Universidade Federal do Rio Grande do Sul
Instituto de Matemática
Programa de Pós-Graduação em Matemática

Estimação e Previsão em Processos SFIEGARCH: Variância Finita e Infinita

Taiane Schaedler Prass

Porto Alegre, Agosto de 2013.

Tese submetida por Taiane Schaedler Prass[1] ao Programa de Pós-Graduação em Matemática como requisito parcial para a obtenção do título de Doutor em Matemática pelo Programa de Pós-Graduação em Matemática da Universidade Federal do Rio Grande do Sul.

Orientadora:

Dra. Sílvia Regina Costa Lopes (PPGMAT - UFRGS)

Comissão Examinadora:

Dra. Chang Chiann (Departamento de Estatística - IME-USP)

Dr. Cleber Bisognin (Departamento de Estatística - UFRGS)

Dr. Jorge A. Achcar (Faculdade de Medicina, Ribeirão Preto - USP)

Dr. Márcio Valk (Departamento de Estatística - UFRGS)

Data de Apresentação: 12 de Agosto de 2013.

[1] Bolsista do Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq

"I suppose it is tempting,
if the only tool you have is a hammer,
to treat everything as if it were a nail."
(Abraham Maslow, 1966)

Agradecimentos

À Sílvia, minha orientadora, pela oportunidade, orientação, apoio e paciência.

À Universidade Federal do Rio Grande do Sul pela oportunidade.

Aos professores do Programa de Pós-Graduação em Matemática pelo comprometimento com a qualidade de ensino da universidade.

Ao CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico pelo apoio financeiro.

Ao LCPM - Laboratório de Computação e Pesquisa em Matemática e ao CESUP - Centro Nacional de Supercomputação pelos recursos computacionais.

À banca examinadora pela atenção e comentários que auxiliaram no aprimoramento deste trabalho.

Aos colegas do LCPM, pelo auxílio e amizade.

À Rosane, nossa querida secretária do Programa de Pós-Graduação, pelo auxilio prestado a todos os alunos.

À minha família pelo apoio incondicional, carinho, incentivo, compreensão e confiança.

Aos amigos que estiveram ao meu lado ao longo dessa caminhada e também aos amigos recentes pelo apoio e incentivo.

Em especial, ao meu esposo Guilherme. Teu carinho, atenção, dedicação, amor e, mais do que tudo, paciência foram essenciais para a conlcusão dessa etapa. Por mais que eu me esforce, não existem palavras que descrevam minha gratidão.

Resumo

Neste trabalho definimos e estudamos uma nova classe de processos estocásticos pertencente à família ARCH. Tais processos são denominados FIEGARCH com sazonalidade ou, simplesmente, SFIEGARCH. Definimos e estudamos as propriedades teóricas dos modelos SFIEGARCH em dois contextos diferentes. Primeiramente, consideramos os processos para os quais as inovações possuem variância finita. Em um segundo momento, estendemos a definição do processo SFIEGARCH para inovações com distribuição α-estável, quando $1 < \alpha < 2$, portanto, com variância infinita. Os processos definidos dessa maneira são denotados α-SFIEGARCH para deixar implícita a relação com as distribuições α-estáveis.

Tanto no caso de variância finita quanto infinita, apresentamos condições necessárias e suficientes para que os processos SFIEGARCH estejam bem definidos. Tratamos da invertibilidade, estacionariedade (fraca e estrita), ergodicidade e representação espectral (no caso de variância finita) destes processos. Discutimos a representação por série infinita para o logaritmo da volatilidade e apresentamos uma fórmula de recorrência para o cálculo dos coeficientes nessa representação, discutindo suas propriedades assintóticas.

Apresentamos uma descrição detalhada dos principais métodos para a estimação, tanto do parâmetro de longa dependência, quanto dos demais parâmetros dos modelos SFIEGARCH. Em particular, mostramos como obter as expressões da função de verossimilhança, quase-verossimilhança, e pseudo-verossimilhança para os processos em questão. Discutimos ainda o método de estimação de Whittle e a extensão do estimador para o caso em que as inovações possuem distribuição α-estável. Consideramos ainda métodos Bayesianos para a estimação dos parâmetros dos modelos SFIEGARCH e α-SFIEGARCH.

Abordamos também a previsão para os processos SFIEGARCH. No caso de variância finita, obtemos preditores h passos à frente considerando-se o método da esperança condicional e uma expansão de Taylor, de ordem 2, para a função logaritmo. No caso dos processos α-SFIEGARCH, derivamos condições necessárias para a existência dos preditores definidos através do método da esperança condicional, bem como obtemos preditores h passos à frente utilizando a medida de dispersão.

Aplicações considerando as série temporais dos log-retornos dos índices Bovespa e S&P500 ilustram a utilização, na prática, dos modelos estudados neste trabalho.

Abstract

In this work we define and study a class of stochastic processes belonging to the so called ARCH-type family of models. The new class of models is called Seasonal FIEGARCH or simply SFIEGARCH. Theoretical properties of these models are studied in two different contexts, namely, finite and infinite variance innovations. The processes defined under the assumption that the innovations have α-stable distribution, with $1 < \alpha < 2$, are denoted by α-SFIEGARCH to make implicit the relationship with the stable distribution.

In both cases finite and infinite variance, we present necessary and sufficient conditions under which SFIEGARCH processes are well defined. We provide conditions for the invertibility, stationarity (weak and/or strict) and ergodicity property of these processes. In the finite variance case we also study the spectral density function of the process. We discuss the infinite series representation for the logarithm of volatility, present a recurrence formula to obtain the coefficients in this representation and analyze their asymptotic behavior.

We present a detailed description on the main estimation procedures. In particular, we show how to obtain the expressions for the likelihood, quasi-likelihood and pseudo-likelihood functions for SFIEGARCH processes. We also summarize the Whittle method for parameter estimation and present a Whittle-type estimator which can be applied when the innovation process has α-stable distribution. Bayesian methods for parameter estimation in the context of both SFIEGARCH and α-SFIEGARCH are also discussed.

Forecasting procedures are also discussed. In particular, for the finite variance case, we obtain h-step ahead predictors considering the conditional expectation technic and a order-2 Taylor expansion for the logarithm function. For α-SFIEGARCH processes, we derive necessary conditions for the existence of the h-step ahead predictor obtained through the conditional expectation and define a predictor based on the dispersion measure.

Applications to the Bovespa and S&P500 stock exchange index log-return time series illustrate the practical application of the models studied in this work.

Índice

Capítulo 1

Introdução

O fenômeno de longa dependência em séries temporais foi observado primeiramente por Hurst (1951, 1957) e, mais tarde, aparece novamente nos trabalhos de Mandelbrot e Wallis (1968) e McLeod e Hipel (1978). Desde então, tal fenômeno tem sido tema de estudo e vários métodos destinados a identificar e estimar o parâmetro de longa dependência H, denominado *coeficiente de Hurst*, aparecem na literatura.

Os primeiros modelos para séries temporais com longa dependência foram introduzidos por Mandelbrot (1965), Mandelbrot e Van Ness (1968) e Mandelbrot e Wallis (1968, 1969), no caso de tempo contínuo, e por Granger e Joyeux (1980) e Hosking (1981), no caso de tempo discreto. Mais recentemente, para modelar a longa dependência na volatilidade de série temporais financeiras, foram propostos modelos não-lineares tais como o FIGARCH (*Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity*), introduzido por Baillie et al. (1996), o FIEGARCH (*Fractionally Integrated Exponential Generalized Autoregressive Conditional Heteroskedasticity*), proposto por Bollerslev e Mikkelsen (1996) e o modelo LMSV (*Long-Memory Stochastic volatility*), introduzido por Breidt et al. (1998).

Outra característica marcante de muitas séries temporais observadas na prática é um comportamento cíclico, em alguns casos associado à longa dependência. Essa característica é muito comum em série temporais hidrológicas (veja, por exemplo, Hosking, 1984; Montanari et al., 1997, 2000; Bisognin e Lopes, 2007; Prass et al., 2011, 2012), em dados sobre queimadas (veja, por exemplo, Prass et al., 2012) e em dados financeiros de alta frequência (veja, por exemplo, Bordignon et al., 2007, 2009; Lopes e Prass, 2013b). Para modelar o comportamento cíclico associado com a longa dependência, foram propostas generalizações dos modelos citados no parágrafo anterior. Alguns exemplos são os modelos SARFIMA (veja Porter-Hudak, 1990; Bisognin, 2007; Bisognin e Lopes, 2007, 2009), k-factor Gegenbauer (veja Giraitis e Leipus, 1995) e k-factor GARMA (veja Woodward et al., 1998), que generalizam o modelo ARFIMA e os modelos PLM-EGARCH (*Periodic Long-Memory Exponential Generalized Autoregressive Conditional Heteroskedasticity*, introduzido por Bordignon et al. (2009) e SFIEGARCH (*Seasonal* FIEGARCH), proposto por (Lopes e Prass, 2013b), que generalizam o modelo FIEGARCH.

Os modelos da *família* ARCH (*Autoregressive Conditional Heteroskedasticity*. No inglês, ARCH-*type models*), introduzidos por Engle (1982), em conjunto com os modelos de volatilidade estocástica (no inglês, *stochastic volatility models*), apresentados por Breidt et al. (1998), constituem as principais classes utilizadas na modelagem da volatilidade de séries temporais financeiras. Um fato amplamente conhecido na literatura é que, para os modelos de volatilidade estocástica, a volatilidade é definida como uma variável latente que não é diretamente observável, o que torna o processo de estimação mais complexo.

