874	0.06738	0.01517	0.00539	0.00345	0.00192	0.00123	0.00079
$Q_1(k)$	0.16178	0.01297	0.00128	0.00026	0.00013	0.00005	0.00003	0.00001
$Q_2(k)$	1.26414	1.01350	1.00129	1.00026	1.00013	1.00005	1.00003	1.00001
M5 := FIEGARCH$(1, d, 1)$								
$\lambda_{d,k}$	0.12291	0.03897	0.01207	0.00531	0.00373	0.00234	0.00164	0.00115
$Q_1(k)$	0.03977	0.00408	0.00041	0.00008	0.00004	0.00002	0.00001	0.00000
$Q_2(k)$	0.97189	0.99720	0.99972	0.99994	0.99997	0.99999	0.99999	1.00000
M6 := FIEGARCH$(1, d, 0)$								
$\lambda_{d,k}$	0.05472	0.01599	0.00435	0.00174	0.00117	0.00070	0.00047	0.00032
$Q_1(k)$	0.02027	0.00219	0.00022	0.00004	0.00002	0.00001	0.00000	0.00000
$Q_2(k)$	0.91632	0.99192	0.99919	0.99984	0.99992	0.99997	0.99998	0.99999

DGP2: Set $Z_0 \sim \text{GED}(\nu)$, with $\nu = 1.5$, and obtain an i.i.d. sample $\{z_t\}_{t=-m}^{N}$.

DGP3: By considering (1), (5) and the equality in (8), the sample $\{x_t\}_{t=1}^{n}$ is obtained through the relation

$$\ln(\sigma_t^2) = \sum_{k=0}^{m} \lambda_{d,k} g(z_{t-1-k}) \quad \text{and} \quad x_t = \sigma_t z_t, \quad \text{for all } t = 1, \cdots, N.$$

Remark 8. For parameter estimation and forecasting procedures we shall consider sub-samples from these time series, with size $n \in \{2000; 5000\}$. The sub-samples of size $n = 2000$ correspond to the last 2000 values of the generated time series (after removing the last 50 values which are used only to compare the out-of-sample forecasting performance of the models). The value $n = 2000$ is the approximated size of the observed time series considered in [12]. The value $n = 5000$ was chosen to analyze the estimators' asymptotic properties.

4.2. Estimation Procedure

In this study we consider the quasi-likelihood method to estimate the parameters of FIEGARCH models for the simulated time series. Given any time series $\{x_t\}_{t=1}^{n}$, this method assumes that $X_t|\mathcal{F}_{t-1}$, for all $t \in \mathbb{Z}$, is normally distributed. The vector of unknown parameters is denoted by

$$\boldsymbol{\eta} = (d; \omega; \theta; \lambda; \alpha_1, \cdots, \alpha_p; \beta_1, \cdots, \beta_q)' \ \in \mathbb{R}^{p+q+4}$$

and the estimator $\hat{\boldsymbol{\eta}}$ of $\boldsymbol{\eta}$ is the value that maximizes

$$\ln(\ell(\boldsymbol{\eta}; x_1, \cdots, x_n)) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\left[\ln(\sigma_t^2) + \frac{x_t^2}{\sigma_t^2}\right]. \tag{30}$$

Since the processes $\{x_t\}_{t<1}$ and $\{z_t\}_{t<1}$ are unknown, we need to consider a set I_0 of initial conditions in order to start the recursion and to obtain the random variables $\ln(\sigma_t^2)$, for $t \in \{1, \cdots, n\}$. Then we use these estimated values to solve (30). For this simulation study we assume, as initial conditions, $g(z_t) = 0$, $\sigma_t^2 = \hat{\sigma}_X^2$ and $x_t := \sigma_t z_t = 0$, whenever $t < 1$, where $\hat{\sigma}_X^2$ is the sample variance of $\{x_t\}_{t=1}^{n}$. This is the initial set suggested by [6]. The random variables $\ln(\sigma_t^2)$, for $t \in \{1, \cdots, n\}$, are then estimated upon considering the set I_0 of initial conditions and the known values $\{x_t\}_{t=1}^{n}$. The infinite sum in the polynomial $\lambda(\cdot)$ is truncated at $m = n$, where n is the available sample size.

4.3. Performance Measures

For any model, let $\hat{\eta}_k$ denote the estimate of η in the k-th replication, where $k \in \{1, \cdots, re\}$, $re = 1{,}000$ and η is any element of the parameter vector given in Table 2. To access the performance of the quasi-likelihood procedure we calculate the mean $\bar{\eta}_i$, the standard deviation (sd), the bias ($bias$), the mean absolute error (mae) and the mean square error (mse) values, defined by

$$\bar{\eta} := \frac{1}{re}\sum_{k=1}^{re}\hat{\eta}_k, \ \ sd := \sqrt{\frac{1}{re}\sum_{k=1}^{re}(\hat{\eta}_k - \bar{\eta})^2}, \ \ bias := \frac{1}{re}\sum_{k=1}^{re} e_k, \ \ mae := \frac{1}{re}\sum_{k=1}^{re}|e_k| \ \text{ and } \ mse := \frac{1}{re}\sum_{k=1}^{re} e_k^2,$$

where $e_k := \hat{\eta}_k - \eta$, for $k \in \{1, \cdots, re\}$.

4.4. Estimation Results

Table 4 summarizes the results of the parameter estimation procedure. Figures 6 - 11 present the kernel estimates of the probability density function of the parameter estimates of each considered model when $n \in \{2000; 5000\}$. These graphs help to illustrate the results presented in Table 4.

By observing Figures 6 - 11, it is easy to see that, for most estimates, the density function is approximately symmetric. For some parameters, we notice the presence of possible outliers, see for instance the graphs for

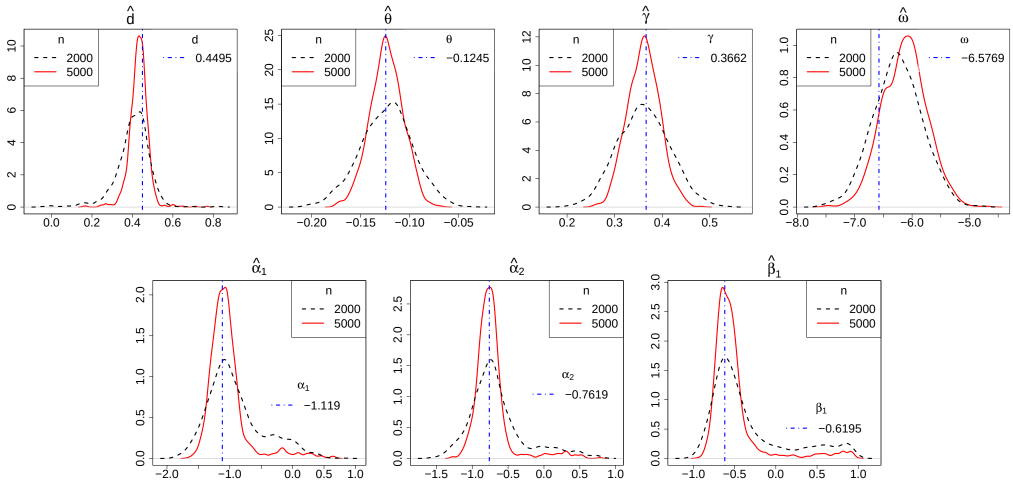


Figure 6: Kernel estimates of the probability density function of the parameter estimates corresponding to model M1, for $n \in \{2000; 5000\}$.

the parameters d (in particular, models M2, M3 and M4), α_i (model M1) and β_j (in particular, models M1 and M2), with $i \in \{1, 2\}$ and $j \in \{1, 2, 3, 4\}$. Although the graphs for $n = 2000$ and $n = 5000$ are similar, one observes that, as expected, the observations tend to concentrate closer to the mean when $n = 5000$.

From Table 4 we conclude that, given the models' complexity, the quasi-likelihood method performs relatively well. Since model M2 presents more parameters than the other models, which implies a higher dimension maximization problem, one would expect that the quasi-likelihood method would present the worst performance in this case. However, in terms of *mae* or *mse* values, the estimation results for model M2 ($p = 0$, $d = 0.2391$ and $q = 4$), M3 ($p = 0$, $d = 0.4312$ and $q = 1$), M4 ($p = 0$, $d = 0.3578$ and $q = 1$) and M6 ($p = 1$, $d = 0.4312$ and $q = 0$) are similar (except for the parameter d in model M6) and the quasi-likelihood method performs better for model M2 (except for the parameter d) than for models M1 ($p = 2$, $d = 0.4495$ and $q = 1$) and M5 ($p = 1$, $d = 0.49$ and $q = 1$).

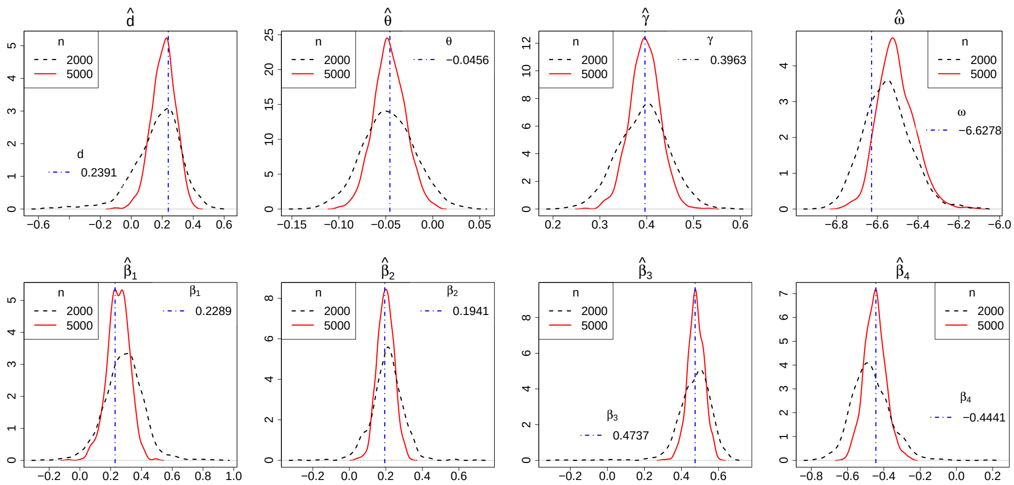


Figure 7: Kernel estimates of the probability density function of the parameter estimates corresponding to model M2, for $n \in \{2000; 5000\}$.

Table 4 also indicates that the quasi-likelihood procedure may perform better for $p = 0$ and $q > 0$ than for $p > 0$ and $q = 0$ (we shall investigate this in a future work). This conclusion is based on the fact that

Table 4: Estimation results for the simulated FIEGARCH models considering $re = 1000$ replications.

Sample Size (n)	2000					5000				
Parameter (η)	$\bar{\eta}$	sd	$bias$	mae	mse	$\bar{\eta}$	sd	$bias$	mae	mse
M1 := FIEGARCH$(2, d, 1)$										
0.4495 (d)	0.4022	0.0854	-0.0473	0.0688	0.0095	0.4309	0.0468	-0.0186	0.0357	0.0025
-0.1245 (θ)	-0.1240	0.0266	0.0005	0.0213	0.0007	-0.1237	0.0168	0.0008	0.0133	0.0003
0.3662 (γ)	0.3612	0.0543	-0.0050	0.0438	0.0030	0.3610	0.0337	-0.0052	0.0271	0.0012
-6.5769 (ω)	-6.2516	0.4270	0.3253	0.4358	0.2881	-6.1284	0.3830	0.4485	0.4930	0.3479
-1.1190 (α_1)	-0.9067	0.4519	0.2123	0.3567	0.2492	-1.0344	0.3259	0.0846	0.2010	0.1134
-0.7619 (α_2)	-0.6517	0.4035	0.1102	0.2832	0.1750	-0.7281	0.2623	0.0338	0.1534	0.0700
-0.6195 (β_1)	-0.3415	0.4774	0.2780	0.3474	0.3052	-0.5052	0.3214	0.1143	0.1764	0.1164
M2 := FIEGARCH$(0, d, 4)$										
0.2391 (d)	0.1683	0.1538	-0.0708	0.1216	0.0287	0.2077	0.0767	-0.0314	0.0650	0.0069
-0.0456 (θ)	-0.0469	0.0275	-0.0013	0.0220	0.0008	-0.0461	0.0169	-0.0005	0.0134	0.0003
0.3963 (γ)	0.3931	0.0536	-0.0032	0.0426	0.0029	0.3959	0.0326	-0.0004	0.0256	0.0011
-6.6278 (ω)	-6.5525	0.1146	0.0753	0.1075	0.0188	-6.5077	0.0905	0.1201	0.1253	0.0226
0.2289 (β_1)	0.2841	0.1284	0.0552	0.1083	0.0195	0.2488	0.0721	0.0199	0.0593	0.0056
0.1941 (β_2)	0.2078	0.0865	0.0137	0.0657	0.0077	0.1990	0.0456	0.0049	0.0367	0.0021
0.4737 (β_3)	0.4710	0.0935	-0.0027	0.0667	0.0088	0.4784	0.0441	0.0047	0.0349	0.0020
-0.4441 (β_4)	-0.4704	0.1063	-0.0263	0.0867	0.0120	-0.4500	0.0592	-0.0059	0.0466	0.0035
M3 := FIEGARCH$(0, d, 1)$										
0.4312 (d)	0.3606	0.1268	-0.0706	0.1043	0.0211	0.3933	0.0648	-0.0379	0.0569	0.0056
-0.1095 (θ)	-0.1111	0.0255	-0.0016	0.0201	0.0007	-0.1090	0.0157	0.0005	0.0125	0.0002
0.3376 (γ)	0.3346	0.0493	-0.0030	0.0394	0.0024	0.3331	0.0300	-0.0045	0.0241	0.0009
-6.6829 (ω)	-6.3686	0.4230	0.3143	0.4271	0.2778	-6.2413	0.3715	0.4416	0.4814	0.3330
0.5454 (β_1)	0.5976	0.1472	0.0522	0.1231	0.0244	0.5822	0.0851	0.0368	0.0731	0.0086
M4 := FIEGARCH$(0, d, 1)$										
0.3578 (d)	0.2950	0.1338	-0.0628	0.1056	0.0218	0.3258	0.0721	-0.0320	0.0569	0.0062
-0.1661 (θ)	-0.1702	0.0248	-0.0041	0.0198	0.0006	-0.1666	0.0156	-0.0005	0.0124	0.0002
0.2792 (γ)	0.2793	0.0415	0.0001	0.0326	0.0017	0.2769	0.0248	-0.0023	0.0197	0.0006
-7.2247 (ω)	-6.9615	0.3122	0.2632	0.3284	0.1667	-6.8766	0.2604	0.3481	0.3689	0.1889
0.6860 (β_1)	0.7160	0.1128	0.0300	0.0915	0.0136	0.7067	0.0665	0.0207	0.0535	0.0048
M5 := FIEGARCH$(1, d, 1)$										
0.4900 (d)	0.4258	0.1273	-0.0642	0.1096	0.0203	0.4453	0.0645	-0.0447	0.0629	0.0062
-0.0215 (θ)	-0.0229	0.0355	-0.0014	0.0282	0.0013	-0.0229	0.0218	-0.0014	0.0175	0.0005
0.3700 (γ)	0.3751	0.0577	0.0051	0.0455	0.0034	0.3742	0.0354	0.0042	0.0285	0.0013
-5.8927 (ω)	-5.7507	0.2688	0.1420	0.2415	0.0924	-5.6414	0.2494	0.2513	0.2902	0.1253
0.1409 (α_1)	0.1152	0.4082	-0.0257	0.3232	0.1673	0.1012	0.3310	-0.0397	0.2613	0.1111
-0.1611 (β_1)	-0.1383	0.3799	0.0228	0.3189	0.1448	-0.1581	0.3213	0.0030	0.2579	0.1032
M6 := FIEGARCH$(1, d, 0)$s										
0.4312 (d)	0.3220	0.1825	-0.1092	0.1706	0.0452	0.3449	0.1135	-0.0863	0.1107	0.0203
-0.1095 (θ)	-0.1132	0.0351	-0.0037	0.0282	0.0012	-0.1114	0.0222	-0.0019	0.0176	0.0005
0.3376 (γ)	0.3368	0.0585	-0.0008	0.0467	0.0034	0.3380	0.0355	0.0004	0.0283	0.0013
-6.6829 (ω)	-6.6233	0.1144	0.0596	0.1014	0.0166	-6.5926	0.0978	0.0903	0.1071	0.0177
0.5454 (α_1)	0.4189	0.2297	-0.1265	0.2109	0.0688	0.4429	0.1492	-0.1025	0.1428	0.0328

models M3 and M6 have the same parameter values (with the necessary adjustments in α_1 and β_1) and all parameters, except ω, were better estimated in model M3 than M6.

By comparing the mae and mse values, given in Table 4, we conclude that the worst performance occurs for models M1 and M5 (in particular, see the estimation results for ω, α_i and β_j, $i = 1, \cdots, p$ and $j = 1, \cdots, q$). This outcome is explained by the fact that the parameter d is very close to the non-stationary

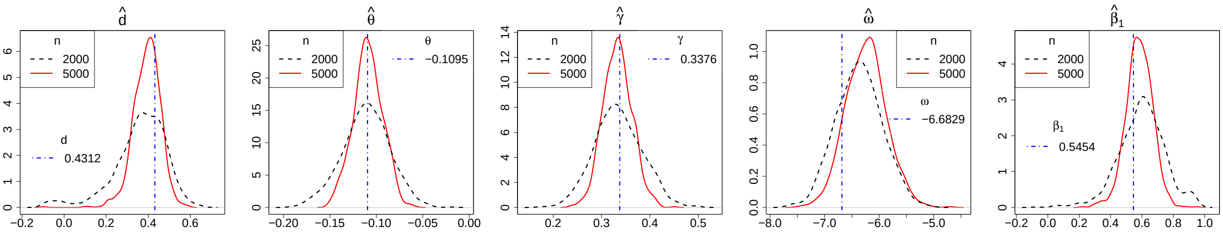


Figure 8: Kernel estimates of the probability density function of the parameter estimates corresponding to model M3, for $n \in \{2000; 5000\}$.

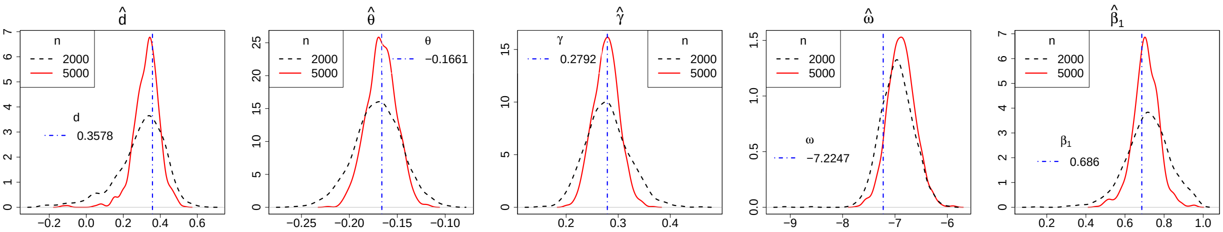


Figure 9: Kernel estimates of the probability density function of the parameter estimates corresponding to model M4, for $n \in \{2000; 5000\}$.

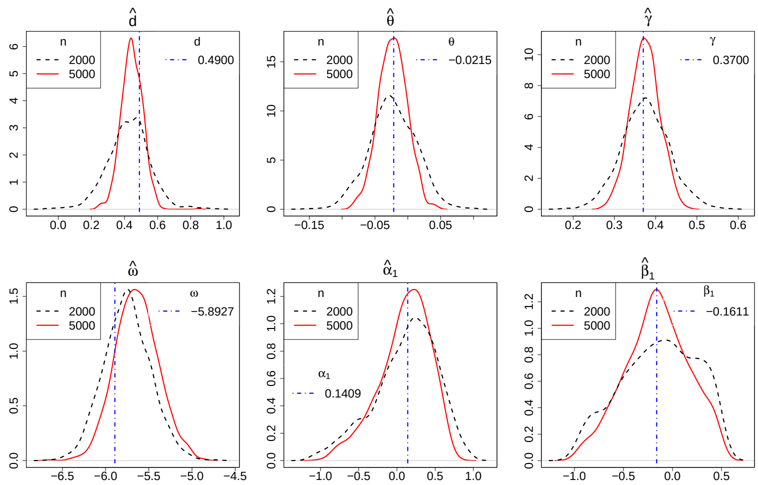


Figure 10: Kernel estimates of the probability density function of the parameter estimates corresponding to model M5, for $n \in \{2000; 5000\}$.

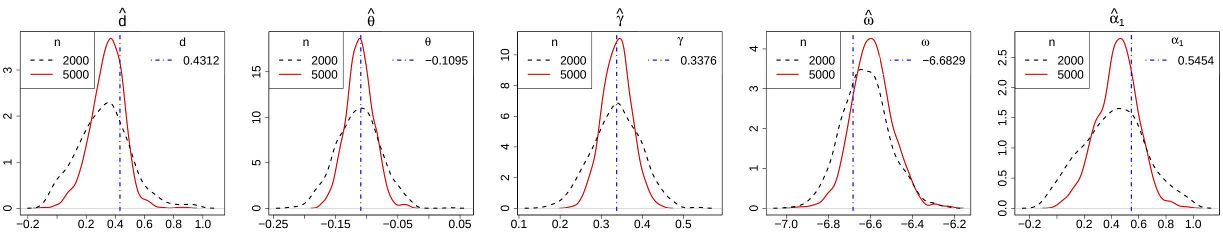


Figure 11: Kernel estimates of the probability density function of the parameter estimates corresponding to model M6, for $n \in \{2000; 5000\}$.

region for model M5 and, for model M1, not only $p = 2$ but also $d = 0.4495$, which implies a more complex model with stronger long-range dependence. The small *bias* values indicate that, for all parameters, the mean estimated value is very close to the true value. Although for $n = 2000$ the standard deviation of several estimates is high if compared with the mean estimated value, as expected, the estimators' performance improves as the sample size increases, except for the parameter ω.

4.5. Forecasting Procedure

To obtain the predicted values, for each replication of model Mi, with $i \in \{1, \cdots, 6\}$, and each sub-sample $\{x_t\}_{t=1}^{n}$, with $n \in \{2000; 5000\}$, we repeat steps **F1 - F5** below.

F1: Replace the true parameter values $\boldsymbol{\eta} = (d; \omega; \theta; \lambda; \alpha_1, \cdots, \alpha_p; \beta_1, \cdots, \beta_q)'$ by the estimated ones, namely, $\hat{\boldsymbol{\eta}} = (\hat{d}; \hat{\omega}; \hat{\theta}; \hat{\lambda}; \hat{\alpha}_1, \cdots, \hat{\alpha}_p; \hat{\beta}_1, \cdots, \hat{\beta}_q)'$, and use the recurrence formula given in Proposition 2 to calculate the corresponding coefficients $\{\hat{\lambda}_{d,k}\}_{k=0}^{n+50}$.

F2: Obtain the time series $\{z_t\}_{t=1}^{n}$ (which corresponds to the residuals of the fitted model) and $\{\sigma_t\}_{t=1}^{n}$. To do so, let $g(z_t) = 0$, whenever $t < 0$, and calculate σ_t and z_t recursively as follows:

$$\sigma_1 = e^{\hat{0}.5\omega}; \quad z_1 = \frac{x_1}{\sigma_1}; \quad \sigma_t = \exp\left\{\frac{\hat{\omega}}{2} + \frac{1}{2}\sum_{k=0}^{n-1}\hat{\lambda}_{d,k}\left[\hat{\theta} z_{t-1-k} + \hat{\gamma}(|z_{t-1-k}| - \sqrt{2/\pi})\right]\right\} \quad \text{and} \quad z_t = \frac{x_t}{\sigma_t},$$

for all $t = 2, \cdots, n$.

F3: In expression (11), replace $\mathbb{E}(|Z_0|)$ and $\mathbb{E}(Z_0|Z_0|)$ by their respective sample estimates, and obtain an estimate $\hat{\sigma}_g^2$ for σ_g^2 given by

$$\hat{\sigma}_g^2 = \hat{\theta}^2 + \hat{\gamma}^2 - \hat{\gamma}^2\left[\frac{1}{n}\sum_{t=1}^{n}|z_t|\right]^2 + 2\,\hat{\theta}\,\hat{\gamma}\left[\frac{1}{n}\sum_{t=1}^{n}z_t|z_t|\right].$$

F4: By considering expressions (23) and (28), obtain the predicted values $\{\tilde{\sigma}_{N+h}^2\}_{h=1}^{50}$,

$$\tilde{\sigma}_{N+1}^2 = \check{\sigma}_{N+1}^2 \qquad \text{and} \qquad \tilde{\sigma}_{N+h}^2 = \check{\sigma}_{N+h}^2\left(1 + \frac{1}{2}\hat{\sigma}_g^2\sum_{k=0}^{h-2}\hat{\lambda}_{d,k}^2\right), \quad \text{for all } h > 1,$$

where

$$\check{\sigma}_{N+h}^2 = \exp\left\{\hat{\omega} + \sum_{k=0}^{n-1}\hat{\lambda}_{d,k+h-1}\left[\hat{\theta} z_{n-k} + \hat{\gamma}(|z_{n-k}| - \hat{\mu}_{|z|})\right]\right\}, \quad \text{for all } h > 0,$$

with $\hat{\mu}_{|z|} := \frac{1}{n}\sum_{t=1}^{n}|z_t|$.

F5: Based on the fact that $\mathbb{E}(X_{N+h}^2|\mathcal{F}_N) = \mathbb{E}(\sigma_{N+h}^2|\mathcal{F}_N)$, set $\tilde{X}_{N+h}^2 := \tilde{\sigma}_{N+h}^2$, for all $h > 0$.