Bollerslev (2010) apresenta uma revisão dos modelos da família ARCH conhecidos na literatura

até a data de preparação do manuscrito. Dentre eles, destacamos os modelos ARCH(p), apresentados por Engle (1982), GARCH(p,q) (*Generalized* ARCH), introduzidos por Bollerslev (1986), EGARCH(p,q) (*Exponential* GARCH), propostos por Nelson (1991), FIGARCH(p,d,q), apresentados por Baillie et al. (1996) e FIEGARCH(p,d,q), propostos por Bollerslev e Mikkelsen (1996). Ressaltamos que, dentre estes, o modelo FIEGARCH é o que apresenta mais vantagens sobre os demais. Em particular, análogo aos modelos ARCH e GARCH, são capazes de descrever cluster de volatilidade; análogo aos modelos EGARCH, capturam a assimetria dos retornos (a volatilidade reage de forma assimétrica aos retornos, tendendo a ser maior para retornos negativos); análogo aos modelos FIGARCH, levam em conta a longa dependência na volatilidade, com a vantagem de serem fracamente estacionários quando $d < 0.5$ (veja Lopes e Prass, 2013a).

Nos últimos anos, com os avanços da informática, especialmente em termos de armazenamento, transmissão e manipulação de dados, a atenção dos pesquisadores tem-se voltado cada vez mais à análise de dados financeiros coletados em intervalos de tempo cada vez menores. O estudo desses dados revelou, por exemplo, a existência tanto de longa dependência, quanto de comportamento cíclico na volatilidade de séries temporais financeiras de alta frequência. Com isso, surgiu a necessidade de desenvolver modelos que captem tais características. Embora os modelos FIEGARCH descrevam a longa dependência na volatilidade, tais modelos não são apropriados para séries financeiras com longa dependência do tipo sazonal. Sendo assim, Bordignon et al. (2007, 2009) propõem os modelos G-GARCH (Gegenbauer-GARCH), PLM-GARCH (*Periodic Long-Memory* GARCH), PLM-LGARCH (*Periodic Long-Memory* log-GARCH) e PLM-EGARCH (*Periodic Long-Memory* EGARCH). Segundo Bordignon et al. (2007), resultados sobre a estacionariedade fraca dos processos G-GARCH são desconhecidos. Bordignon et al. (2009) afirma, sem apresentar uma prova formal, que o processo PLM-GARCH não é fracamente estacionário mas, sob certas condições, obtém-se estacionariedade estrita. Propriedades dos processos PLM-LGARCH e PLM-EGARCH não são discutidas pelos autores.

Neste trabalho estudamos uma nova classe de modelos pertencentes à família ARCH. As principais notações utilizadas no decorrer do trabalho são apresentadas no Capítulo 2. Em particular, nesse capítulo definimos *longa dependência para processos com variância infinita*. No Capítulo 3 resumimos os principais resultados relacionados à distribuições α-estáveis. A teoria sobre distribuições estáveis univariadas foi desenvolvida, essencialmente, em torno dos anos 1920 e 1930 por Paul Lévy e Aleksabder Yakovlevich Khinchine. Um estudo detalhado sobre a teoria pode ser encontrado em clássicos como Gnedenko e Kolmogorov (1954) e Feller (1971) ou em trabalhos mais recentes como Zolotarev (1986) e Samorodnitsky e Taqqu (1994). Dentre os principais resultados deste capítulo, ressaltamos a representação assintótica das caudas da distribuição estável; a existências dos momentos de ordem r para variáveis estáveis; as diferentes parametrizações para a função característica; o cálculo da função densidade de probabilidade estável e o método para gerar variáveis aleatórias estáveis.

No Capítulo 4 introduzimos os processos FIEGARCH *com sazonalidade*, denotados por SFIEGARCH. A proposta com este modelo é generalizar os processos FIEGARCH introduzidos por Bollerslev e Mikkelsen (1996) e estudados detalhadamente em Prass (2008) e Lopes e Prass (2013a) (Apêndice C). Definimos e estudamos as propriedades teóricas dos modelos SFIEGARCH em dois contextos diferentes. Primeiramente, consideramos os processos para os quais as inovações possuem variância finita. Tais processos são apresentados na Seção 4.1 e discutidos com detalhes em Lopes e Prass (2013b) (Apêndice D). Em um segundo momento, estendemos a definição do processo SFIEGARCH para inovações com distribuição α-estável, quando $1 < \alpha < 2$, portanto, com variância infinita. Os processos definidos dessa maneira são denotados α-SFIEGARCH para deixar implícita a relação com distribuições α-estáveis. Tais processos são apresentados na Seção 4.3 e discutidos com detalhes em Prass, Lopes e Crato (2013) (Apêndice I).

Tanto no caso de variância finita quanto infinita, apresentamos condições necessárias e suficientes para que os processos SFIEGARCH estejam bem definidos. Tratamos da invertibilidade, estacionariedade (fraca e estrita), ergodicidade e representação espectral (no caso de variância finita) destes processos. Além disso, discutimos a representação por série infinita para a variável $\ln(h_t^2)$, para $t \in \mathbb{Z}$, onde h_t é a volatilidade estocástica associada à um processo SFIEGARCH. Apresentamos uma fórmula de recorrência para os coeficientes nessa representação e estudamos o comportamento assintótico destes, no sentido teórico e de simulações. Analisamos a influência do ponto de truncamento da soma infinita, na geração de processos SFIEGARCH, conduzindo um breve estudo de simulação considerando diferentes pontos de truncamento e comparando as amostras geradas dessa maneira.

Dedicamos o Capítulo 5 à descrição dos principais métodos para a estimação, tanto do parâmetro de longa dependência d, quanto dos demais parâmetros dos modelos SFIEGARCH, estudados no Capítulo 4. Ressaltamos que os métodos heurísticos apresentados nesse capítulo são, em geral, utilizados como uma primeira estimativa do parâmetro de longa dependência H, denominado $coeficiente\ de\ Hurst$. Discutimos também métodos paraméricos e semi-paramétricos para a estimação dos parâmetros do modelo. No domínio do tempo, apresentamos os estimadores da máxima verossimilhança (MLE), da quase-máxima verossimilhança (QMLE), da pseudo-máxima verossimilhança (PMLE) e, no domínio da frequência, o estimador de Whittle. Embora o estimador de Whittle seja mais flexível no sentido de que não são necessárias hipóteses quanto à distribuição dos dados, no caso dos processos SFIEGARCH, tal estimador torna-se computacionalmente complexo. Apresentamos também a versão modificada do estimador de Whittle, que pode ser aplicada para processos com distribuição α-estável.

O Capítulo 5 inclui ainda uma seção que trata de métodos Bayesianos para a estimação dos parâmetros do modelo. A descrição detalhada do procedimento adotado é dada em Prass, Lopes e Achcar (2013) (Apêndice H) e em Prass, Lopes e Crato (2013) (Apêndice I). Métodos Bayesianos têm sido utilizados com frequência na análise de dados financeiros, principalmente devido à complexidade da função de verossimilhança para modelos de volatilidade estocástica (veja, por exemplo, Meyer e Yu, 2000). A principal diferença do método Bayesiano para a análise clássica é que, no primeiro caso, os parâmetros do modelo são tratados como variáveis aleatórias enquanto que, no segundo caso, são tratados como constantes reais (ou complexas, de acordo com o problema).

No Capítulo 6 abordamos a previsão para os processos SFIEGARCH. Dividimos nosso estudo em dois casos. Na Seção 6.1 tratamos da previsão para processos SFIEGARCH quando as inovações possuem variância finita. Os resultados apresentados nessa seção fazem parte dos artigos Lopes e Prass (2013a,b) (Apêndices C e D, respectivamente). Na Seção 6.2 consideramos os processos α-SFIEGARCH$(p, d, q)_s$ definidos na Seção 4.3. Dentre os resultados apresentados na Seção 6.2, destacamos a obtenção de condições necessárias para a existência do momento de ordem r das variáveis aleatórias X_t e h_t, para todo $t \in \mathbb{Z}$ (dadas na definição dos processos α-SFIEGARCH); a obtenção dos preditores h passos à frente, através do método da esperança condicional; a definição da medida de dispersão $\mathrm{disp}(\cdot)$ e a utilização de tal medida para obter preditores h-passos à frente para os processos associados à um α-SFIEGARCH.

Conclusões e futuros trabalhos são discutidos no Capítulo 7.

Capítulo 2

Conceitos Preliminares

Neste capítulo introduzimos a notação utilizada no decorrer do trabalho. As definições e resultados aqui apresentados podem ser encontradas na literatura básica de séries temporais e processos estocásticos. As definições que envolvem funções podem ser encontradas em livros de análise. Embora tais conceitos sejam básicos, eles são necessários para o desenvolvimento deste trabalho. Não apresentamos provas para os resultados aqui introduzidos mas sempre que necessário indicamos uma bibliografia.

2.1 Conjuntos, Funções e Sequências Reais

No decorrer deste trabalho os símbolos $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$ e $\mathbb{C}$ são utilizados unicamente para denotar o *conjunto dos números naturais, inteiros, reais* e *complexos*, respectivamente. Por convenção, assumimos que o número zero pertence aos naturais. Além disso, dado um conjunto $T \neq \emptyset$ arbitrário, que contenha o zero, definimos $T^* := T \setminus \{0\}$. O símbolo "$\mathfrak{i}$" é utilizado para representar o valor complexo $\sqrt{-1}$ e $\mathfrak{Re}(z)$ denota a parte real de z, para qualquer $z \in \mathbb{C}$.

Dado um conjunto ou evento aleatório A, denotamos por $\mathbb{I}_A(\cdot)$ a *função indicadora de* A, ou seja,

$$\mathbb{I}_A(z) = \begin{cases} 1, & \text{se} \quad z \in A, \\ 0, & \text{se} \quad z \notin A. \end{cases}$$

As funções *maior* e *menor inteiro* são representadas por $\lfloor \cdot \rfloor$ e $\lceil \cdot \rceil$, respectivamente. A *função sinal* $\operatorname{sgn}(\cdot)$ é definida, para todo $t \in \mathbb{R}$, através da expressão

$$\operatorname{sgn}(t) = \begin{cases} 1, & \text{se} \quad t > 0, \\ 0, & \text{se} \quad t = 0, \\ -1, & \text{se} \quad t < 0. \end{cases} \tag{2.1}$$

Observamos que a função $\operatorname{sgn}(\cdot)$ e a função valor absoluto $|\cdot|$ estão relacionadas através da igualdade $\operatorname{sgn}(t) = t|t|$, para qualquer $t \in \mathbb{R}$. $\Gamma(\cdot,\cdot)$ e $\Gamma(\cdot)$ denotam, respectivamente, a função *Gama incompleta* e *Gama*, dadas por

$$\Gamma(x,y) = \int_y^\infty t^{x-1}\, e^{-t}\, dt \quad \text{e} \quad \Gamma(x) = \Gamma(x,0) = \int_0^\infty t^{x-1}\, e^{-t}\, dt, \quad \text{para todo } x, y \in \mathbb{R}. \tag{2.2}$$

A notação assintótica utilizada no decorrer do trabalho é descrita na Tabela 2.1. Observamos que as definições apresentadas nesta tabela são também válidas quando $f(\cdot)$ e $g(\cdot)$ são substituídas por duas sequências de números reais $\{a_n\}_{n\in\mathbb{N}}$ e $\{b_n\}_{n\in\mathbb{N}}$. Nesse caso, sempre assumimos $a = \infty$.