4.6. Forecasting Results

In what follows we discuss the simulation results related to forecasting based on the fitted FIEGARCH models. To access the models' forecast performance, during the generating process, we create 50 extra values for each simulated time series. Those values are used here to compare with the h-step ahead forecast, for $h \in \{1, \cdots, 50\}$.

Table 5 presents the mean over 1000 simulated values of σ_{N+h}^2 and X_{N+h}^2 obtained from model Mi, for each $i \in \{1, \cdots, 6\}$, and the corresponding h-step ahead predicted values $\tilde{\sigma}_{N+h}^2 := \tilde{X}_{N+h}^2$, for $h \in \{1, \cdots, 5\}$, forecasting origin $N = 5000$ and sub-samples $n \in \{2000; 5000\}$. This table also reports the mean square error (mse) of forecast, defined as

$$mse(Y_{N+h}) := \frac{1}{re}\sum_{k=1}^{re}\big(Y_{N+h}^{(k)} - \check{Y}_{N+h}^{(k)}(n)\big)^2, \quad \text{for any} \;\; h \in \{1, \cdots, 5\} \;\; \text{and} \;\; n \in \{2000; 5000\},$$

where $re = 1000$ is the number of replications, Y_{N+h} denotes the true value of σ^2_{N+h} (or X^2_{N+h}) and $\check{Y}^{(k)}_{N+h}(n)$ is the predicted value obtained in the k-th replication, for $k \in \{1, \cdots, re\}$, based on the model fitted to the sub-sample with size n. Notice that, due to the small magnitude of the sample means, all values in Table 5 are multiplied by 100.

From Table 5 (see also Figure 12 below) we conclude that,

- when we consider σ^2_{N+h}, the predicted values are relatively close to the simulated ones, which is indicated by the small mse values, for all models and any $h \in \{1, \cdots, 6\}$;
- the mse value increases as h increases. This result is expected and it is theoretically explained in Proposition 4 which shows that

$$\mathbb{E}\big(\big[\ln(\sigma^2_{n+h}) - \hat{\ln}(\sigma^2_{n+h})\big]^2\big) = \sigma^2_g \sum_{k=0}^{h-2}\lambda^2_{d,k} \overset{h\to\infty}{\longrightarrow} \sigma^2_g \sum_{k=0}^{\infty}\lambda^2_{d,k},$$

  where $\sigma^2_g := \mathbb{E}([g(Z_0)]^2)$ is given in (11);
- when we consider X^2_{N+h}, the mse is usually high, if compared to the mean simulated and mean predicted values. Therefore, we conclude that $\tilde{X}^2_{n+h} := \tilde{\sigma}^2_{n+h}$ is a poor estimator for X^2_{n+h}. This result is not a surprise since the main purpose of FIEGARCH models is to estimate the logarithm of the conditional variance of the process and not the process itself;
- as expected, in all cases, the models' forecasting performance improves as n increases. Notice, however, that the difference in the mse values, from $n = 2000$ to $n = 5000$, is small (recall that the values are multiplied by 100). This is so because the coefficients $\lambda_{d,k}$ converges to zero, as k goes to infinity. Therefore, it is expected that, for some $m \in \mathbb{N}$ and any $M > 0$, using the last m or the last $m + M$ known values to calculate the h-step ahead forecast value for the process will not considerably change the results.

Figure 12 shows the mean taken over 1000 replications for:

- the simulated values σ^2_{N+h} and X^2_{N+h} obtained from model Mi, for each $i \in \{1, \cdots, 6\}$, $N = 5000$ and $h \in \{1, \cdots, 50\}$;
- the one-step ahead forecast values $\check{\sigma}^2_{N^*+1} := \tilde{\sigma}^2_{N^*+1}$ (denoted in the graphs by $\hat{\sigma}^2_{N+h-1}(1)$), for $N^* = N + h$, $N = 5000$ and $h \in \{1, \cdots, 50\}$. The predictor $\check{\sigma}^2_N(1)$ is obtained directly from the sub-sample $\{x_t\}_{t=1}^n$, by following steps **F1 -F5** (this figure only reports the graphs for the case $n = 5000$). The remaining predicted values $\{\check{\sigma}^2_{N+h-1}(1)\}_{h=2}^{50}$ are calculated by updating the forecasting origin from $N = 5000$ to $N^* = N + h - 1$, that is, by introducing the observations $\{X_{N+h}\}_{h=1}^{49}$, one at a time, and following steps **F2 -F5**;
- the h-step ahead forecast values considering the predictors $\tilde{\sigma}^2_{N+h}$ and $\check{\sigma}^2_{N+h}$ (denoted in the graphs by $\sigma^2_N(h)$). These values are obtained by following steps **F1 -F5** with forecasting origin $N = 5000$ (without update). For all graphs the size of the sub-sample used for parameter estimation and forecasting is $n = 5000$.

The dashed lines in Figure 12 correspond to the limiting constants $L_1(i)$ and $L_2(i)$, for $i \in \{1, \cdots, 6\}$, described in the sequel.

Table 5: Mean simulated values for σ^2_{N+h} and X^2_{N+h}, obtained from model Mi, the corresponding mean predicted values $\tilde{\sigma}^2_{N+h} = \tilde{X}^2_{N+h}$ and the mean square error of forecast, for $h \in \{1, \cdots, 5\}$ and $i \in \{1, \cdots, 6\}$. The forecasting origin is $N = 5000$ and $n \in \{2000; 5000\}$ is the sub-sample size used to fit the models and to obtain the predicted values. All values reported correspond to the calculated values multiplied by a scaling constant (except h). The scaling constant is equal to 10^2, for σ^2_{N+h}, X^2_{N+h} and $\tilde{\sigma}^2_{N+h}$, and to 10^4, for the mse values. The number of considered replications is $re = 1000$.

n			2000			5000		
h	σ^2_{N+h}	X^2_{N+h}	Predictor	$mse(\sigma^2_{N+h})$	$mse(X^2_{N+h})$	Predictor	$mse(\sigma^2_{N+h})$	$mse(X^2_{N+h})$
M1 := FIEGARCH$(2, d, 1)$								
1	0.1698	0.1575	0.1652	0.0010	0.0993	0.1634	0.0003	0.0969
2	0.1635	0.1473	0.1640	0.0038	0.0900	0.1611	0.0032	0.0875
3	0.1636	0.1540	0.1655	0.0078	0.1122	0.1632	0.0075	0.1116
4	0.1629	0.1490	0.1662	0.0122	0.1117	0.1633	0.0114	0.1101
5	0.1641	0.1542	0.1665	0.0147	0.1906	0.1642	0.0141	0.1903
M2 := FIEGARCH$(0, d, 4)$								
1	0.1387	0.1284	0.1359	0.0004	0.0521	0.1369	0.0002	0.0515
2	0.1383	0.1246	0.1395	0.0024	0.0506	0.1398	0.0021	0.0501
3	0.1357	0.1299	0.1374	0.0027	0.0551	0.1381	0.0024	0.0547
4	0.1378	0.1276	0.1390	0.0029	0.0562	0.1399	0.0028	0.0559
5	0.1356	0.1253	0.1409	0.0036	0.0568	0.1414	0.0034	0.0570
M3 := FIEGARCH$(0, d, 1)$								
1	0.1487	0.1380	0.1452	0.0007	0.0833	0.1439	0.0002	0.0848
2	0.1456	0.1287	0.1459	0.0026	0.0681	0.1442	0.0022	0.0674
3	0.1426	0.1350	0.1466	0.0052	0.0773	0.1447	0.0045	0.0777
4	0.1438	0.1200	0.1473	0.0075	0.0619	0.1453	0.0068	0.0600
5	0.1411	0.1354	0.1479	0.0076	0.1316	0.1459	0.0069	0.1309
M4 := FIEGARCH$(0, d, 1)$								
1	0.0932	0.0894	0.0918	0.0005	0.0411	0.0910	0.0002	0.0416
2	0.0905	0.0809	0.0918	0.0013	0.0275	0.0908	0.0010	0.0270
3	0.0885	0.0810	0.0917	0.0027	0.0293	0.0907	0.0022	0.0291
4	0.0886	0.0764	0.0918	0.0040	0.0251	0.0908	0.0036	0.0242
5	0.0876	0.0831	0.0919	0.0042	0.0461	0.0909	0.0037	0.0456
M5 := FIEGARCH$(1, d, 1)$								
1	0.2898	0.2669	0.2808	0.0012	0.2096	0.2795	0.0005	0.2087
2	0.2883	0.2800	0.2833	0.0069	0.2489	0.2817	0.0065	0.2494
3	0.2908	0.2836	0.2844	0.0081	0.2452	0.2821	0.0081	0.2461
4	0.2909	0.2963	0.2847	0.0077	0.3178	0.2827	0.0076	0.3174
5	0.2923	0.2971	0.2852	0.0097	0.3695	0.2832	0.0096	0.3704
M6 := FIEGARCH$(0, d, 1)$								
1	0.1271	0.1143	0.1242	0.0001	0.0367	0.1247	0.0001	0.0367
2	0.1260	0.1140	0.1265	0.0013	0.0379	0.1265	0.0013	0.0377
3	0.1259	0.1228	0.1262	0.0014	0.0471	0.1263	0.0013	0.0471
4	0.1284	0.1188	0.1263	0.0018	0.0479	0.1264	0.0017	0.0477
5	0.1261	0.1192	0.1264	0.0015	0.0473	0.1265	0.0014	0.0474

From Figure 12 we observe that, for all models, the means for the one-step ahead forecast values $\breve{\sigma}^2_{N^*+1}$, show the same behavior over time as the means for the true values $\sigma^2_{N^*+1}$, where $N^* = N+h-1$, $N = 5000$ and $h \in \{1, \cdots, 50\}$. As expected, due to the error carried from the parameter estimation (specially, from the distribution misspecification), we observe a small forecasting bias, which decreases as h increases. The decrease in the forecasting bias, as the forecasting origin is updated, can be attributed to the fact that we start the recurrence formula (step **F2**) assuming $\mathbb{E}(|Z_0|) = \sqrt{2/\pi}$ and as the new observations X_{N+h} are introduced, the constant $\mathbb{E}(|Z_0|)$ is replaced by its sample estimate (step **F3**), which provides more accurate

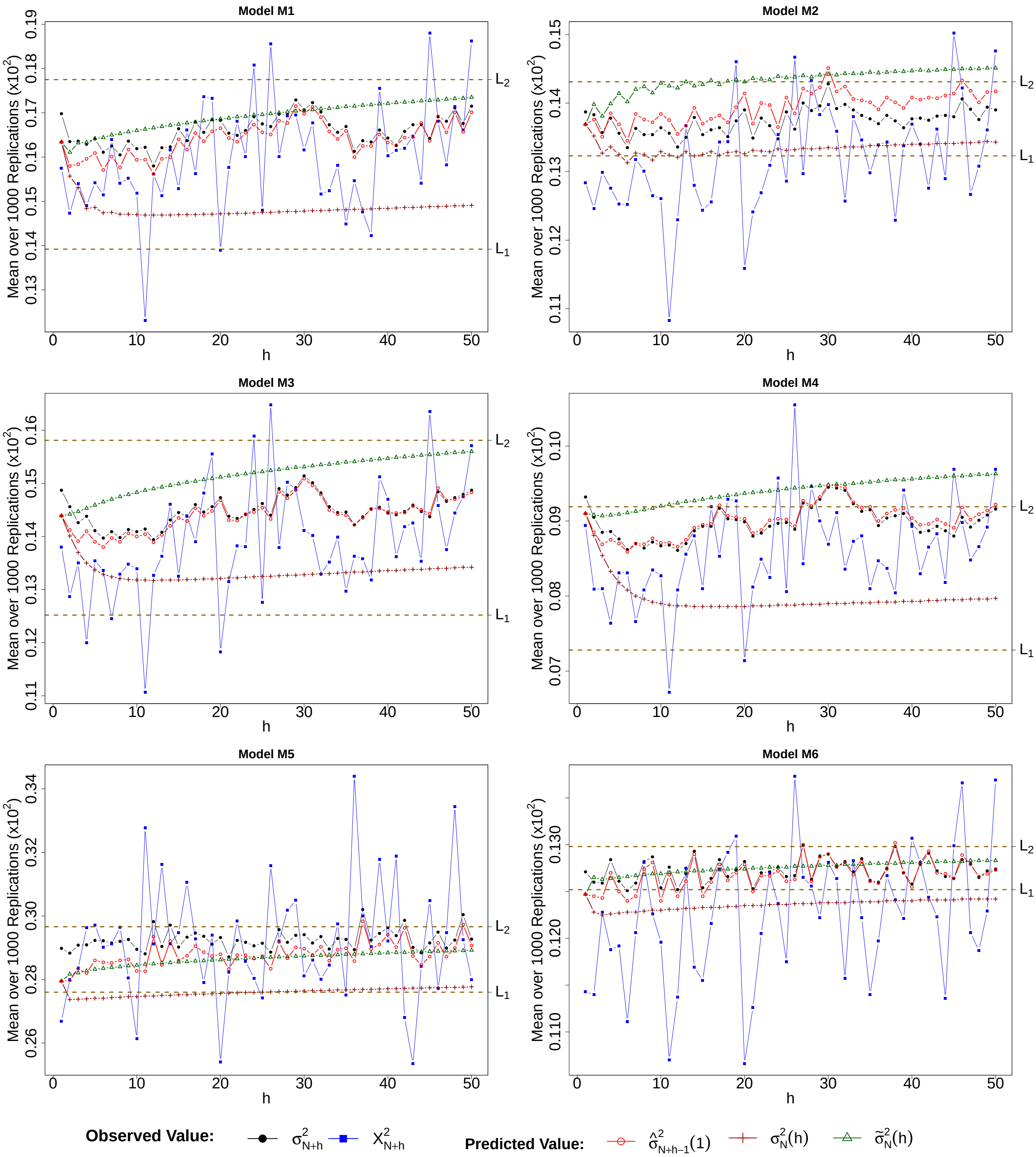


Figure 12: For each model Mi, $i \in \{1, \cdots, 6\}$: the simulated values for σ^2_{N+h}; the one-step ahead forecast $\check{\sigma}^2_{N^*+1} := \tilde{\sigma}^2_{N^*+1}$ (denoted in the graphs by $\hat{\sigma}^2_{N+h-1}(1)$), obtained by updating the forecasting origin to $N^* = N + h - 1$; the h-step ahead forecast values considering the predictors $\tilde{\sigma}^2_{N+h}$ and $\check{\sigma}^2_{N+h}$ (denoted in the graphs by $\sigma^2_N(h)$), with forecasting origin N. For all models $h \in \{1, \cdots, 50\}$, $N = 5000$ and the size of the sub-sample used for parameter estimation and forecasting is $n = 5000$. All values in the graphs correspond to the mean taken over 1000 replications.

values for $g(Z_t)$ as t increases ($t > N$).

Regarding the h-step ahead predictors $\breve{\sigma}^2_{i,n+h}$ and $\tilde{\sigma}^2_{i,n+h}$, Figure 12 shows that the estimation bias is higher if we consider the former one. This figure also shows that, for all models, the predicted value converges to a constant as h increases. This is expected since the h-step ahead predictor is defined in terms of the conditional expectation. In fact, from expression (26), $\breve{\sigma}^2_{N+h}$ converges to $L_1(i) := e^{\omega(i)}$ as h goes to infinity, where $\omega(i)$ denotes the parameter ω for model Mi and hence, from expression (29),

$$\tilde{\sigma}^2_{N+h} := \breve{\sigma}^2_{N+h}\left(1 + \frac{1}{2}\sigma^2_g \sum_{k=0}^{h-2} \lambda^2_{d,k}\right) \xrightarrow{h\to\infty} e^{\omega(i)}\left(1 + \frac{1}{2}\sigma^2_g(i) \sum_{k=0}^{\infty} \lambda^2_{d,k}(i)\right)$$
$$\approx e^{\omega(i)}\left(1 + \frac{1}{2}\sigma^2_g(i) \sum_{k=0}^{m} \lambda^2_{d,k}(i)\right) := L_2(i), \qquad (31)$$

for each $i \in \{1, \cdots, 6\}$ and m sufficiently large. The values of $\omega(i)$ (also given in Table 2), $L_1(i)$ and $L_2(i)$, for $m = 50,000$ and $i \in \{1, \cdots, 6\}$, are presented in Table 6.

Table 6: Values of $\omega(i)$, $L_1(i) := e^{\omega(i)}$ and $L_2(i)$, defined in (31), for $m = 50,000$ and $i \in \{1, \cdots, 6\}$.

i	1	2	3	4	5	6
$\omega(i)$	-6.5769	-6.6278	-6.6829	-7.2247	-5.8927	-6.6829
$L_1(i) \times 100$	0.1392	0.1323	0.1252	0.0728	0.2760	0.1252
$L_2(i) \times 100$	0.1775	0.1431	0.1581	0.0919	0.2966	0.1298

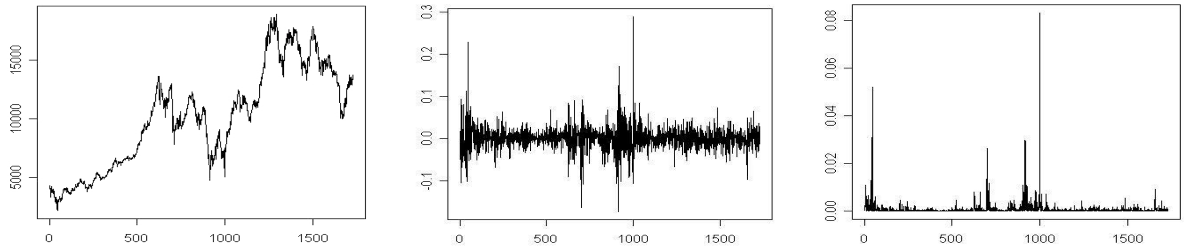
Upon comparing the values of $L_1(i)$ and $L_2(i)$, given in Table 6 (also reported in Figure 12 as L_1 and L_2), for each $i \in \{1, \cdots, 6\}$, respectively, with the limits $\lim_{h\to\infty} \breve{\sigma}^2_{N+h}$ and $\lim_{h\to\infty} \tilde{\sigma}^2_{N+h}$ (see Figure 12), we conclude that these values are close to each other, for all models. A small difference between $L_1(i)$ and $\lim_{h\to\infty} \breve{\sigma}^2_{N+h}$ (respectively, $L_2(i)$ and $\lim_{h\to\infty} \tilde{\sigma}^2_{N+h}$) is expected since the former is calculated using the true parameter values while $\breve{\sigma}^2_{N+h}$ is obtained by considering the estimates for the parameter values.

5. Analysis of an Observed Time Series

This section presents the analysis of the São Paulo Stock Exchange Index (Bovespa Index or IBovespa) log-return time series. We consider the FIEGARCH model, fully described in this paper, and we compare its forecasting performance with other ARCH-type models. The total number of observations for the IBovespa time series is $n = 1737$. We consider the first 1717 observations to fit the models and we reserve the last 20 ones to compare with the out-of-sample forecast.

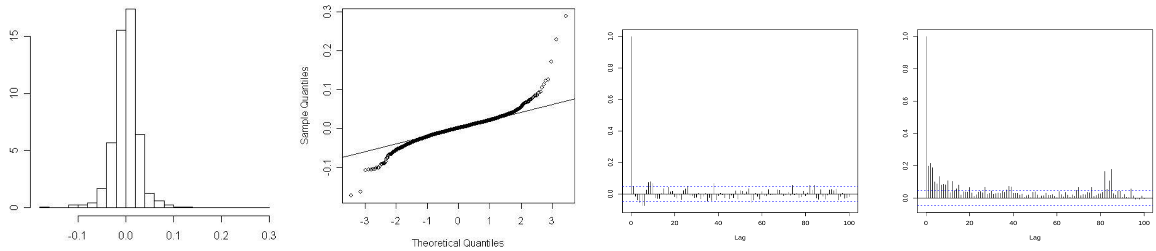
Figure 13 (a) presents IBovespa time series $\{P_t\}_{t=1}^{1718}$, in the period of January/1995 to December/2001. We observe a strong decay in the index value close to $t = 1000$ (that is, January 15, 1999). This period is characterized by the Real (the Brazilian currency) devaluation. Figures 13 (b) and (c) present, respectively, the IBovespa log-return time series, $\{r_t\}_{t=1}^{1717}$, and the square of the log-return time series, $\{r_t^2\}_{t=1}^{1717}$, in the same period. Observe that the log-return time series presents the stylized facts of financial time series such as apparent stationarity, mean around zero and clusters of volatility. Also, in Figure 14 we observe that, while the log-return time series presents almost no correlation, the sample autocorrelation of the square of the log-return time series assumes high values for several lags, pointing to the existence of heteroskedasticity and possibly long-range dependence. Notice that the periodogram of $\{\ln(r_t^2)\}_{t=1}^{1717}$, presented in Figure 4 (c), also indicates possibly long-range dependence in the conditional variance. Regarding the histogram and the QQ-plot, we observe that the distribution of the log-return time series seems approximately symmetric and leptokurtic.

To investigate whether the stationarity property holds for the time series $\{r_t\}_{t=1}^{1717}$ we apply the runs test (or Wald-Wolfwitz test), as described in [30]. Due to the magnitude of the data we multiply the time



(a) $\{P_t\}_{t=1}^{1718}$ (b) $\{r_t\}_{t=1}^{1717}$ (c) $\{r_t^2\}_{t=1}^{1717}$

Figure 13: Time series: (a) Bovespa index; (b) IBovespa log-returns; (c) square of the IBovespa log-returns, in the period of January/1995 to December/2001.



(a) (b) (c) (d)

Figure 14: (a) Histogram; (b) QQ-plot; (c) sample autocorrelation of the IBovespa log-return time series; (d) sample autocorrelation of the square of the IBovespa log-return time series.

series values by 100 before applying the test. The p-values for the test considering the moments of order[6] $r \in \{1, \cdots, 10\}$ are reported in Figure 15. For comparison, this figure also shows the p-values of the test applied to the simulated time series presented in Figure 1. Notice that, for all $r \in \{1, \cdots, 10\}$ the test does not reject the null hypothesis of stationarity at a 5% significance level.

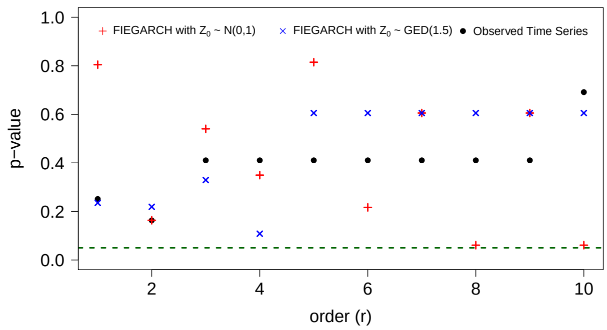


Figure 15: The p-values for the stationarity hypothesis test considering the moments of order $r \in \{1, \cdots, 10\}$. The dashed line corresponds to p-value = 0.05.

To analyze if the ergodicity property holds for the time series $\{r_t\}_{t=1}^{1717}$ we perform the test described in [31]. For comparison, we also apply this test to the simulated time series (only for sample size $n = 2000$) considered in Section 4. The test results are given in Table 7. The reported values are the proportion of p-values smaller than 0.05 and 0.10 in a total of 100 repetitions of step 3 of the Algorithm 1 given in

[6]For $r > 10$ the values of $\{r_t^r\}_{t=1}^{1,717}$ are too close to zero and the test always returns the same p-value as for $r = 10$.

[31]. Moreover, for the simulated time series, the values in Table 7 correspond to the mean taken over 1000 replications. Notice that the proportion of p-values smaller than 0.05 (respectively, 0.10) is always higher for the simulated time series (known to be ergodic) then for the observed time series. Given that the proportion of p-values smaller than 0.05 and 0.10 is close to the expected, we conclude that the ergodicity property holds for $\{r_t\}_{t=1}^{1717}$.

Table 7: Proportion of p-values smaller than 0.05 and 0.10 in a total of 100 repetitions of step 3 of the Algorithm 1 given in [31] for the simulated time series obtained from model Mi, with $i \in \{1, \cdots, 6\}$, and for the observed time series $\{r_t\}_{t=1}^{1717}$.

p-values	M1	M2	M3	M4	M5	M6	$\{r_t\}_{t=1}^{1717}$
0.05	0.10	0.08	0.09	0.09	0.07	0.07	0.05
0.10	0.17	0.13	0.14	0.15	0.13	0.12	0.11

The analysis of the sample autocorrelation function suggests an ARMA(p_1, q_1)-FIEGARCH(p_2, d, q_2) model. While an ARMA model accounts for the correlation among the log-returns, a FIEGARCH model takes into account the long-range dependence (in the conditional variance) and the heteroskedasticity characteristics of the time series. To select the best ARMA(p_1, q_1)-FIEGARCH(p_2, d, q_2) model for the data we initially considered all possible models with $p_1, q_1 \in \{0, 1, 2, 3\}$ and $p_2, q_2 \in \{0, 1, 2\}$ and applied the quasi-likelihood method to estimate the unknown parameters. Then we eliminated the models with correlated residuals and selected the best models, with respect to the log-likelihood, Bayesian (BIC), Akaike (AIC) and Hannan-Quinn (HQC) information criteria (in this step three models were selected). The models order and the corresponding AIC, BIC and HQC values are reported in Table 8. Boldface indicates that the model was the best with respect to the corresponding criterion.

Table 8: Log-likelihood value and Bayesian (BIC), Akaike (AIC) and Hannan-Quinn (HQC) information criteria values for three competitive ARMA(p_1, q_1)-FIEGARCH(p_2, d, q_2) models fitted to the IBovespa log-return time series.

Order					Criterion			
p_1	q_1	p_2	d	q_2	Log-likelihood	BIC	AIC	HQC
3	2	1	0.3651	1	**4142.260**	-8202.588	-8262.520	-8240.344
0	1	0	0.3578	1	4138.552	**-8232.414**	-8265.104	-8253.008
0	2	0	0.3785	1	4141.197	-8230.256	**-8268.394**	**-8254.282**

Note: Boldface indicates that the model was the best, among all combinations of $p_1, q_1 \in \{0, 1, 2, 3\}$ and $p_2, q_2 \in \{0, 1, 2\}$, with respect to the corresponding criterion.