Tabela 2.1: Notação assintótica considerando-se $f, g : \mathbb{R} \to \mathbb{R}$ duas funções quaisquer e $a \in \mathbb{R} \cup \{-\infty, \infty\}$.

Notação	Significado
$f(x) = O(g(x))$, quando $x \to a$	existe um número real $c > 0$ tal que $\lvert f(x)\rvert \leq c\lvert g(x)\rvert$, quando $x \to a$
$f(x) = o(g(x))$, quando $x \to a$	$\lim\limits_{x\to a} \dfrac{f(x)}{g(x)} = 0$
$f(x) \sim g(x)$, quando $x \to a$	$f(x) - g(x) = o(g(x))$, quando $x \to a$. De forma equivalente, $\lim\limits_{x\to a} \dfrac{f(x)}{g(x)} = 1$
$f(x) \approx g(x)$, quando $x \to \infty$	para qualquer $\varepsilon > 0$, existe $x_0 \in \mathbb{R}$ tal que, se $x \geq x_0$, então $\lvert f(x) - g(x)\rvert < \varepsilon$

2.2 Processos Estocásticos

As definições que seguem estão relacionadas a processos estocásticos e séries temporais. Tais definições são importantes para o desenvolvimento do Capítulo 4. Assumimos que o leitor esteja familiarizado com a seguinte notação utilizada no trabalho.

- Dado $\mathfrak{p} \geq 0$, $\ell^{\mathfrak{p}}$ denota o conjunto de todas as sequências reais $\{a_n\}_{n\in\mathbb{N}}$ que satisfazem $\sum_{n=0}^{\infty} |a_n|^{\mathfrak{p}} < \infty$.
- Ω denota o espaço amostral.
- $\mathcal{F}$ é a σ-álgebra de eventos aleatórios.
- $\mathbb{P}$ é a medida de probabilidade definida em $\mathcal{F}$.
- Dado o espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$, denotamos por $\mathcal{L}^{\mathfrak{p}}$ o espaço $\mathcal{L}^{\mathfrak{p}}(\Omega, \mathcal{F}, \mathbb{P})$, isto é, o conjunto de todas as variáveis aleatórias X, definidas sobre Ω, satisfazendo a condição

$$\mathbb{E}(|X|^{\mathfrak{p}}) = \int_{\Omega} |X(\omega)|^{\mathfrak{p}} \mathbb{P}(d\omega) < \infty, \quad \text{para qualquer } \mathfrak{p} \geq 0.$$

Ressaltamos alguns fatos importantes que serão empregados no Capítulo 4

- se $\{a_n\}_{n\in\mathbb{N}} \in \ell^{\mathfrak{p}}$, então $\{a_n\}_{n\in\mathbb{N}} \in \ell^{r}$, para todo $r \geq \mathfrak{p}$.
- Como consequência imediata da desigualdade de Hölder, temos

$$\left[\mathbb{E}(|X|^{s})\right]^{1/s} \leq \left[\mathbb{E}(|X|^{r})\right]^{1/r}, \quad \text{para todo } 0 < s \leq r.$$

  Esse resultado é também conhecido como desigualdade de Lyapunov e nos permite concluir que, se $X \in \mathcal{L}^r$ então $X \in \mathcal{L}^s$.

- Ambos, $\mathcal{L}^{\mathfrak{p}}$ e $\ell^{\mathfrak{p}}$, são espaços lineares completos.

Nas definições que seguem, $T \neq \emptyset$ representa um conjunto de índices arbitrário, podendo ser discreto ou contínuo.

Definição 2.1. (Processo Estocástico). Um *processo estocástico* é uma família de variáveis aleatórias $\{X_t\}_{t\in\mathrm{T}}$ definidas em um mesmo espaço de probabilidades $(\Omega, \mathcal{F}, \mathbb{P})$. Para cada $\omega \in \Omega$ fixo, $\{X_t(\omega)\}_{t\in T}$ é denominado uma *trajetória* ou uma *realização* do processo estocástico $\{X_t\}_{t\in\mathrm{T}}$.

Informalmente, uma *série temporal* é um conjunto de observações medidas em diferentes tempos. Formalmente, dizemos que uma série temporal é uma realização, ou parte de uma realização $\{X_t(\omega)\}_{t\in T_0}$, de um processo estocástico $\{X_t\}_{t\in T}$, com $T_0 \subseteq T$.

Observação 2.1. No decorrer deste trabalho, a menos que explicitado o contrário, consideramos $T \subseteq \mathbb{Z}$ e $T_0 = \{1, \cdots, n\}$, para algum $n \in \mathbb{N}$. Além disso, como um abuso de notação, diremos que $\{X_t\}_{t=1}^{n}$ é uma série temporal, deixando implícita a dependência em $\omega \in \Omega$.

2.2.1 Processos Ergódicos

Informalmente, dizemos que um processo $\{X_t\}_{t\in\mathbb{Z}}$ é *ergódico* se todas as suas propriedades estatísticas podem ser determinadas a partir de uma única realização desse processo. Sendo assim, ergodicidade é uma propriedade muito importante para aplicações práticas e para a análise de séries temporais.

No que segue apresentamos alguns conceitos básicos da teoria ergódica, como a definição formal de processo ergódico e os principais teoremas relacionados à ergodicidade. Por simplicidade, na maioria das definições e teoremas que segue, consideramos $T = \mathbb{N}$, mas todos os resultados apresentados podem ser facilmente generalizados para $T = \mathbb{Z}$. Consideramos apenas processos estritamente estacionários, no sentido da Definição 2.5.

Seja S um conjunto qualquer munido de uma σ-álgebra $\mathcal{A}$, $\{X_t\}_{t\in\mathbb{N}}$ um processo estocástico tal que $X_t \in S$, para todo $t \in \mathbb{N}$ e $\Omega = S^{\mathbb{N}}$ o espaço amostral dos caminhos. Então, um cilindro em Ω é um conjunto da forma

$$\{X_{t_1} \in A_1, X_{t_2} \in A_2, \cdots, X_{t_k} \in A_k\},$$

onde A_i é um subconjunto de S e $t_i \in \mathbb{N}$, para todo $1 \leq i \leq k$.

Denote por $\mathcal{F}$ a σ-álgebra gerada pelos cilindros em Ω. O processo $\{X_t\}_{t\in\mathbb{N}}$ define, de maneira única (veja Billingsley, 1995), uma probabilidade $\mathbb{P}$ sobre cilindros da seguinte forma,

$$\mathbb{P}(X_{t_1} \in A_1, X_{t_2} \in A_2, \cdots, X_{t_k} \in A_k), \quad \text{para} \quad k, t_i \in \mathbb{N},$$

onde $t_1 < t_2 < \cdots < t_k$ e A_i é subconjunto de S, para todo $1 \leq i \leq k$. A função $\mathbb{P}(\cdot)$ é denominada *probabilidade associada ao processo estocástico* $\{X_t\}_{t\in\mathbb{N}}$.

Lembramos que, dados dois espaços mensuráveis $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ quaisquer, uma transformação $\varphi : \Omega_1 \to \Omega_2$ é dita *mensurável* se, e somente se, para todo $A \in \mathcal{F}_2$, $\varphi^{-1}(A) \in \mathcal{F}_1$. Além disso, dado um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ e uma transformação $\varphi : \Omega \to \Omega$, dizemos que $\varphi(\cdot)$ *preserva a probabilidade* $\mathbb{P}$ se, e somente se, $\mathbb{P}(\varphi^{-1}(A)) = \mathbb{P}(A)$, para todo $A \in \mathcal{F}$. Diz-se também que $\mathbb{P}$ é *invariante* para φ.

Definição 2.2. (Ergodicidade). Seja $\varphi : \Omega \to \Omega$ uma transformação que preserva a probabilidade $\mathbb{P}$. Então, $\mathbb{P}$ é *ergódica* se, e somente se, para todo $A \in \mathcal{F}$ tal que $\varphi^{-1}(A) = A$ tem-se $\mathbb{P}(A) = 0$ ou $\mathbb{P}(A) = 1$.

O teorema que segue estabelece equivalências para a condição de ergodicidade de um processo estocástico. A caracterização no item (iv) do Teorema 2.1 indica que um processo ergódico é aquele para o qual a média de uma única série temporal, extraída deste processo, permite estimar o seu valor esperado.

Teorema 2.1. *Seja $\{X_t\}_{t\in\mathbb{N}}$ um processo fortemente estacionário para a probabilidade $\mathbb{P}$. As seguintes condições são equivalentes:*

(i) *$\{X_t\}_{t\in\mathbb{N}}$ é ergódico;*

(ii) *Para todo conjunto $A\in\mathcal{F}$ de sequências $(x_0,x_1,\cdots)$, temos*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{I}_A(X_i,X_{i+1},\cdots)=\mathbb{P}\big((X_0,X_1,\cdots)\in A\big);$$

(iii) *Para todo $k\in\mathbb{N}$ e todo conjunto $A\in\mathcal{F}$ de vetores $(x_0,x_1,\cdots,x_k)$, temos*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{I}_A(X_i,X_{i+1},\cdots,X_{i+k})=\mathbb{P}\big((X_0,X_1,\cdots,X_k)\in A\big);$$

(iv) *Para toda função $g:\Omega\to\mathbb{R}$ mensurável temos,*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}g\big(X_i,X_{i+1},\cdots\big)=\mathbb{E}\big(g(X_0,X_1,\cdots)\big),$$

sempre que a esperança existir;

(v) *Para todo $k\in\mathbb{N}$ e toda função mensurável $g:\mathbb{R}^{k+1}\to\mathbb{R}$, temos*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}g\big(X_i,X_{i+1},\cdots,X_{i+k}\big)=\mathbb{E}\big(g(X_0,X_1,\cdots,X_k)\big),$$

sempre que a esperança existir.