As shown in Table 8, the values of the selection criteria did not vary much amongst the tested models so we choose the most parsimonious one, namely, ARMA(0,1)-FIEGARCH$(0, d, 1)$. We compare the forecasting performance of this model with other ARCH-type models and with a radial basis function model (a detailed description of this approach is given in the sequel). For this comparison the order of the ARMA(p_1, q_1) part of the model was not changed, that is, we fixed $p_1 = 0$ and $q_1 = 1$ for all ARCH-type models. The EGARCH(p_2, q_2) model was set to have the same values for p_2 and q_2 as the FIEGARCH model so we could investigate the influence of the long-range dependence parameter d. For the GARCH(p_2, q_2) model we choose the smallest values of p_2 and q_2 for which the residuals of the model are not correlated. The same was done for the ARCH(p_2) model (which resulted in $p_2 = 6$). The ARCH(1) model was presented only for comparison. The estimated coefficients for the ARCH-type models are given in Table 9, with the corresponding log-likelihood value. Notice that, the FIEGARCH model fitted to this time series present the same parameter values as model M4 considered in the Monte Carlo simulation study in Section 4.

In what follows we consider the so-called radial basis function network model to estimate the conditional mean and the conditional volatility of $\{r_t\}_{t\in\mathbb{Z}}$ (the same approach is considered, for instance, in [32] and [33]). We recall that the radial basis function network is an artificial neural network that uses radial basis functions as activation functions. To fit a radial basis model to the data (no exogenous variables are considered) we

Table 9: Fitted models and their respective log-likelihood, BIC, AIC and HQC values. The number in parenthesis corresponds to the standard error of the estimate.

Estimate	ARMA(0,1) + ARCH(1)	ARMA(0,1) + ARCH(6)	ARMA(0,1) + GARCH(1,1)	ARMA(0,1) + EGARCH(0,1)	ARMA(0,1) + FIEGARCH(0,d,1)
$\hat{\theta}_1$	-0.1138 (0.0200)	-0.0642 (0.0267)	-0.0647 (0.0266)	-0.0751 (0.0254)	-0.0776 (0.0257)
$\hat{\omega}$	0.0004 (0.0000)	0.0002 (0.0000)	0.0000 (0.0000)	-7.4694 (0.0969)	-7.2247 (0.2143)
$\hat{\alpha}_1$	0.6071 (0.0581)	0.2307 (0.0417)	0.2019 (0.0247)	-	-
$\hat{\alpha}_2$	-	0.1540 (0.0333)	-	-	-
$\hat{\alpha}_3$	-	0.1852 (0.0390)	-	-	-
$\hat{\alpha}_4$	-	0.1145 (0.0348)	-	-	-
$\hat{\alpha}_5$	-	0.0641 (0.0290)	-	-	-
$\hat{\alpha}_6$	-	0.0635 (0.0257)	-	-	-
$\hat{\beta}_1$	-	-	0.7659 (0.0271)	0.9373 (0.0103)	0.6860 (0.0986)
$\hat{d}$	-	-	-	-	0.3578 (0.0810)
$\hat{\theta}$	-	-	-	-0.1653 (0.0197)	-0.1661 (0.0224)
$\hat{\gamma}$	-	-	-	0.2782 (0.0300)	0.2972 (0.0332)
log-likelihood	3934.337	4060.372	4072.622	4137.625	4138.552
BIC	-7846.329	-8061.157	-8115.451	-8238.008	-8232.414
AIC	-7862.674	-8104.744	-8137.244	-8265.250	-8265.104
HQC	-7856.626	-8088.616	-8129.180	-8255.170	-8253.008

assume that $\{r_t\}_{t\in\mathbb{Z}}$ can be written as

$$r_t = \phi(\boldsymbol{y}_{t-1}) + \psi(\boldsymbol{y}_{t-1})Z_t := \phi(\boldsymbol{y}_{t-1}) + \varepsilon_t, \quad \text{for all } t \in \mathbb{Z},$$

with $\boldsymbol{y}_{t-1} = (r_{t-1}, \cdots, r_{t-p})$, for some $p > 0$, $\varepsilon_t := \psi(\boldsymbol{y}_{t-1})Z_t$, $\mathbb{E}(Z_t) = 0$ and $\mathbb{E}(Z_t^2) = 1$. Under these assumptions, $\mathbb{E}(r_t|\boldsymbol{y}_{t-1}) = \phi(\boldsymbol{y}_{t-1})$ and $\mathbb{E}(\varepsilon_t^2|\boldsymbol{y}_{t-1}) = \psi^2(\boldsymbol{y}_{t-1})$, for all $t \in \mathbb{Z}$. Therefore, we use neural networks Φ_n and $\boldsymbol{\Psi}_n$ to approximate, respectively, $\phi(\boldsymbol{y})$ and $\psi^2(\boldsymbol{y})$, and to obtain

$$\hat{\phi}(\boldsymbol{y}) = \Phi_n(\boldsymbol{y}; \hat{\boldsymbol{w}}_1) \quad \text{and} \quad \hat{\psi}^2(\boldsymbol{y}) = \Psi_n(\boldsymbol{y}; \hat{\boldsymbol{w}}_2), \quad \text{for all } \boldsymbol{y} \in \mathbb{R}^p,$$

where

$$\hat{\boldsymbol{w}}_1 = \arg\min\bigg\{\frac{1}{n-p}\sum_{t=p+1}^{n}\Big[r_t - \Phi_n(\boldsymbol{y}_{t-1}; \boldsymbol{w})\Big]^2\bigg\} \quad \text{and} \quad \hat{\boldsymbol{w}}_2 = \arg\min\bigg\{\frac{1}{n-p}\sum_{t=p+1}^{n}\Big[\hat{\varepsilon}_t^2 - \Psi_n(\boldsymbol{y}_{t-1}; \boldsymbol{w})\Big]^2\bigg\},$$

with $\hat{\varepsilon}_t = r_t - \hat{\phi}(\boldsymbol{y}_{t-1})$, for all $t \in \mathbb{Z}$. In both cases, we consider one hidden layer containing J neurons, for some $J \in \mathbb{N}$, that is,

$$\Phi_n(\boldsymbol{y}; \boldsymbol{w}_1) = \sum_{i=1}^{J} a_i\rho_i(||\boldsymbol{y} - \boldsymbol{c}_i||) \quad \text{and} \quad \Psi_n(\boldsymbol{y}; \boldsymbol{w}_2) = \sum_{i=1}^{J} a_i^*\rho_i^*(||\boldsymbol{y} - \boldsymbol{c}_i^*||), \quad \text{for all } \boldsymbol{y} \in \mathbb{R}^p,$$

with $\boldsymbol{w}_1 = (a_1, \cdots, a_J, b_1, \cdots, b_J, \boldsymbol{c}_1, \cdots, \boldsymbol{c}_J)$, $\boldsymbol{w}_2 = (a_1^*, \cdots, a_J^*, b_1^*, \cdots, b_J^*, \boldsymbol{c}_1^*, \cdots, \boldsymbol{c}_J^*)$, $a_i, b_i, a_i^*, b_i^* \in \mathbb{R}$, $\boldsymbol{c}_i, \boldsymbol{c}_i^* \in \mathbb{R}^p$, $||\cdot||$ the Euclidean norm, $\rho_i(z) = e^{-(b_i z)^2}$ and $\rho_i^*(z) = e^{-(b_i^* z)^2}$, for any $z \in \mathbb{R}$ and $i \in \{1, \cdots, J\}$.

To obtain a h-step ahead predictor for r_{n+h}^2 given $\{r_t\}_{t=1}^n$, we observe that, for all $t \in \mathbb{Z}$,

$$\mathbb{E}\big(r_t|\{r_k\}_{k<t}\big) = \mathbb{E}(r_t|\boldsymbol{y}_{t-1}) = \phi(\boldsymbol{y}_{t-1}) \quad \text{and} \quad \mathrm{Var}\big(r_t|\{r_k\}_{k<t}\big) = \mathrm{Var}(r_t|\boldsymbol{y}_{t-1}) = \psi^2(\boldsymbol{y}_{t-1}).$$

Therefore, $\mathbb{E}\big(r_t^2|\{r_k\}_{k<t}\big) = \mathbb{E}(r_t^2|\boldsymbol{y}_{t-1}) = \varphi(\boldsymbol{y}_{t-1}) = \psi^2(\boldsymbol{y}_{t-1}) + \phi^2(\boldsymbol{y}_{t-1})$, for some $\varphi : \mathbb{R}^p \to \mathbb{R}^p$. Thus, once $\phi(\cdot)$ and $\psi^2(\cdot)$ are estimated, the predictors $\hat{r}_{n+h}$ and $\hat{r}_{n+h}^2$ can be obtained recursively as

$$\begin{aligned} \hat{r}_{n+1} &= \hat{\phi}(\boldsymbol{y}_n) \quad \text{and} \quad \hat{r}_{n+1}^2 = \hat{\psi}^2(\boldsymbol{y}_n) + \hat{\phi}^2(\boldsymbol{y}_n), \\ \hat{r}_{n+h} &= \hat{\phi}(\hat{\boldsymbol{y}}_{n+h-1}) \quad \text{and} \quad \hat{r}_{n+h}^2 = \hat{\psi}^2(\hat{\boldsymbol{y}}_{n+h-1}) + \hat{\phi}^2(\hat{\boldsymbol{y}}_{n+h-1}), \quad \text{for all } h > 1, \end{aligned}$$

where $\hat{\boldsymbol{y}}_{n+h-1} = (\hat{r}_{n+h-1}, \cdots, \hat{r}_{n+h-1-p})$, with $\hat{r}_{n+h-1-k} = r_{n+h-1-k}$, whenever $n+h-1-k \leq n$.

Tables 10 and 11 present some statistics to access the out-of-sample forecasting performance, respectively, of ARCH-type and radial basis models. The values in these tables correspond to the mean absolute error (mae), the mean percentage error (mpe) and the maximum absolute error (max_{ae}) of forecast, respectively defined as

$$mae = \frac{1}{20}\sum_{h=1}^{20}|e_{n+h}|, \quad mpe := \frac{1}{20}\sum_{h=1}^{20}\frac{|e_{n+h}|}{r^2_{n+h}} \quad \text{and} \quad max_{ae} := \max_{h\in\{1,\cdots,20\}}\{|e_{n+h}|\},$$

where $e_{n+h} := \hat{r}^2_{n+h} - r^2_{n+h}$, for $h \in \{1, \cdots, 20\}$ and $n = 1717$, is the h-step ahead forecast error. Note that, when considering the ARMA model combined with ARCH-type models, the ARMA(0,1) part of the models implies that $r_t = X_t - \theta_1 X_{t-1}$, where $X_t = \sigma_t Z_t$, for all $t \in \mathbb{Z}$. Since we define $\hat{r}^2_{t+h} = \mathbb{E}(r^2_{t+h}|\mathcal{F}_t)$ and σ^2_t is $\mathcal{F}_{t-1}$-measurable, for all $t \in \mathbb{Z}$, by elementary calculations we conclude that, $\hat{r}^2_{n+1} = \sigma^2_{n+1} + \theta_1^2 X_n^2$ and $\hat{r}^2_{n+h} = \hat{\sigma}^2_{n+h} + \theta_1^2\hat{\sigma}^2_{n+h-1}$, for all $h > 1$, with $\hat{\sigma}^2_{n+1} = \sigma^2_{n+1}$. For EGARCH and FIEGARCH models, $\hat{\sigma}^2_{n+1}$ is replaced by $\tilde{\sigma}^2_{n+1}$, given in expression (28), and $\check{\sigma}^2_{n+h} := \exp\{\hat{\ln}(\sigma^2_{n+h})\}$, where $\hat{\ln}(\sigma^2_{n+h})$ is defined in Proposition 4.

Table 10: Mean absolute error (mae), mean percentage error (mpe) and maximum absolute error (max_{ae}) of forecasting for the models in Table 9.

Model	ARMA(0,1) + ARCH(1)	ARMA(0,1) + ARCH(6)	ARMA(0,1) + GARCH(1,1)	ARMA(0,1) + EGARCH(1,1)		ARMA(0,1) + FIEGARCH(1,d,1)	
Predictor	$\hat{\sigma}^2_{t+h}$	$\hat{\sigma}^2_{t+h}$	$\hat{\sigma}^2_{t+h}$	$\tilde{\sigma}^2_{t+h}$	$\check{\sigma}^2_{t+h}$	$\tilde{\sigma}^2_{t+h}$	$\check{\sigma}^2_{t+h}$
mae	0.00053	0.00045	0.00043	0.00045	0.00044	0.00045	0.00043
mpe	109.40844	68.97817	60.29677	71.33057	61.26625	68.42884	59.88066
max_{ae}	0.00094	0.00094	0.00094	0.00082	0.00087	0.00084	0.00088

Note: The high mpe values are due to 5 observations close to zero.

From Table 10 we conclude that, given its high mpe value, the ARMA(0,1)-ARCH(1) does not fit the data well. In fact, the squared residuals from this model are still correlated and we use the model only for comparison. The ARMA(0,1)-ARCH(6) model performs similarly to the ARMA(0,1)-GARCH(1,1) model, in terms of both, mae and max_{ae} values, presenting a higher mpe value. However, the latter is more parsimonious. Although the log-likelihood value is higher (and the max_{ae} value is smaller) for the ARMA(0,1)-EGARCH(0,1) model, the mae and the mpe values are smaller for the ARMA(0,1)-GARCH(0,d,1) model. Overall, the ARMA(0,1)-FIEGARCH(0,d,1) performs slightly better than the other models.

The fact that all models present a similar perfomance confirms the following, already known in the literature.

- In practice, ARCH(p) models perform relatively well for most applications.
- GARCH(p, q) models are more parsimonious than the ARCH ones. For instance, notice that similar results were obtained here by considering an ARCH(6) model and a GARCH$(1, 1)$ model.
- For EGARCH(p, q) models the conditional variance is defined in terms of the logarithm function and less (usually none) restrictions have to be imposed during parameter estimation. Moreover, EGARCH models are not necessarily more parsimonious than ARCH/GARCH ones since they also carry information on the returns' asymmetry (θ and γ parameters).
- FIEGARCH(p, d, q) models can describe not only the same characteristics as ARCH, GARCH and EGARCH models do, but also the long-range dependence in the volatility. Also, the performance of all models will be very similar if the volatility presents high persistence. For instance, notice that for the ARCH(6) model $\alpha_1 + \cdots + \alpha_6 = 0.812$, for the GARCH$(1, 1)$ model $\alpha_1 + \beta_1 = 0.9678$ and for the EGARCH model $\beta_1 = 0.9373$, which implies high persistence in the volatility. Moreover, for the

FIEGARCH model, we found $d = 0.3578$ with standard error equal to 0.0810, which indicates that the parameter d is statistically different from zero and thus, there is evidence of long-range dependence in the volatility.

- Given their definition, it is expected that EGARCH and FIEGARCH models will provide better forecasts for $\ln(\sigma^2_{t+h})$ than for σ^2_{t+h} and, consequently, for X^2_{t+h}.

Table 11: Mean absolute error (mae), mean percentage error (mpe) and maximum absolute error (max_{ae}) of forecasting for radial basis models with $J \in \{5, 10, \cdots, 45\}$ hidden neurons and $p \in \{1, 5, 10, 15\}$.

p	J	mae	mpe	max_{ae}	p	J	mae	mpe	max_{ae}
1	5	0.00189	168.16694	0.00276	10	5	0.00046	84.07916	0.00096
	10	0.00464	360.57740	0.02105		10	0.00209	211.49929	0.00288
	15	0.00306	205.95363	0.01798		15	0.00076	40.16931	0.00156
	20	0.00284	405.17466	0.00406		20	0.00251	353.29510	0.00329
	25	0.00106	69.24385	0.00193		25	0.00099	65.04972	0.00177
	30	0.00077	35.08914	0.00165		30	0.00214	309.03589	0.00292
	35	0.00117	81.84698	0.00204		35	0.00047	60.11370	0.00083
	40	0.00082	40.86115	0.00169		40	0.00224	214.27183	0.00302
	45	0.00044	7.76332	0.00130		45	0.00043	46.54092	0.00084
5	5	**0.00040**	49.60723	0.00090	15	5	**0.00040**	20.88682	0.00111
	10	0.00050	92.13256	0.00092		10	0.00063	42.05418	0.00164
	15	0.00058	111.93650	0.00109		15	0.00110	185.41861	0.00212
	20	0.00040	21.32100	0.00116		20	0.00277	326.16372	0.00378
	25	0.00052	5.95880	0.00138		25	0.00045	63.80141	0.00082
	30	0.00046	4.61686	0.00129		30	0.00047	**3.95304**	0.00123
	35	0.00041	19.79905	0.00116		35	0.00045	72.19310	0.00098
	40	**0.00040**	31.93826	0.00107		40	0.00044	63.04763	**0.00079**
	45	0.00120	88.07146	0.00207		45	0.00271	363.47039	0.00343

Note: Boldface indicates the best model for each criterion.

From Table 11 we observe that

- in terms of mae or max_{ae}, both radial basis and ARCH-type (see Table 10) models have a similar performance. In this case, ARCH-type models seem a better choice given the smaller number of parameter to be estimated;

- for each p there exists at least one J for which the mpe value for the radial basis model is much smaller then for any ARCH-type models. However, given the similarity regarding mae, the small mpe values only indicate that radial basis models provide a better forecast for observations that are close to zero.

6. Conclusions

Here we presented complete mathematical proofs for the stationarity, the ergodicity, the conditions for the causality and invertibility properties, the autocorrelation and spectral density functions' decay and the convergence order for the polynomial coefficients that describe the volatility for any FIEGARCH(p, q, d) process. We proved that if $\{X_t\}_{t\in\mathbb{Z}}$ is a FIEGARCH(p, d, q) process and $\mathbb{E}(\big[\ln(Z_0^2)\big]^2) < \infty$, then $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is an ARFIMA$(q, d, 0)$ process with correlated innovations. Expressions for the kurtosis and the asymmetry measures of any stationary FIEGARCH(p, d, q) process were also provided.

We also proved that if $\{X_t\}_{t\in\mathbb{Z}}$ is a FIEGARCH(p, d, q) process then it is a martingale difference with respect to the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$, where $\mathcal{F}_t := \sigma(\{Z_s\}_{s\leq t})$. The h-step ahead forecast for the processes $\{X_t\}_{t\in\mathbb{Z}}$, $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ and $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ were given with their respective mean square error of forecast. Since $\mathbb{E}(\sigma^2_{t+h}|\mathcal{F}_t)$ cannot be easily calculated for FIEGARCH models, we also discussed some alternative estimators for the h-step ahead forecast of σ^2_{t+h}, for all $h > 0$.

We presented a Monte Carlo simulation study showing how to perform the generation, the estimation and the forecasting of six different FIEGARCH models. The parameter selection of these six models was related to the real time series analyzed in [11]. Parameter estimation was performed by considering the well known quasi-likelihood method. We concluded that, given the complexity of FIEGARCH models, the quasi-likelihood method performs relatively well, which was indicated by the small *bias*, *mae* and *mse* values for the estimates. Regarding the h-step ahead forecast for the processes $\{\sigma_t^2\}_{t\in\mathbb{Z}}$ and $\{X_t^2\}_{t\in\mathbb{Z}}$, we observed that the mean square error of forecast decreases as the sample size increases. However, while the conditional variance was well estimated, which was indicated by the small *mae* values, the estimator $\tilde{X}_{n+h}^2 := \tilde{\sigma}_{t+h}^2$, which is an approximation for $\hat{X}_{t+h}^2 := \mathbb{E}(X_{n+h}^2|\mathcal{F}_n) = \hat{\sigma}_{n+h}^2$, did not perform well in predicting X_{n+h}^2. This result was expected since the purpose of the model is to forecast the logarithm of the conditional variance and not the process $\{X_t\}_{t\in\mathbb{Z}}$ itself.

Finally, we presented the analysis of the São Paulo Stock Exchange Index (Bovespa Index or IBovespa) log-return time series. We compared the forecasting performance of FIEGARCH models, fully described in this paper, with other ARCH-type models. All models presented a similar performance which was attributed to the fact that the ARCH, GARCH and EGARCH models indicated high persistence in the volatility. We also compared the forecasting performance of ARCH-type with radial basis models. Given the similarity regarding the mean (and maximum) absolute error of forecast we concluded that both classes show a similar forecasting performance. Comparing the mean percentage error of forecasts we concluded that radial basis models provide a better forecast for observations which are close to zero.

Acknowledgments

S.R.C. Lopes was partially supported by CNPq-Brazil, by CAPES-Brazil, by INCT em *Matemática* and by Pronex *Probabilidade e Processos Estocásticos* - E-26/170.008/2008 -APQ1. T.S. Prass was supported by CNPq-Brazil. The authors are grateful to the (Brazilian) National Center of Super Computing (CESUP-UFRGS) for the computational resources.

Appendix: Proofs

In this section we provide the proofs of all propositions, lemmas, corollaries and theorems stated in Sections 2 and 3, in the same order as they appear in the text.

Proof of Proposition 1:
See [10]. □

Proof of Theorem 1:
See theorem 2.1 in [4]. □

Proof of Theorem 2:
See theorem 2.2 in [4]. □

Proof of Theorem 3:
Denote $\beta(z)^{-1}$ by $f(z)$. Since $\beta(\cdot)$ has no roots in the closed disk $\{z : |z| \leq 1\}$, one has

$$\beta(z)^{-1} := f(z) = \sum_{k=0}^{\infty} f_k z^k, \quad \text{where } f_k = \frac{f^{(k)}(0)}{k!}, \text{ for all } k \in \mathbb{N}. \tag{A.1}$$

From expressions (7), (8) and (A.1) it follows that

$$\lambda(z) = \sum_{k=0}^{\infty} \Bigg[\sum_{i=0}^{\min\{p,k\}} (-\alpha_i) \sum_{j=0}^{k-i} \pi_{d,k-i-j} f_j \Bigg] z^k. \tag{A.2}$$

From (A.2), one has

$$\lambda_{d,k} = \sum_{i=0}^{\min\{p,k\}} (-\alpha_i) \sum_{j=0}^{k-i} \pi_{d,k-i-j} f_j, \quad \text{for all } k \in \mathbb{N}.$$

In particular, $\lambda_{d,k} = \sum_{i=0}^{p}(-\alpha_i) \sum_{j=0}^{k-i} \pi_{d,j} f_{k-i-j}$, for all $k > p$.

Moreover, since $f_k \to 0$, as $k \to \infty$, it follows that for all $\varepsilon > 0$, there exists $k_0 > 0$, such that, for a given $m > 0$ and for all $k > k_0$, $|\pi_{d,j} f_{k-i-j}| < \frac{\varepsilon}{m}$, for all $0 \le j \le m$ and $0 \le i \le p$. Hence, for k sufficiently large,

$$\lambda_{d,k} \sim \sum_{i=0}^{p}(-\alpha_i) \sum_{j=m+1}^{k-i} \pi_{d,j} f_{k-i-j}.$$

Notice that, since $\pi_{d,k} \sim \frac{1}{\Gamma(d)\, k^{1-d}}$, as $k \to \infty$, one can choose m_0 such that $\pi_{d,k} \sim \pi_{d,k-i} \sim \pi_{d,j}$, for all $m_0 < m+1 \le j \le k-i$ and $0 \le i \le p$. Consequently,

$$\lambda_{d,k} \sim \pi_{d,k} \sum_{i=0}^{p}(-\alpha_i) \sum_{j=0}^{k-i-(m+1)} f_j \sim \pi_{d,k} \sum_{i=0}^{p}(-\alpha_i) \sum_{j=0}^{\infty} f_j.$$

However, $\sum_{j=0}^{\infty} f_j = f(1) = \dfrac{1}{\beta(1)}$ and $\pi_{d,k} \sim \frac{1}{\Gamma(d)\, k^{1-d}}$, as $k \to \infty$. So, we have

$$\lambda_{d,k} \sim \pi_{d,k} \frac{\alpha(1)}{\beta(1)} \sim \frac{1}{\Gamma(d)\, k^{1-d}} \frac{\alpha(1)}{\beta(1)}.$$

It follows that $\lambda_{d,k} \to 0$ and $\lambda_{d,k} k^{1-d} \to \frac{1}{\Gamma(d)} \frac{\alpha(1)}{\beta(1)}$, as $k \to \infty$. Hence, $\lambda_{d,k} = O(k^{d-1})$, as $k \to \infty$, which concludes the proof. □

Proof of Proposition 2:

Let $\lambda(\cdot)$ be defined by (8). Consequently,

$$\alpha(z) = \beta(z)(1-z)^d \sum_{k=0}^{\infty} \lambda_{d,k} z^k. \tag{A.3}$$

By defining β_k^* as in expression (17), for all $k \in \mathbb{N}$, and upon considering expression (6), observing that $\delta_{d,0} = -1 = \beta_0$, the right hand side of expression (A.3) can be rewritten as

$$\begin{aligned}\beta(z)(1-z)^d \sum_{k=0}^{\infty} \lambda_{d,k} z^k &= \left[\sum_{k=0}^{\infty} \left(\sum_{j=0}^{k} -\beta_j^* \delta_{d,k-j}\right) z^k\right] \sum_{k=0}^{\infty} \lambda_{d,k} z^k = \sum_{k=0}^{\infty} \left[\sum_{i=0}^{k} \lambda_{d,i} \left(-\sum_{j=0}^{k-i} \beta_j^* \delta_{d,k-i-j}\right)\right] z^k \\ &= \sum_{k=0}^{\infty} \left[\lambda_{d,k} - \sum_{i=0}^{k-1} \lambda_{d,i} \sum_{j=0}^{k-i} \beta_j^* \delta_{d,k-i-j}\right] z^k. \end{aligned} \tag{A.4}$$