Prova: Veja Karlin e Taylor (1975). ∎

A seguir definimos a função de distribuição n-dimensional, para todo n finito, também denominada *função de distribuição conjunta*, de um processo estocástico. O processo estocástico $\{X_t\}_{t\in\mathbb{Z}}$ estará especificado se conhecermos suas distribuições finito-dimensionais, para todo $n\geq 1$.

Definição 2.3. (Função de Distribuição Finito-Dimensional). Seja $\{X_t\}_{t\in\mathbb{Z}}$ um processo estocástico. As *funções de distribuição finito-dimensionais* de $\{X_t\}_{t\in\mathbb{Z}}$ são as funções $\{F_{\mathbf{t}}(\cdot);\mathbf{t}\in\mathcal{T}\}$ definidas por

$$F_{\mathbf{t}}(\mathbf{x})=\mathbb{P}(X_{t_1}\leq x_1,\cdots,X_{t_m}\leq x_m),$$

onde $\mathbf{t}=(t_1,\cdots,t_m)'\in\mathcal{T}$, $\mathbf{x}=(x_1,\cdots,x_m)'\in\mathbb{R}^m$ e $\mathcal{T}$ é o conjunto de todos os vetores $\big\{\mathbf{t}=(t_1,\cdots,t_m)'\in\mathbb{Z}^m:t_1<\cdots<t_m,\ m\in\mathbb{N}-\{0\}\big\}$.

Note que, se tomarmos $m=1$ na definição anterior, obtemos as funções de distribuição unidimensionais para o processo estocástico $\{X_t\}_{t\in\mathbb{Z}}$, que denotamos por $F_{X_t}(\cdot)$ ao invés de $F_t(\cdot)$, para todo $t\in\mathbb{Z}$. Em particular, se as variáveis aleatórias X_t são identicamente distribuídas, para todo $t\in\mathbb{Z}$, escreveremos apenas $F_X(\cdot)$. Além disso, adotamos a seguinte convenção,

- a sigla i.i.d. denota *independente e identicamente distribuídas*;
- o símbolo $\stackrel{d}{=}$ denota igualdade em distribuição. Também escreveremos $X\sim Y$, para dizer que as variáveis aleatórias X e Y tem a mesma função de distribuição. Também utilizamos a notação $X\sim F$ para dizer que a variável aleatória X tem função de distribuição F;

- o símbolo $\stackrel{d}{\Longrightarrow}$ denota *convergência em distribuição*, também denominada *convergência fraca*;
- dizemos que uma variável aleatória X possui um *domínio de atração* se existe uma sequência $\{Y_n\}_{n\geq 1}$ de variáveis aleatórias i.i.d. e sequências reais $\{a_n\}_{n\geq 1}$ e $\{d_n\}_{n\geq 1}$, com $d_n \neq 0$ para todo $n \geq 1$, tais que
$$\frac{Y_1 + \cdots + Y_n}{d_n} + a_n \stackrel{d}{\Longrightarrow} X. \tag{2.3}$$
Observamos ainda que,
  - se $X \sim F$, dizer que X possui um domínio de atração é equivalente a dizer que F possui um domínio de atração;
  - se $\{Y_n\}_{n\geq 1}$ é uma sequência de variáveis aleatórias i.i.d. para a qual a expressão (2.3) é satisfeita e Y é uma cópia independente de Y_n, para qualquer $n \geq 1$, então dizemos que a variável aleatória Y pertence ao domínio de atração de X;
  - se $X \sim F$ e se $Y \sim G$, dizer que Y pertence ao domínio de atração de X é equivalente a dizer que G pertence ao domínio de atração de F.

Na prática, a caracterização de uma variável aleatória através de suas distribuições finito-dimensionais torna-se muito difícil e, em alguns casos, impossível. Uma alternativa é caracterizar a variável aleatória através de sua função característica. Tal caracterização é de grande importância e é amplamente utilizada no Capítulo 3. No que segue, apresentamos a definição desta função e suas principais propriedades.

2.2.2 Função Característica

A função característica fornece uma maneira alternativa de descrever uma variável aleatória X. Assim como a função de distribuição, a função característica determina, por completo, o comportamento probabilístico de X. Como veremos a seguir, esses dois métodos são equivalentes, no sentido que, dada uma variável aleatória X, existe uma correspondência única entre as funções de distribuição e característica de X.

Definição 2.4. (Função Característica). Seja X uma variável aleatória. A *função característica* da variável aleatória X é a função $\varphi_X \colon \mathbb{R} \to \mathbb{C}$, definida por

$$\varphi_X(t) := \mathbb{E}\big(e^{\mathrm{i}tX}\big) = \int_{\mathbb{R}} e^{\mathrm{i}tx}\, dF_X(x), \quad \text{para todo} \quad t \in \mathbb{R}, \tag{2.4}$$

onde $F_X(\cdot)$ é a função de distribuição de X e a integral dada no lado direito da expressão (2.4) é a integral de Riemann-Stieltjes.

A proposição que segue apresenta as propriedades da função característica de uma variável aleatória X. Tais resultados são empregados nas provas dos teoremas e proposições apresentados no Capítulo 3. Mais detalhes e provas para esses resultados podem ser encontrados em Roussas (2005) e Tucker (1967).

Proposição 2.1. *A função característica $\varphi_X(\cdot)$ de uma variável aleatória X apresenta as seguintes propriedades,*

(i) *como a função característica $\varphi_X(\cdot)$ é definida a partir de uma integral de uma função contínua e limitada, sob um espaço cuja medida é finita, a função característica de uma variável aleatória sempre existe;*

(ii) $\varphi_X(\cdot)$ *é uniformemente contínua em* $\mathbb{R}$*;*

(iii) $\varphi_X(0) = 1$*;*

(iv) *a função característica é limitada. Mais especificamente,* $|\varphi_X(t)| \leq 1$*, para todo* $t \in \mathbb{R}$*;*

(v) $\varphi_X(\cdot)$ *é Hermitiana, isto é,* $\varphi_X(-t) = \overline{\varphi_X(t)}$*. Em particular, a função característica de uma variável aleatória simétrica (em torno da origem) é uma função real e par;*

(vi) *existe uma bijeção entre a função de distribuição e a função característica, isto é, para quaisquer variáveis aleatórias* X_1 *e* X_2*,*

$$F_{X_1}(\cdot) = F_{X_2}(\cdot) \quad \Longleftrightarrow \quad \varphi_{X_1}(\cdot) = \varphi_{X_2}(\cdot);$$

(vii) *se a variável aleatória* X *possui momento até ordem* k *finito, então a função característica é* k *vezes continuamente diferenciável e*

$$\mathbb{E}(X^k) = (-\mathfrak{i})^k \varphi_X^{(k)}(0);$$

(viii) *se* $\varphi_X(\cdot)$ *é* k *vezes diferenciável em zero, então a variável aleatória* X *possui todos os momentos até ordem* k *finitos, se* k *é par, mas somente até a ordem* $k-1$ *finitos, se* k *é ímpar;*

(ix) *se* $X_1, \cdots, X_n$ *são variáveis aleatórias independentes e* $a_1, \cdots, a_n$ *são constantes então, a função característica da combinação linear* $\sum_{i=1}^n a_i X_i$ *é dada por*

$$\varphi_{a_1 X_1 + \cdots + a_n X_n}(t) = \varphi_{X_1}(a_1 t) \times \cdots \times \varphi_{X_n}(a_n t).$$

Prova: Veja Roussas (2005, Capítulo 11). ∎

Pela expressão (2.4), dada a função de distribuição $F_X(\cdot)$ de uma variável aleatória X, a sua função característica fica unicamente determinada. O resultado que segue é de grande importância pois mostra que, dada a função característica de uma variável aleatória, a sua função de distribuição pode ser recuperada através da chamada *fórmula da inversão*.

Teorema 2.2. *Sejam* $F_X(\cdot)$ *e* $\varphi_X(\cdot)$*, respectivamente, as funções de distribuição e característica da variável aleatória* X*. Então,*

(i) *para todo* $a, b \in \mathbb{R}$ $(a < b)$ *e* $T > 0$*,*

$$\frac{F_X(b) - F_X(b-)}{2} - \frac{F_X(a) - F_X(a-)}{2} = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-\mathfrak{i}ta} - e^{-\mathfrak{i}tb}}{\mathfrak{i}t} \varphi_X(t) dt;$$

(ii) *se* a *e* b *são pontos de continuidade de* $F_X(\cdot)$ *e* $T > 0$*,*

$$F_X(b) - F_X(a) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-\mathfrak{i}ta} - e^{-\mathfrak{i}tb}}{\mathfrak{i}t} \varphi_X(t) dt;$$

(iii) *se* X *é uma variável aleatória discreta, que assume o valor* x_j*, com probabilidade* $P(X = x_j) = p(x_j)$*, para* $j \geq 1$*, então*

$$p(x_j) = \lim_{T\to\infty} \frac{1}{2\pi T} \int_{-T}^{T} e^{-\mathfrak{i}tx_j} \varphi_X(t) dt, \quad \textit{para} \quad j \geq 1,\ T > 0.$$

Prova: Veja Roussas (2005, página 238). ∎

Se a função de distribuição $F_X(\cdot)$ da variável aleatória X é absolutamente contínua, então, existe uma função $p_X(\cdot)$ tal que

$$F_X(x) = \int_{-\infty}^{x} p_X(t)dt, \quad \text{para todo } x \in \mathbb{R}.$$

A função $p_X(\cdot)$ é denominada *função densidade de probabilidade* de $F_X(\cdot)$. Mais precisamente, $p_X(\cdot)$ é, quase toda parte (com respeito a medida de Lebesgue), unicamente determinada pelo teorema de Radon-Nikodym. De fato, $p_X(\cdot)$ é a derivada de Radon-Nikodym da medida de Lebesgue-Stieltjes determinada por $F_X(\cdot)$, com respeito a medida de Lebesgue.