Now, by setting α_k^* as in expression (17), for all $k \in \mathbb{N}$, from expression (A.4) one concludes that the equality (A.3) holds if and only if,

$$-\alpha_0^* = \lambda_{d,0} \quad \text{and} \quad -\alpha_k^* = \lambda_{d,k} - \sum_{i=0}^{k-1} \lambda_{d,i} \sum_{j=0}^{k-i} \delta_{d,k-i-j} \beta_j^*, \quad \text{for all } k \ge 1.$$

Therefore, expression (16) holds. It is easy to see that by replacing the coefficients $\lambda_{d,k}$, given by (16), in the expression (A.4), for all $k \in \mathbb{N}$, we get $\sum_{k=0}^{\infty}(-\alpha_k^*) z^k = \alpha(z)$, which completes the proof. □

Proof of Corollary 1:

Let $\{\lambda_{d,k}\}_{k\in\mathbb{N}}$ be defined by (8) and rewrite (5) as (9). Observe that, by Theorem 3, the condition $d < 0.5$ implies that $\sum_{k=0}^{\infty} \lambda_{d,k}^2 < \infty$. Therefore, the results follow from Theorem 1 by taking $\omega_t := \omega$, for all $t \in \mathbb{Z}$, and $\lambda_k := \lambda_{d,k-1}$, for all $k \geq 1$. $\square$

Proof of Corollary 2:

Let $\{\lambda_{d,k}\}_{k\in\mathbb{N}}$ be defined by (8) and rewrite (5) as (9). Define $\omega_t := \omega$, for all $t \in \mathbb{Z}$, and $\lambda_k := \lambda_{d,k-1}$, for all $k \geq 1$. Observe that, from Theorem 3, $d < 0.5$ implies $\sum_{k=1}^{\infty} \lambda_k^2 < \infty$. Therefore, the assumptions of Theorem 2 hold and the results follow. $\square$

Proof of Proposition 3:

Let $\{X_t\}_{t\in\mathbb{Z}}$ be any stationary FIEGARCH(p,d,q) process and $\lambda(\cdot)$ be the polynomial defined by (8). Notice that, since $\{g(Z_t)\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables, from (5) it follows that

$$\mathbb{E}(\sigma_0^r) = e^{\frac{r\omega}{2}} \prod_{k=0}^{\infty} \mathbb{E}\Big(\exp\Big\{\frac{r}{2}\lambda_{d,k} g(Z_0)\Big\}\Big), \quad \text{for all } r > 0. \tag{A.5}$$

From the fact that σ_t and Z_t are independent random variables one has

$$\mathbb{E}(|X_t|^r) = \mathbb{E}(|X_0|^r) = \mathbb{E}(|Z_0|^r)\mathbb{E}(|\sigma_0|^r), \quad \text{for all } t \in \mathbb{Z} \text{ and } r > 0.$$

Thus, given $r > 0$, $\mathbb{E}(X_0^r) < \infty$ if and only if $\mathbb{E}(\sigma_0^r)$ and $\mathbb{E}(Z_0^r)$ are both finite. Therefore, if $\mathbb{E}(X_0^3) < \infty$ (analogously, $\mathbb{E}(X_0^4) < \infty$), the asymmetry (analogously, the kurtosis) measure exists, and expression (A.5) converges, for any $r \leq 3$ (analogously, $r \leq 4$). Upon replacing (A.5) in (18) we conclude the proof. $\square$

Proof of Theorem 4:

Assume that $\mathbb{E}([\ln(Z_0^2)]^2) < \infty$ and $d < 0.5$. Let $\{\lambda_{d,k}\}_{k\in\mathbb{Z}}$ be given by (8) and rewrite (5) as (9).

Observe that $\mathbb{E}([\ln(Z_0^2)]^2) < \infty$ implies $\mathbb{E}(|\ln(Z_0^2)|) < \infty$ and thus $|\ln(Z_t^2)|$ is finite with probability one, for all $t \in \mathbb{Z}$. Since $d < 0.5$, it follows that $\ln(\sigma_t^2)$ is finite with probability one, for all $t \in \mathbb{Z}$ (see Corollary 1). Therefore, $\ln(X_t^2)$ is finite with probability one, for all $t \in \mathbb{Z}$, and hence the stochastic process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is well defined. The strict stationarity and ergodicity of $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ follow immediately from the measurability of $\ln(Z_t^2) + \ln(\sigma_t^2)$ and the i.i.d. property of $\{Z_t\}_{t\in\mathbb{Z}}$ (see [28]). To prove that $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is also weakly stationary notice that $\mathbb{E}([\ln(Z_0^2)]^2) < \infty$ implies $\mathrm{Var}(\ln(Z_t^2)) < \infty$, $d < 0.5$ implies $\mathrm{Var}(\ln(\sigma_t^2)) < \infty$ (see Corollary 1) and the independence of $\{Z_t\}_{t\in\mathbb{Z}}$ implies that $\ln(Z_t^2)$ and $\ln(\sigma_t^2)$ are independent random variables. Hence,

$$\mathrm{Var}(\ln(X_t^2)) = \mathrm{Var}(\ln(\sigma_t^2)) + \mathrm{Var}(\ln(Z_t^2)) < \infty, \quad \text{for all} \quad t \in \mathbb{Z}.$$

To complete the proof it remains to show that the autocovariance function $\gamma_{\ln(X^2)}(h)$, for all $h \in \mathbb{Z}$, is given by expression (19). From the definition of $\ln(X_t^2)$, it follows that

$$\begin{aligned}\mathrm{Cov}\big(\ln(X_{t+h}^2), \ln(X_t^2)\big) = {} & \mathrm{Cov}\big(\ln(\sigma_{t+h}^2), \ln(\sigma_t^2)\big) + \mathrm{Cov}\big(\ln(Z_{t+h}^2), \ln(Z_t^2)\big) \\ & + \mathrm{Cov}\big(\ln(\sigma_{t+h}^2), \ln(Z_t^2)\big) + \mathrm{Cov}\big(\ln(Z_{t+h}^2), \ln(\sigma_t^2)\big).\end{aligned} \tag{A.6}$$

Theorem 1 shows that

$$\mathrm{Cov}\big(\ln(\sigma_{t+h}^2), \ln(\sigma_t^2)\big) = \sigma_g^2 \sum_{k=0}^{\infty} \lambda_{d,k}\lambda_{d,k+|h|}, \quad \text{for all } h \in \mathbb{Z}.$$

From the independence of the random variables $\ln(Z_t^2)$, for all $t \in \mathbb{Z}$, and from expression (9), we have

$$\mathrm{Cov}\big(\ln(Z_{t+h}^2), \ln(Z_t^2)\big) = \begin{cases} 0, & \text{if } h \neq 0; \\ \mathrm{Var}\big(\ln(Z_t^2)\big), & \text{if } h = 0 \end{cases} \quad \text{and} \quad \mathrm{Cov}\big(\ln(\sigma_{t+h}^2), \ln(Z_t^2)\big) = \begin{cases} \lambda_{d,h-1}\mathcal{K}, & \text{if } h > 0; \\ 0, & \text{if } h \leq 0. \end{cases}$$

where $\mathcal{K} = \mathrm{Cov}\big(g(Z_0), \ln(Z_0^2)\big)$. Now, since

(a) when $h < 0$,

$$\mathrm{Cov}\big(\ln(\sigma_{t+h}^2), \ln(Z_t^2)\big) = \mathrm{Cov}\big(\ln(\sigma_{t-|h|}^2), \ln(Z_t^2)\big) = 0$$

and, by setting $u = t + h$,

$$\mathrm{Cov}\big(\ln(Z_{t+h}^2), \ln(\sigma_t^2)\big) = \mathrm{Cov}\big(\ln(Z_u^2), \ln(\sigma_{u+|h|}^2)\big) = \lambda_{d,|h|-1}\mathcal{K}\mathbb{I}_{\mathbb{Z}^*}(h).$$

(b) when $h > 0$

$$\mathrm{Cov}\big(\ln(\sigma_{t+h}^2), \ln(Z_t^2)\big) = \lambda_{d,h-1}\mathcal{K}\mathbb{I}_{\mathbb{Z}^*}(h) = \lambda_{d,|h|-1}\mathcal{K}\mathbb{I}_{\mathbb{Z}^*}(h)$$

and, by setting $u = t + h$,

$$\mathrm{Cov}\big(\ln(Z_{t+h}^2), \ln(\sigma_t^2)\big) = \mathrm{Cov}\big(\ln(Z_u^2), \ln(\sigma_{u-|h|}^2)\big) = 0$$

one concludes that

$$\mathrm{Cov}\big(\ln(\sigma_{t+h}^2), \ln(Z_t^2)\big) + \mathrm{Cov}\big(\ln(Z_{t+h}^2), \ln(\sigma_t^2)\big) = \mathrm{Cov}\big(\ln(\sigma_{t+|h|}^2), \ln(Z_t^2)\big) = \lambda_{d,|h|-1}\mathcal{K}\mathbb{I}_{\mathbb{Z}^*}(h).$$

By replacing these results on expression (A.6) we conclude that the autocovariance function of $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is given by (19). □

Proof of Theorem 5:

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a FIEGARCH process. From expressions (1) and (5) we have

$$\beta(\mathcal{B})(1-\mathcal{B})^d(\ln(X_t^2) - \omega) = \varepsilon_t, \quad \text{for all } t \in \mathbb{Z},$$

where

$$\varepsilon_t = \alpha(\mathcal{B})g(Z_{t-1}) + \beta(\mathcal{B})(1-\mathcal{B})^d \ln(Z_t^2), \quad \text{for all } t \in \mathbb{Z}.$$

In particular, if $d > 0$, we have $\beta(\mathcal{B})(1-\mathcal{B})^d\omega = 0$ and $\beta(\mathcal{B})(1-\mathcal{B})^d \ln(X_t^2) = \varepsilon_t$, for all $t \in \mathbb{Z}$.

Now, suppose that $\mathbb{E}([\ln(Z_0^2)]^2) < \infty$. Since $\{Z_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables and $0 \leq |\mathbb{E}(\ln(Z_0^2))| \leq \mathbb{E}(|\ln(Z_0^2)|) \leq [\mathbb{E}([\ln(Z_0^2)]^2)]^{1/2}$, one concludes that $\mathbb{E}(\ln(Z_t^2)) = \mathbb{E}(\ln(Z_0^2)) < \infty$, for all $t \in \mathbb{Z}$. Therefore, $\beta(\mathcal{B})(1-\mathcal{B})^d\mathbb{E}(\ln(Z_t^2)) = 0$ and $\alpha(\mathcal{B})\mathbb{E}(g(Z_{t-1})) = 0$. Consequently, $\mathbb{E}(\varepsilon_t) = 0$, for all $t \in \mathbb{Z}$.

Let $\phi(\cdot)$ be defined by expression (22). Assume, for the moment, that $\mathrm{Var}(\varepsilon_t^2) < \infty$, for all $t \in \mathbb{Z}$. It follows that

$$\begin{aligned}\mathrm{Cov}(\varepsilon_{t+h}, \varepsilon_t) = &\,\mathrm{Cov}\big(\alpha(\mathcal{B})g(Z_{t+h-1}), \alpha(\mathcal{B})g(Z_{t-1})\big) + \mathrm{Cov}\big(\phi(\mathcal{B})\ln(Z_{t+h}^2), \phi(\mathcal{B})\ln(Z_t^2)\big) \\ &+ \mathrm{Cov}\big(\alpha(\mathcal{B})g(Z_{t+h-1}), \phi(\mathcal{B})\ln(Z_t^2)\big) + \mathrm{Cov}\big(\phi(\mathcal{B})\ln(Z_{t+h}^2), \alpha(\mathcal{B})g(Z_{t-1})\big). \qquad \text{(A.7)}\end{aligned}$$

Since $\{g(Z_t)\}_{t\in\mathbb{Z}}$ is a white noise process we have

$$\mathrm{Cov}\big(\alpha(\mathcal{B})g(Z_{t+h-1}), \alpha(\mathcal{B})g(Z_{t-1})\big) = \begin{cases} \mathrm{Var}\big(g(Z_0)\big)\displaystyle\sum_{i=|h|}^{p} \alpha_i\alpha_{i-|h|}, & \text{if} \quad |h| \leq p; \\ 0, & \text{if} \quad |h| > p, \end{cases}$$

which does not depend on $t \in \mathbb{Z}$. From the independence of the random variables Z_t, for all $t \in \mathbb{Z}$, one has

$$\mathrm{Cov}\big(\alpha(\mathcal{B})g(Z_{t+h-1}), \phi(\mathcal{B})\ln(Z_t^2)\big) = \begin{cases} \mathcal{K}\displaystyle\sum_{i=0}^{p} \alpha_i\phi_{i-h+1}, & \text{if} \quad h < 1; \\ \mathcal{K}\displaystyle\sum_{i=h-1}^{p} \alpha_i\phi_{i-h+1}, & \text{if} \quad 1 \leq h \leq p+1; \\ 0, & \text{if} \quad h > p+1 \end{cases}$$

and

$$\operatorname{Cov}\big(\phi(\mathcal{B})\ln(Z_{t+h}^2),\alpha(\mathcal{B})g(Z_{t-1})\big)=\begin{cases}0, & \text{if} \quad h<-(p+1);\\ \mathcal{K}\displaystyle\sum_{i=|h|-1}^{p}\alpha_i\phi_{i+h+1}, & \text{if} \quad -(p+1)\leq h\leq -1;\\ \mathcal{K}\displaystyle\sum_{i=0}^{p}\alpha_i\phi_{i+h+1}, & \text{if} \quad h>-1,\end{cases}$$

where $\mathcal{K}:=\operatorname{Cov}\big(g(Z_0),\ln(Z_0^2)\big)$ does not depend on $t\in\mathbb{Z}$. Also, from the independence of the random variables $\ln(Z_t^2)$, for all $t\in\mathbb{Z}$, we have

$$\operatorname{Cov}\big(\phi(\mathcal{B})\ln(Z_{t+h}^2),\phi(\mathcal{B})\ln(Z_t^2)\big)=\operatorname{Var}\big(\ln(Z_0^2)\big)\sum_{i=|h|}^{\infty}\phi_i\phi_{i-|h|},\quad \text{for all } h\in\mathbb{Z},$$

which does not depend on $t\in\mathbb{Z}$.