Corolário 2.1. *Seja $\varphi_X(\cdot)$ a função característica da variável aleatória X. Se $\varphi_X(\cdot)$ é tal que $\int_{-\infty}^{\infty} |\varphi_X(t)|dt < \infty$, então a derivada $F_X'(x) = p_X(x)$ existe, é limitada, contínua em $\mathbb{R}$ e é dada por*

$$p_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\mathrm{i}tx} \varphi_X(t)dt. \tag{2.5}$$

Prova: Veja Roussas (2005, página 244). ∎

A expressão (2.5) é denominada *fórmula da inversão*. O resultado apresentado no Corolário 2.1 é de grande importância para o estudo de distribuições estáveis (veja Capítulo 3), para as quais a função densidade de probabilidade em geral não possui uma expressão fechada.

2.2.3 Processos Estacionários

Outro problema enfrentado na prática é como lidar com a dependência entre duas ou mais variáveis aleatórias e como caracterizar e modelar tal dependência. Ao escolher um modelo para uma série temporal, geralmente impomos restrições ao tipo de dependência com que estamos tratando, já que uma estrutura de dependência arbitrária é matematicamente intratável. Em geral, assumimos que os processos estocásticos são estacionários.

Definição 2.5. (Estacionariedade). Seja $\{X_t\}_{t\in\mathbb{Z}}$ um processo estocástico. Então,

(a) $\{X_t\}_{t\in\mathbb{Z}}$ é *estritamente estacionário* se as distribuições finito-dimensionais forem invariantes sob translações de tempo, isto é,

$$\mathbb{P}(X_{t_1+h} \leq x_1, \cdots, X_{t_k+h} \leq x_k) = \mathbb{P}(X_{t_1} \leq x_1, \cdots, X_{t_k} \leq x_k),$$

para todo inteiro positivo k e para quaisquer $t_1, \cdots, t_k$, $h \in \mathbb{Z}$.

(b) $\{X_t\}_{t\in\mathbb{Z}}$ é *fracamente estacionário* (*covariância estacionário* ou *estacionário de segunda ordem*) se,

(i) $\mathbb{E}\big(|X_t|^2\big) < \infty$, para todo $t \in \mathbb{Z}$;

(ii) $\mathbb{E}(X_t) = \mu$, para todo $t \in \mathbb{Z}$ e $\mu \in \mathbb{R}$ constante;

(iii) $\mathrm{Cov}(X_r, X_s) = \mathrm{Cov}(X_{r+t}, X_{s+t})$, para todo r, s e $t \in \mathbb{Z}$.

Observação 2.2. Se $\{X_t\}_{t\in\mathbb{Z}}$ é um processo estocástico tal que $\mathrm{Var}(X_t) < \infty$, para todo $t \in \mathbb{Z}$, utilizamos a seguinte notação:

(1) $\gamma_X(\cdot,\cdot)$, para denotar a *função de autocovariância* de $\{X_t\}_{t\in\mathbb{Z}}$, definida por

$$\gamma_X(r,s) = \mathrm{Cov}(X_r, X_s) = \mathbb{E}\Big[\big(X_r - \mathbb{E}(X_r)\big)\big(X_s - \mathbb{E}(X_s)\big)\Big],$$

onde $\gamma_X(r,r) = \mathrm{Var}(X_r)$, para todo $r, s \in \mathbb{Z}$.

(2) $\rho_X(\cdot,\cdot)$, para denotar a *função de autocorrelação* de $\{X_t\}_{t\in\mathbb{Z}}$, definida por

$$\rho_X(r,s) = \mathrm{Corr}(X_r, X_s) = \frac{\gamma_X(r,s)}{\sqrt{\mathrm{Var}(X_r)}\sqrt{\mathrm{Var}(X_s)}},$$

para todo $r, s \in \mathbb{Z}$.

(3) Se o processo estocástico $\{X_t\}_{t\in\mathbb{Z}}$ é estacionário (estritamente ou fracamente), então $\gamma_X(r,s) = \gamma_X(r-s,0)$, para todo $r, s \in \mathbb{Z}$, e as funções de autocovariância e autocorrelação de $\{X_t\}_{t\in\mathbb{Z}}$ são, equivalentemente definidas como

$$\gamma_X(h) := \gamma_X(h,0) = \mathrm{Cov}(X_{t+h}, X_t) \quad \text{e} \quad \rho_X(h) := \frac{\gamma_X(h)}{\gamma_X(0)}, \quad \text{para todo} \quad t, h \in \mathbb{Z}.$$

O seguinte teorema fornece condições necessárias e suficientes para que uma dada função $\gamma(\cdot)$ seja uma função de autocovariância, associada à algum processo estocástico.

Teorema 2.3. *Seja $\gamma : \mathbb{Z} \to \mathbb{C}$ uma função complexa definida sobre os inteiros. Então, $\gamma(\cdot)$ é a função de autocovariância de um processo estocástico estacionário se e somente se é uma função Hermitiana e não-negativa definida, isto é, se e somente se*

(i) $\gamma(h) = \overline{\gamma(-h)}$, *para todo inteiro* $h > 0$,

(ii) *para todos os vetores* $\boldsymbol{a} = (a_1, \cdots .a_m)' \in \mathbb{C}^m$ *e* $\boldsymbol{t} = (t_1, \cdots, t_m)' \in \mathbb{Z}^m$ *tem-se*

$$\sum_{i=1}^{m}\sum_{j=1}^{m} a_i \gamma(t_i - t_j)\bar{a}_j \geq 0.$$

Prova: Veja Brockwell e Davis (1991, teorema 4.1.1.) ∎

O Teorema 2.3 caracteriza a função de autocovariância de um processo estocástico como uma função Hermitiana e não negativa. Por outro lado, a propriedade de ser Hermitiana segue imediatamente da propriedade de ser não-negativa definida (Brockwell e Davis, 1991, página 115). Sendo assim, o teorema que segue, denominado na literatura Teorema de Herglotz, é de grande importância pois fornece condições necessárias e suficientes para que uma dada função $\gamma(\cdot)$ seja não-negativa definida e, portanto, uma função de autocovariância associada à algum processo estocástico. O Teorema 2.4 também define as funções de distribuição e densidade espectral de um processo estocástico, bem como associa tais funções à função de autocovariância.

Teorema 2.4. (Teorema de Herglotz). *Uma função $\gamma : \mathbb{Z} \longrightarrow \mathbb{C}$ é não-negativa definida se, e somente se,*

$$\gamma(h) = \int_{(-\pi,\pi]} e^{\mathrm{i}h\lambda} d\mathrm{F}(\lambda), \quad \textit{para todo } h \in \mathbb{Z}, \tag{2.6}$$

onde $\mathrm{F}(\cdot)$ é uma função limitada em $[-\pi,\pi]$, contínua à direita, não-decrescente e $\mathrm{F}(-\pi) = 0$.

Prova: Veja Brockwell e Davis (1991, página 118). ∎

A função $\mathrm{F}(\cdot)$, dada na expressão (2.6), é denominada *função de distribuição espectral* de $\gamma(\cdot)$. Se $\mathrm{F}(\cdot)$ pode ser escrita como

$$\mathrm{F}(\lambda) = \int_{-\pi}^{\lambda} f(\omega) d\omega, \quad \text{para todo } \lambda \in (-\pi, \pi],$$

então $f(\cdot)$ é denominada *função densidade espectral* de $\gamma(\cdot)$. Além disso, quando especificamos o processo estocástico estacionário $\{X_t\}_{t\in\mathbb{Z}}$ ao qual $\gamma(\cdot)$ está associada, utilizamos a notação $\gamma_X(\cdot)$, $\mathrm{F}_X(\cdot)$ e $f_X(\cdot)$. Nesse caso, também é correto afirmar que $\mathrm{F}_X(\cdot)$ e $f_X(\cdot)$ são, respectivamente, a função de distribuição e densidade espectral do processo $\{X_t\}_{t\in\mathbb{Z}}$. Além disso, as seguintes propriedades são válidas (veja, por exemplo, Brockwell e Davis, 1991; Wei, 1990),

- Se a função densidade espectral $f_X(\cdot)$ existe ela é uma função real contínua, não negativa e periódica, com período igual a 2π, tal que $f_X(-\lambda) = f_X(\lambda)$, para todo $\lambda \in [0,\pi]$, e $\int_{-\pi}^{\pi} f_X(\lambda)\, d\lambda < \infty$.
- Se $\mathrm{F}_X(\cdot)$ é absolutamente contínua então $f_X(\cdot)$ existe e satisfaz

$$f_X(\lambda) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\gamma_X(h)e^{-\mathrm{i}\lambda h} = \frac{1}{2\pi}\Big[\gamma_X(0) + 2\sum_{h=1}^{\infty}\gamma_X(h)\cos(\lambda h)\Big], \tag{2.7}$$

  para todo $\lambda \in [0,\pi]$.
- Pelo Teorema de Herglotz e pelo item anterior, é possível concluir que as funções densidade espectral e de autocovariância do processo estacionário $\{X_t\}_{t\in\mathbb{Z}}$ formam um par pela transformada de Fourier, com uma sendo unicamente determinada pela outra.

No que segue, apresentamos a definição da função periodograma. Se $\{X_t\}_{t\in\mathbb{Z}}$ é estacionário, o periodograma é um estimador não-viciado mas inconsistente para a função densidade espectral $f_X(\cdot)$. Além disso, embora a função densidade espectral não esteja definida para processos $\{X_t\}_{t\in\mathbb{Z}}$ não estacionários, a função periodograma $I_n(\cdot)$, calculada nas frequências de Fourier $\lambda_j = \frac{2\pi j}{n}$, para todo $j \in \{0,\cdots,\lfloor n/2\rfloor\}$, está sempre bem definida (Crato, 1996).

Definição 2.6. (Função Periodograma). Seja $\{X_t\}_{t=1}^n$ uma série temporal. Então, a *função periodograma* de $\{X_t\}_{t=1}^n$ é definida como

$$I_n(\lambda) = \frac{1}{2\pi n}\Big|\sum_{t=1}^{n} X_t e^{-\mathrm{i}\lambda t}\Big|^2 = J(-\lambda)J(\lambda) \quad \text{e} \quad I_n(-\lambda) := I_n(\lambda), \text{ para todo } \lambda \in [0,\pi], \tag{2.8}$$

onde $\quad J(\lambda) := \dfrac{1}{\sqrt{2\pi n}}\displaystyle\sum_{t=1}^{n} X_t e^{-\mathrm{i}\lambda t}.$

Embora a função periodograma seja definida para todo $\lambda \in [0,\pi]$, na prática avaliamos esta função apenas para um conjunto discreto de frequências pois obter os valores de $I_n(\lambda)$ para todo $\lambda \in [0,\pi]$ é computacionalmente impossível. Em geral, $I_n(\cdot)$ é calculado apenas para as frequências de Fourier $\lambda_j = \frac{2\pi j}{n}$, para $j \in \{0,\cdots,\lfloor n/2\rfloor\}$. Alguns autores (veja, por exemplo, Brockwell e Davis, 1991; Crato, 1996) optam por definir a função periodograma apenas nessas frequências e extender a definição para as demais frequências $\lambda \in [0,\pi]$ através da relação

$$I_{n,j}(\lambda) := I_n(\lambda_j), \quad \lambda_j - \frac{\pi}{n} < \lambda \leq \lambda_j + \frac{\pi}{n}, \quad \text{para} \quad j \in \Big\{0,1,\cdots,\Big\lfloor\frac{n}{2}\Big\rfloor\Big\},$$

com $I_n(\lambda_j)$ dado pela expressão (2.8).

Observamos que, para as frequências de Fourier, ambas as definições coincidem. Além disso para tais frequências (com excessão de $\lambda_0 = 0$) a seguinte igualdade é válida (veja Brockwell e Davis, 1991, proposição 10.1.2.)

$$I_n(\lambda_j) = \frac{1}{2\pi}\Big[\hat{\gamma}_X(0) + 2\sum_{h=1}^{n-1}\hat{\gamma}_X(h)\cos(\lambda_j h)\Big], \quad \text{para todo } \lambda_j = \frac{2\pi j}{n}, \quad j \in \{1,\cdots,\lfloor n/2\rfloor\},$$

onde $\hat{\gamma}_X(\cdot)$ é o estimador amostral da função de autocovariância $\gamma_X(\cdot)$, dado por

$$\hat{\gamma}_X(h) = \begin{cases} \frac{1}{n}\sum_{t=1}^{n-h}(X_{t+h} - \bar{X})(X_t - \bar{X}), & \text{se} \quad 0 \leq h < n, \\ \hat{\gamma}_X(-h), & \text{se} \quad -n < h \leq 0, \end{cases}$$

e $\bar{X}$ é a *média amostral*, definida como $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

2.2.4 Processos com Longa Dependência

A presença de longa dependência em séries temporais foi primeiramente observada por Hurst (1951, 1957), ao analisar a série temporal de níveis do rio Nilo, e posteriormente por Mandelbrot e Wallis (1968) e McLeod e Hipel (1978), também na área de hidrologia.

Os primeiros modelos para séries temporais com longa dependência foram introduzidos por Mandelbrot (1965), Mandelbrot e Van Ness (1968) e Mandelbrot e Wallis (1968, 1969), no caso de tempo contínuo, e por Granger e Joyeux (1980) e Hosking (1981), no caso de tempo discreto.

Recentemente, os economistas observaram que há evidências de que processos com longa dependência descrevem, de modo satisfatório, dados econômicos e financeiros tais como taxas de juros e inflação. Entretanto, para tais séries temporais, a longa dependência não é observada na média e sim na volatilidade. Para modelar tal característica foram introduzidos modelos não-lineares tais como o FIGARCH (Baillie et al., 1996), o FIEGARCH (Bollerslev e Mikkelsen, 1996) e o modelo LMSV (Breidt et al., 1998).

Prass (2008), Lopes e Prass (2013a) e Prass e Lopes (2013) apresentam um estudo detalhado dos processos FIEGARCH envolvendo suas propriedades teóricas, simulação e análise de séries temporais reais. Lopes (2008) trata da estimação dos parâmetros do modelo e da previsão de valores futuros para processos estocásticos que apresentam longa dependência, tanto na média, quanto na volatilidade.

A propriedade de longa dependência para um processo estocástico fracamente estacionário pode ser definida de diferentes maneiras, podendo ser no domínio do tempo ou no domínio da frequência (equivalentemente). Uma abordagem mais formal para a definição de longa dependência é dada a seguir.

Definição 2.7. (Longa Dependência para Variância Finita). Seja $\{X_t\}_{t\in\mathbb{Z}}$ um processo estocástico fracamente estacionário com função de autocorrelação $\rho_X(\cdot)$ e função densidade espectral $f_X(\cdot)$. No domínio do tempo, dizemos que $\{X_t\}_{t\in\mathbb{Z}}$ apresenta *longa dependência* se existem constantes reais $a \in (0,1)$ e $c_1 \neq 0$ tais que

$$\rho_X(h) \sim c_1 h^{-a}, \quad \text{quando } h \to \infty.$$

Equivalentemente, no domínio da frequência, dizemos que $\{X_t\}_{t\in\mathbb{Z}}$ apresenta *longa dependência* se existem constantes reais $b \in (0,1)$ e $c_2 > 0$ tais que

$$f_X(\lambda) \sim c_2 |\lambda|^{-b}, \quad \text{quando } \lambda \to 0.$$

Note que o decaimento hiperbólico da função de autocorrelação de um processo estocástico fracamente estacionário com longa dependência implica que

$$\sum_{h=-m}^{m} |\rho_X(h)| \to \infty, \quad \text{quando } m \to \infty.$$

Sendo assim, essa propriedade é muitas vezes utilizada na literatura para caracterizar longa dependência.

Outra conclusão importante, que segue imediatamente da expressão (2.7) e da Definição 2.7, é que, se $f_X(\cdot)$ existe, então,

$$f_X(0) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\gamma_X(h) = \infty.$$

Portanto, para processos estocásticos com longa dependência, a função densidade espectral possui um pólo na frequência $\lambda = 0$.

Observamos ainda que na literatura encontramos definições alternativas para longa dependência. Uma definição amplamente utilizada é dizer que o processo $\{X_t\}_{t\in\mathbb{Z}}$ apresenta longa dependência se

$$\rho_X(h) \sim L_1(h)h^{-a}, \quad \text{quando } h \to \infty, \tag{2.9}$$

onde $a \in (0,1)$ e $L_1 : \mathbb{R} \to \mathbb{R}$ é uma função de variação lenta no infinito, isto é, L_1 é uma função limitada em um intervalo finito, mensurável e

$$\lim_{x\to\infty}\frac{L_1(cx)}{L_1(x)} = 1, \quad \text{para todo } c > 0.$$

Ou, equivalentemente, se $f_X(\lambda) \sim L_2(\lambda)h^{-\beta}$, quando $\lambda \to 0$, onde $\beta \in (0,1)$ e L_2 é uma função de variação lenta no zero. Note que a Definição 2.7 é um caso particular dessa definição com $L(x) \equiv c_1$, para todo $x \in \mathbb{R}$. Embora a expressão (2.9) seja mais geral, a Definição 2.7 cobre todos os processos estudados neste trabalho e, portanto, será a definição que utilizaremos.

Considere o processo estocástico $\{X_t\}_{t\in\mathbb{Z}}$, definido por

$$X_t = \sum_{k=0}^{\infty} c_k\varepsilon_{t-k}, \quad \text{para todo} \quad t \in \mathbb{Z}, \tag{2.10}$$

onde $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ é uma sequência de variáveis aleatórias i.i.d., com $\varepsilon_0 \sim \mathcal{N}(0,1)$, e $\{c_k\}_{k\in\mathbb{N}}$ é uma sequência de constantes reais tais que o processo $\{X_t\}_{t\in\mathbb{Z}}$ esteja bem definido.

Note que, como $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ é uma sequência de variáveis aleatórias i.i.d., para que o processo $\{X_t\}_{t\in\mathbb{Z}}$ seja estritamente (e fracamente) estacionário é suficiente que $\{c_k\}_{k\in\mathbb{N}} \in \ell^2$ (veja Stout, 1974, página 182). Nese caso, a função de autocovariância $\gamma_X(\cdot)$ de $\{X_t\}_{t\in\mathbb{Z}}$ é dada por

$$\gamma_X(h) = \mathrm{Cov}(X_t, X_{t+h}) = \mathrm{Cov}\Big(\sum_{k=0}^{\infty} c_k\varepsilon_{t-k}, \sum_{j=0}^{\infty} c_j\varepsilon_{t+h-j}\Big) = \sum_{k=0}^{\infty} c_k c_{k+h},$$

para todo $h \in \mathbb{Z}$. Consequentemente, a função de autocorrelação $\rho_X(\cdot)$ do processo $\{X_t\}_{t\in\mathbb{Z}}$ é dada por

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \frac{\sum_{k=0}^{\infty} c_k c_{k+h}}{\sum_{k=0}^{\infty} c_k^2}, \quad \text{para todo } h \in \mathbb{Z}.$$

Assim, se os coeficientes $\{c_k\}_{k\in\mathbb{N}}$ são tais que a função $\rho_X(\cdot)$ apresenta decaimento hiperbólico então, o processo $\{X_t\}_{t\in\mathbb{Z}}$ apresenta a característica de longa dependência.