Therefore, all four terms in expression (A.7) do not depend on $t\in\mathbb{Z}$ and expression (21) holds. Now, to validate expression (21) we only need to show that $\operatorname{Var}(\varepsilon_t)<\infty$, for all $t\in\mathbb{Z}$. Notice that, since $\mathbb{E}(\varepsilon_t)=0$, it follows that $\mathbb{E}(\varepsilon_t^2)=\operatorname{Var}(\varepsilon_t)=\gamma_{\varepsilon}(0)$. Upon replacing $h=0$ in (A.7) one obtains

$$\gamma_{\varepsilon}(0)=\operatorname{Var}\big(g(Z_0)\big)\sum_{i=0}^{p}\alpha_i^2+2\mathcal{K}\sum_{i=0}^{p}\alpha_i\phi_{i+1}+\operatorname{Var}\big(\ln(Z_0^2)\big)\sum_{i=0}^{\infty}\phi_i^2.$$

By hypothesis, $\mathbb{E}([\ln(Z_0^2)]^2)<\infty$ and $d\in(-0.5,0.5)$. It follows that $\operatorname{Var}(\ln(Z_0^2))<\infty$ and $\sum_{i=0}^{\infty}\phi_i^2<\infty$. We also know that $\operatorname{Var}(g(Z_0))<\infty$. In order to show that $\mathcal{K}<\infty$, notice that $\mathcal{K}:=\operatorname{Cov}(g(Z_0),\ln(Z_0^2))=\mathbb{E}(g(Z_0)\ln(Z_0^2))$ and, since $\mathbb{E}(Z_0^2)=1$ and $\operatorname{Var}(\ln(Z_0^2))<\infty$, from Hölder's inequality, we have $\mathbb{E}(|Z_0|)<\infty$ and $\mathbb{E}(\ln(Z_0^2))<\infty$. Then from (2) it follows that

$$\mathbb{E}\big(g(Z_0)\ln(Z_0^2)\big)=\theta\,\mathbb{E}\big(Z_0\ln(Z_0^2)\big)+\gamma\mathbb{E}\big(|Z_0|\ln(Z_0^2)\big)-c,$$

where $c:=\gamma\,\mathbb{E}(|Z_0|)\mathbb{E}(\ln(Z_0^2))<\infty$. By using the fact that $2ab\leq a^2+b^2$, for all $a,b\in\mathbb{R}$, one concludes that

$$\big|\mathbb{E}(Z_t\ln(Z_t^2))\big|\leq\frac{1}{2}\left[\mathbb{E}(Z_t^2)+\mathbb{E}([\ln(Z_t^2)]^2)\right]<\infty\quad\text{and}\quad\big|\mathbb{E}(|Z_t|\ln(Z_t^2))\big|\leq\frac{1}{2}\left[\mathbb{E}(Z_t^2)+\mathbb{E}([\ln(Z_t^2)]^2)\right]<\infty.$$

Hence $\mathbb{E}(g(Z_0)\ln(Z_0^2))<\infty$ and, consequently, $\operatorname{Cov}(g(Z_0),\ln(Z_0^2))<\infty$ and $\gamma_{\varepsilon}(0)<\infty$. Therefore, the result follows. □

Proof of Lemma 1:

From definition, σ_t is a $\mathcal{F}_{t-1}$-measurable function. Moreover, for all $t\in\mathbb{Z}$, $\mathbb{E}(X_t)=\mathbb{E}(\mathbb{E}(X_t|\mathcal{F}_{t-1}))$ and $\mathbb{E}(X_t|\mathcal{F}_{t-1})=\mathbb{E}(\sigma_tZ_t|\mathcal{F}_{t-1})=\sigma_t\mathbb{E}(Z_t|\mathcal{F}_{t-1})=0$. Therefore, the process $\{X_t\}_{t\in\mathbb{Z}}$ is a martingale difference with respect to $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$. □

Proof of Lemma 2:

From Lemma 1, a FIEGARCH(p,d,q) process is a martingale difference. It follows that $\hat{X}_{n+h}=\mathbb{E}(X_{n+h}|\mathcal{F}_n)=0$, for all $h>0$. From definition, $\mathbb{E}(X_{n+1}^2|\mathcal{F}_n)=\sigma_{n+1}^2$. Therefore, the 1-step ahead forecast of X_{n+1}^2, given $\mathcal{F}_n$, is σ_{n+1}^2. Moreover, if $\mathbb{E}(X_t^4)<\infty$, for all $t\in\mathbb{Z}$, then this is the best forecast value in mean square error sense. □

Proof of Proposition 4:

Observe that $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is a causal ARFIMA(q,d,p) process and, if $\alpha(z)\neq 0$ in the closed disk $\{z:|z|\leq 1\}$, it is also invertible. Thus, the result follows by mimicking the proof of theorem 5.5.1 in [23]. □

Proof of Proposition 5:

By expression (25), $\check{\ln}(X^2_{n+h}) := \hat{\ln}(\sigma^2_{n+h})$, for all $h > 0$. Thus, from expression (1) and from Proposition 4, we have

$$\begin{aligned}\mathbb{E}\Big(\big[\ln(X^2_{n+h}) - \check{\ln}(X^2_{n+h})\big]^2\Big) &= \mathbb{E}\Big(\big[\ln(X^2_{n+h}) - \ln(\hat{\sigma}^2_{n+h})\big]^2\Big) = \mathbb{E}\Big(\big[\ln(\sigma^2_{n+h}) + \ln(Z^2_{n+h}) - \ln(\hat{\sigma}^2_{n+h})\big]^2\Big) \\ &= \mathbb{E}\bigg(\bigg[\sum_{k=0}^{h-2} \lambda_{d,k}\, g(Z_{n+h-1-k}) + \ln(Z^2_{n+h})\bigg]^2\bigg). \qquad \text{(A.8)}\end{aligned}$$

By expanding the right hand side of expression (A.8) and using the fact that $\{Z_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables, the proposition follows immediately. $\square$

References

[1] R.F. Engle, Autoregressive Conditional Heteroskedasticity with Estimates of Variance of U.K. Inflation, Econometrica, 50 (1982) 987-1008.

[2] F. Breidt, N. Crato, P.J.F. de Lima, On The Detection and Estimation of Long Memory in Stochastic Volatility, Journal of Econometrics, 83 (1998) 325-348.

[3] Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroskedasticity". *Journal of Econometrics*, vol. 31, 307-327.

[4] D.B. Nelson, Conditional Heteroskedasticity in Asset Returns: A New Approach, Econometrica, 59 (1991) 347-370.

[5] R. Baillie, T. Bollerslev, H. Mikkelsen, Fractionally Integrated Generalised Autoregressive Conditional Heteroscedasticity, Journal of Econometrics, 74 (1996) 3-30.

[6] T. Bollerslev, H.O. Mikkelsen, Modeling and Pricing Long Memory in Stock Market Volatility, Journal of Econometrics, 73 (1996) 151-184.

[7] S.A. Jayasuriya, A Sub Period Analysis of Long Memory in Stock Return Volatility for Emerging Markets, 9th Global Conference on Business and Economics Proceedings, Cambridge University, U.K., (2009) 1-31.

[8] E. Ruiz, H. Veiga, Modelling Long-Memory Volatilities with Leverage Effect: A-LMSV versus FIEGARCH, Computational Statistics and Data Analysis, 52(6) (2008) 2846-2862.

[9] S. Saadi, D. Gandhi, S. Dutta,Testing for Nonlinearity and Modeling Volatility in Emerging Capital Markets: The Case of Tunisia, International Journal of Theoretical and Applied Finance, 9(7) (2006) 1021-1050.

[10] T.S Prass, Análise e Estimação de Medidas de Risco em Processos FIEGARCH, Master Thesis in the Graduate Program in Mathematics, Federal University of Rio Grande do Sul, Porto Alegre, 2008.

[11] T.S. Prass, S.R.C. Lopes, Risk Measures Estimation on FIEGARCH Processes, arXiv:1305.5238 (2013).

[12] T.S. Prass, S.R.C. Lopes, VaR, Teste de Estresse e MaxLoss na Presença de Heteroscedasticidade e Longa Dependência na Volatilidade, Revista Brasileira de Estatística, (2013), Forthcoming.

[13] S.R.C. Lopes, Long-range Dependence in Mean and Volatility: Models, Estimation and Forecasting, in: V. Sidoravicius and M.E. Vares (eds.), In and Out of Equilibrium 2 (Progress in Probability), Birkhäuser, Boston, 2008, 60, pp. 497-525.

[14] W. Palma, Long-Memory Time Series: Theory and Methods, John Wiley, New Jersey, 2007.

[15] S. Lee, B. Hansen, Asymptotic Properties of the Maximum Likelihood Estimator and Test of the Stability on the GARCH and IGARCH Models, Econometric Theory, 10 (1994) 29-52.

[16] R. Lumsdaine, Asymptotic Properties of the Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) Models, Econometrica, 64 (1996) 575-596.

[17] I. Berkes, L. Horváth, P. Kokoszka, GARCH Processes: Structure and Estimation, Bernoulli, 9 (2003) 201-228.

[18] I. Berkes, L. Horváth, The Rate of Consistency of the Quasi-Maximum Likelihood Estimator, Statistics and Probability Letters, 61 (2003) 133-143.

[19] P. Hall, Q. Yao, Inference in ARCH and GARCH Models with Heavy-tailed Errors, Econometrica, 71 (2003) 285-317.

[20] D. Straumann, T. Mikosch, Quasi-Maximum-Likelihood Estimation in Conditionally Heteroskedastic Time Series: A Stochastic Recurrence Equations Approach, The Annals of Statistics, 34(5) (2006) 2449-2495.

[21] E. Zivot, J. Wang, Modeling Financial Time Series with S-PLUS, 2nd edition, Springer-Verlag, New York, 2005.

[22] S. Laurent, J.P Peters, G@RCH 4.0, Estimating and Forecasting ARCH models, Timberlake Consultants, 2005.

[23] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, Second Edition, Springer-Verlag, New York, 1991.

[24] P. Bloomfield, On Series Representations for Linear Predictors, The Annals of Probability, 13(1) (1985) 226-233.

[25] P. Bondon, W. Palma, A Class of Antipersistent Processes, Journal of Time Series Analysis, 28 (2007) 261-273.

[26] A. Perez, P. Zaffaroni, Finite-sample Properties of Maximum Likelihood and Whittle Estimators in EGARCH and FIEGARCH Models, Quantitative and Qualitative Analysis in Social Sciences, 2(1) (2008) 78-97

[27] C.M. Hurvich, E. Moulines, P. Soulier, Estimating Long Memory in Volatility, Econometrica 73(4) (2005) 1283-1328.

[28] R. Durrett, Probability: Theory and Examples, Wadsworth & Brooks/Cole, Pacific Grove, 1991.

[29] V.A. Reisen, S.R.C. Lopes, Some Simulations and Applications of Forecasting Long Memory Time Series Models, Journal of Statistical Planning and Inference, 80(2) (1999) 269-287.

[30] J. Grazzini, Analysis of the Emergent Properties: Stationarity and Ergodicity, Journal of Artificial Societies and Social Simulation 15(2) (2012) 7.

[31] I. Domowitz, M.A El-Gamal, A Consistent Nonparametric Test of Ergodicity for Time Series with Applications, Journal of Econometrics 102 (2001) 365-398.
[32] E. Giacomini, Neural Networks in Quantitative Finance. Master's Thesis, University of Berlin, (2003).
[33] A.K. Dhamija, V.K Bhalla, Financial Time Series Forecasting: Comparison of Neural Networks and ARCH Models, International Research Journal of Finance and Economics, 49 (2010) 185-202.