Tendo em vista que longa dependência foi definida em termos do comportamento assintótico da função de autocorrelação e/ou da função densidade espectral e que para processos estocásticos com variância infinita tais funções não estão definidas, levanta-se a seguinte questão: "*como definir longa dependência para um processo estocástico com variância infinita?*". No que segue, apresentamos uma construção que nos permite responder essa questão. Tal construção baseia-se na discussão apresentada em Kokoszka e Taqqu (1995).

Considere um sequência $\{Z_t\}_{t\in\mathbb{Z}}$ de variáveis aleatórias i.i.d., com média zero e variância infinita e defina o processo estocástico $\{Y_t\}_{t\in\mathbb{Z}}$ através da expressão

$$Y_t = \sum_{k=0}^{\infty} c_k Z_{t-k}, \quad \text{para todo} \quad t \in \mathbb{Z}, \tag{2.11}$$

onde $\{c_k\}_{k\in\mathbb{N}}$ são os mesmos coeficientes da expressão (2.10). Obviamente, para que ambos os processos estocásticos $\{X_t\}_{t\in\mathbb{Z}}$ e $\{Y_t\}_{t\in\mathbb{Z}}$ estejam bem definidos, os coeficientes $\{c_k\}_{k\in\mathbb{N}}$ devem satisfazer certas restrições, que podem variar de acordo com as hipóteses assumidas para o processo $\{Z_t\}_{t\in\mathbb{Z}}$ (para mais detalhes, veja a Seção 4.3 do Capítulo 4).

Como longa dependência foi definida em termos da função de autocovariância ou da função densidade espectral, é necessário esclarecer o sentido de longa dependência, no contexto da expressão (2.11), onde as variáveis aleatórias do processo $\{Z_t\}_{t\in\mathbb{Z}}$ apresentam variância infinita. Tal resultado é apresentado na definição que segue.

Definição 2.8. (Longa Dependência para Variância Infinita). Sejam $\{X_t\}_{t\in\mathbb{Z}}$ e $\{Y_t\}_{t\in\mathbb{Z}}$ os processos estocásticos definidos pelas expressões (2.10) e (2.11), respectivamente. Então, diremos que o processo estocástico $\{Y_t\}_{t\in\mathbb{Z}}$, com variância infinita, apresenta *longa dependência* se o correspondente processo $\{X_t\}_{t\in\mathbb{Z}}$, com variância finita, apresenta longa dependência.

Capítulo 3

Distribuição α-Estável

A teoria de distribuições estáveis univariadas foi desenvolvida, essencialmente, em torno dos anos 1920 e 1930 por Paul Lévy e Aleksabder Yakovlevich Khinchine. Um estudo detalhado sobre a teoria pode ser econtrado em clássicos como Gnedenko e Kolmogorov (1954) e Feller (1971) ou em trabalhos mais recentes como Zolotarev (1986) e Samorodnitsky e Taqqu (1994).

Na literatura encontramos diferentes maneiras para caracterizar uma variável aleatória com distribuição estável. Na Definição 3.1 apresentamos quatro definições que são equivalentes (para mais detalhes veja Samorodnitsky e Taqqu, 1994).

Definição 3.1. (Distribuição Estável). Seja X uma variável aleatória. Então, dizemos que X tem *distribuição estável* se,

(a) para quaisquer A e B positivos, existem constantes $C > 0$ e $D \in \mathbb{R}$ tais que

$$AX_1 + BX_2 \stackrel{d}{=} CX + D, \tag{3.1}$$

onde X_1 e X_2 são cópias independentes de X.

(b) para todo $n \geq 2$ existem $C_n > 0$ e $D_n \in \mathbb{R}$ tais que

$$X_1 + \cdots + X_n \stackrel{d}{=} C_n X + D_n, \tag{3.2}$$

onde $X_1, \cdots, X_n$ são cópias independentes de X.

(c) X possui um domínio de atração;

(d) existem parâmetros $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\sigma \geq 0$ e $\mu \in \mathbb{R}$ tais que a função característica de X é dada por

$$\varphi_X(t) = \begin{cases} \exp\left\{\mathrm{i}\mu t - \sigma^\alpha |t|^\alpha \left[1 - \mathrm{i}\beta \operatorname{sgn}(t) \tan(\frac{\pi\alpha}{2})\right]\right\}, & \text{se } \alpha \neq 1, \\ \exp\left\{\mathrm{i}\mu t - \sigma|t| \left[1 + \mathrm{i}\beta \frac{2}{\pi} \operatorname{sgn}(t) \ln(|t|)\right]\right\}, & \text{se } \alpha = 1, \end{cases} \tag{3.3}$$

onde $\operatorname{sgn}(\cdot)$ é a função sinal dada por (2.1).

Observação 3.1. Feller (1971) prova que, para qualquer variável aleatória X, com distribuição estável, existe um número $\alpha \in (0, 2]$ tal que a constante C dada na expressão (3.1) satisfaz

$$C^\alpha = A^\alpha + B^\alpha. \tag{3.4}$$

O parâmetro α da expressão (3.4) coincide com o parâmetro α que aparece na Definição 3.1. Zolotarev (1986) prova que, se X é uma variável aleatória que satisfaz a expressão (3.2), para $n = 2$ e $n = 3$, então a função de distribuição de X é necessariamente estável.

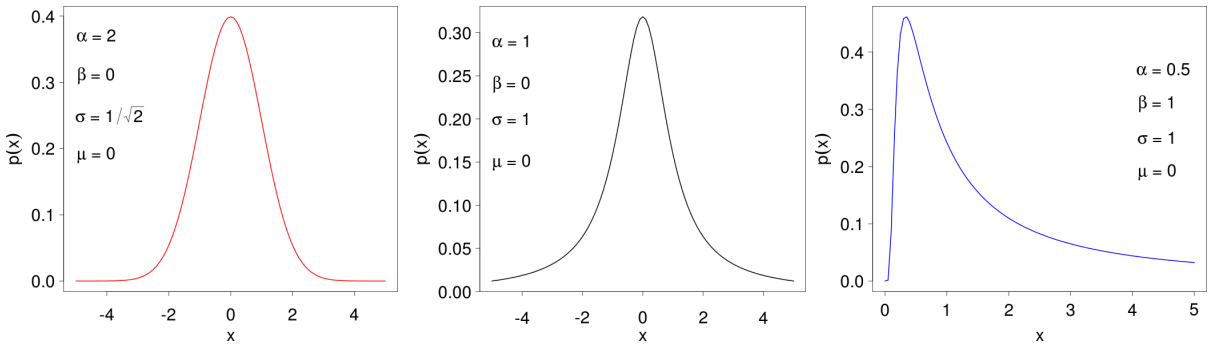
Observação 3.2. O parâmetro α é o *índice de estabilidade*, também conhecido na literatura como *expoente característico* ou *índice da cauda*. Os coefficientes β, σ e μ são conhecidos, respectivamente como *parâmetro de simetria*, *parâmetro de escala* e *parâmetro de locação*. Além disso, utilizaremos a notação $X \sim S_\alpha(\beta, \sigma, \mu)$ para dizer que a variável X tem função de distribuição estável com parâmetros α, β, σ e μ.

Observação 3.3. Note que, se $\alpha = 2$, β é irrelevante pois $\tan(\frac{\pi\alpha}{2}) = \tan(\pi) = 0$. Portanto, sem perda de generalidade, à $\alpha = 2$ associa-se $\beta = 0$.

Ressaltamos que na literatura existem somente três casos conhecidos para os quais a função de densidade de probabilidade estável possui fórmula fechada. São eles

- quando $\alpha = 2$. Nesse caso a distribuição $S_\alpha(\beta, \sigma, \mu)$ coincide com a distribuição Gaussiana com média μ e variância $2\sigma^2$.
- quando $\alpha = 1$ e $\beta = 0$. Nesse caso temos a distribuição de Cauchy com parâmetros de locação μ e de escala σ.
- quando $\alpha = 0.5$ e $\beta = 1$. Nesse caso temos a distribuição de Lévy com parâmetros de locação μ e escala σ.

Na Figura 3.1 apresentamos exemplos da função densidade de probabilidade estável para os casos mencionados acima. Na Figura 3.1(a) consideramos $\alpha = 2$, $\beta = \mu = 0$ e $\sigma = 1/\sqrt{2}$, ou seja, uma Gaussiana padrão. Na Figura 3.1(b) fixamos $\alpha = 1$, $\beta = \mu = 0$ e $\sigma = 1$, ou seja, uma Cauchy padrão. Na Figura 3.1(c) assumimos $\alpha = 0.5$, $\beta = 1$, $\sigma = 1$ e $\mu = 0$, ou seja, uma Lévy padrão.



(a) $S_2(0, 1/\sqrt{2}, 0)$ **(b)** $S_1(0, 1, 0)$ **(c)** $S_{0.5}(1, 1, 0)$

Figura 3.1: Função Densidade de Probabilidade: (a) Gaussiana com média zero e desvio padrão 1; (b) Cauchy; (c) Lévy.

Neste trabalho adotamos as seguintes definições, comumente utilizadas na literatura.

- Dizemos que X é uma variável estável se X tem distribuição estável. Uma variável estável com índice α é chamada α-*estável*.
- Dizemos que a variável aleatória X é *estritamente estável* se $D = 0$ na expressão (3.1). Ressaltamos que (veja Samorodnitsky e Taqqu, 1994),
  - se $X \sim S_\alpha(\beta, \sigma, \mu)$ e $\alpha \neq 1$, então X é estritamente estável se, e somente se, $\mu = 0$;
  - se $\alpha \neq 1$ e $\mu \neq 0$, então $X - \mu$ é estritamente estável (consequência imediata do item anterior);
  - se $\alpha = 1$, X é estritamente estável se, e somente se, $\beta = 0$.

- Dizemos que a variável aleatória X é estável simétrica se X e $-X$ tem a mesma distribuição.
- Utiliza-se a notação $X \sim S\alpha S$ para dizer que a variável aleatória X é α-estável simétrica. Mostra-se que (veja Samorodnitsky e Taqqu, 1994)
  - $X \sim S_\alpha(\beta, \sigma, \mu)$ é simétrica se, e somente se, $\beta = 0$ e $\mu = 0$;
  - X é simétrica em torno de μ se, e somente se, $\beta = 0$;
  - se $X \sim S\alpha S$, então a função característica de X é dada por

$$\varphi_X(t) = e^{-\sigma^\alpha |t|^\alpha}, \quad \text{todo } t \in \mathbb{R} \quad \text{e qualquer } \alpha \in (0, 2].$$

3.1 Propriedades das Variáveis Aleatórias Estáveis

Na proposição que segue apresentamos importantes propriedades das variáveis aleatórias estáveis. Tais propriedades são utilizadas na prova do Lema 3.1.

Proposição 3.1. *Sejam X, X_1 e X_2 variáveis aleatórias, com $X \sim S_\alpha(\beta, \sigma, \mu)$ e $X_i \sim S_\alpha(\beta_i, \sigma_i, \mu_i)$, para $i = 1, 2$. Então, as seguintes afirmações são verdadeiras*

(i) $X_1 + X_2 \sim S_\alpha(\beta, \sigma, \mu)$, *onde*

$$\sigma = (\sigma_1^\alpha + \sigma_2^\alpha)^{\frac{1}{\alpha}}, \quad \beta = \frac{\beta_1 \sigma_1^\alpha + \beta_2 \sigma_2^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha} \quad e \quad \mu = \mu_1 + \mu_2.$$

(ii) $X + a \sim S_\alpha(\beta, \sigma, \mu + a)$, *para qualquer $a \in \mathbb{R}$.*

(iii) *Para qualquer real $a \neq 0$*

$$\begin{array}{ll} aX \sim S_\alpha(\mathrm{sgn}(a)\beta, |a|\sigma, a\mu), & se \quad \alpha \neq 1, \\ aX \sim S_1\left(\mathrm{sgn}(a)\beta, |a|\sigma, a\mu - \frac{2}{\pi} a \ln(|a|)\sigma\beta\right), & se \quad \alpha = 1. \end{array}$$

(iv) *Para qualquer $0 < \alpha < 2$,*

$$X \sim S_\alpha(\beta, \sigma, 0) \iff -X \sim S_\alpha(-\beta, \sigma, 0).$$

Prova: Veja Samorodnitsky e Taqqu (1994). ∎

Como consequência direta da Proposição 3.1 temos,

- quando $\alpha \neq 1$,

$$X \sim S_\alpha(\beta, \sigma, \mu) \iff \frac{X - \mu}{\sigma} \sim S_\alpha(\beta, 1, 0);$$

- quando $\alpha = 1$,

$$X \sim S_1(\beta, \sigma, \mu) \iff \frac{X}{\sigma} \sim S_1\left(\beta, 1, \left[\frac{\mu}{\sigma} - \frac{2}{\pi}\ln\left(\frac{1}{\sigma}\right)\beta\right]\right) \iff \frac{X}{\sigma} \sim S_1\left(\beta, 1, \left[\frac{\mu}{\sigma} + \beta\frac{2}{\pi}\ln(\sigma)\right]\right)$$
$$\iff \frac{X - [\mu + \beta\frac{2}{\pi}\sigma\ln(\sigma)]}{\sigma} \sim S_1(\beta, 1, 0).$$

Ou seja,

$$X \sim S_\alpha(\beta, \sigma, \mu) \iff \frac{X - \mu^*}{\sigma} \sim S_\alpha(\beta, 1, 0), \quad \text{com} \quad \mu^* = \begin{cases} \mu, & \text{se} \quad \alpha \neq 1, \\ \mu + \beta\frac{2}{\pi}\sigma\ln(\sigma), & \text{se} \quad \alpha = 1. \end{cases} \tag{3.5}$$

Mais especificamente, pelas propriedades da função densidade de probabilidade de uma variável aleatória, segue que

$$p_X(x;\alpha,\beta,\sigma,\mu) = \frac{1}{\sigma} p_{X^*}\Big(\frac{x-\mu^*}{\sigma};\alpha,\beta,1,0\Big), \quad \text{para todo } x \in \mathbb{R},$$

onde $p_X(\cdot;\alpha,\beta,\sigma,\mu)$ e $p_{X^*}(\cdot;\alpha\beta,1,0)$ são as funções densidade de probabilidade das variáveis aleatórias $X \sim S_\alpha(\beta,\sigma,\mu)$ e $X^* \sim S_\alpha(\beta,1,0)$, respectivamente.

Por simplicidade de notação, definimos

$$p_X(x;\alpha,\beta) := p_X(x;\alpha,\beta,1,0), \quad \text{para todo } x \in \mathbb{R}.$$

No Lema 3.1, que segue, apresentamos uma fórmula fechada para os parâmetros da distribuição de uma combinação linear finita de variáveis aleatórias α-estáveis.

Lema 3.1. *Sejam $\{a_t\}_{t\in\mathbb{N}}$ uma sequência de constantes reais e $\{X_t\}_{t\in\mathbb{N}}$ uma sequência de variáveis aleatórias independentes, com $X_t \sim S_\alpha(\beta_t,\sigma_t,\mu_t)$, para cada $t \in \mathbb{N}$. Então, para todo $n \geq 1$,*

$$\sum_{t=1}^{n} a_t X_t \sim S_\alpha(\beta_{(n)},\sigma_{(n)},\mu_{(n)}),$$

onde,

$$\mu_{(n)} = \sum_{t=1}^{n}\mu(t), \quad \sigma_{(n)} = \Big(\sum_{t=1}^{n}(|a_t|\sigma_t)^\alpha\Big)^{\frac{1}{\alpha}} \quad e \quad \beta_{(n)} = \frac{\sum_{t=1}^{n}|a_t|^\alpha \mathrm{sgn}(a_t)\beta_t\sigma_t^\alpha}{\sum_{t=1}^{n}(|a_t|\sigma_t)^\alpha},$$

$$com\ \mu(t) = \begin{cases} a_t\mu_t, & se \quad \alpha \neq 1, \\ a_t\mu_t - \frac{2}{\pi}a_t\ln(|a_t|)\sigma_t\beta_t, & se \quad \alpha = 1, \end{cases} \quad para\ cada \quad t \in \mathbb{N}.$$

Prova: Para a prova utilizamos indução em n. Note que, para $n = 1$ e $n = 2$ o resultado segue de imediato da Proposição 3.1.

Suponha que o resultado é válido para $n = k$, isto é, $\sum_{t=1}^{k} a_t X_t \sim S_\alpha(\beta_{(k)},\sigma_{(k)},\mu_{(k)})$, com

$$\mu_{(k)} = \sum_{t=1}^{k}\mu(t), \quad \sigma_{(k)} = \Big(\sum_{t=1}^{k}(|a_t|\sigma_t)^\alpha\Big)^{\frac{1}{\alpha}} \quad \text{e} \quad \beta_{(k)} = \frac{\sum_{t=1}^{k}|a_t|^\alpha \mathrm{sgn}(a_t)\beta_t\sigma_t^\alpha}{\sum_{t=1}^{k}(|a_t|\sigma_t)^\alpha}. \tag{3.6}$$

Pela Proposição 3.1, $a_{k+1}X_{k+1} \sim S_\alpha\big(\mathrm{sgn}(a_{k+1})\beta_{k+1}, |a_{k+1}|\sigma_{k+1}, \mu(k+1)\big)$ e $a_{k+1}X_{k+1} + \sum_{t=1}^{k} a_t X_t \sim S_\alpha(\beta,\sigma,\mu)$, onde

$$\begin{aligned} \mu &= \mu(k+1) + \mu_{(k)}, \quad \sigma = \big[(|a_{k+1}|\sigma_{k+1})^\alpha + (\sigma_{(k)})^\alpha\big]^{\frac{1}{\alpha}}, \\ \beta &= \frac{\mathrm{sgn}(a_{k+1})\beta_{k+1}(|a_{k+1}|\sigma_{k+1})^\alpha + \beta_{(k)}(\sigma_{(k)})^\alpha}{(|a_{k+1}|\sigma_{k+1})^\alpha + (\sigma_{(k)})^\alpha}. \end{aligned} \tag{3.7}$$

Além disso,

$$(\sigma_{(k)})^\alpha = \sum_{t=1}^{k}(|a_t|\sigma_t)^\alpha \quad \Longrightarrow \quad \beta_{(k)}(\sigma_{(k)})^\alpha = \sum_{t=1}^{k}|a_t|^\alpha \mathrm{sgn}(a_t)\beta_t\sigma_t^\alpha. \tag{3.8}$$

Portanto, substituindo-se as igualdades (3.6) nas equações (3.7) e utilizando-se (3.8), obtemos $\mu = \mu_{(k+1)}$, $\sigma = \sigma_{(k+1)}$ e $\beta = \beta_{(k+1)}$ e o resultado segue para $n = k+1$. ∎

O Teorema 3.1 fornece expressões para as constantes C_n e D_n dadas na expressão (3.2). Uma prova alternativa para esse teorema, que não envolve o Lema 3.1, pode ser encontrada em Feller (1971).

Teorema 3.1. *Sejam C_n e D_n as constantes dadas na expressão (3.2). Então,*

$$C_n = n^{1/\alpha} \quad e \quad D_n = \begin{cases} \mu(n - n^{1/\alpha}), & se \quad \alpha \neq 1, \\ \frac{2}{\pi} n \ln(n)\sigma\beta, & se \quad \alpha = 1, \end{cases}$$

onde $\alpha \in [0, 2]$, $\beta \in [-1, 1]$ e $\sigma \geq 0$ são constantes reais.

Prova: Sejam $X, X_1, X_2, \cdots$ i.i.d. $S_\alpha(\beta, \sigma, \mu)$. Então, pelo Lema 3.1,

$$X_1 + \cdots + X_n \sim S_\alpha(\beta_{(n)}, \sigma_{(n)}, \mu_{(n)}), \quad \text{para todo} \quad n \geq 1,$$

onde

$$\beta_{(n)} = \frac{n\beta\sigma^\alpha}{n\sigma^\alpha} = \beta, \quad \sigma_{(n)} = (n\sigma^\alpha)^{1/\alpha} = n^{1/\alpha}\sigma \quad \text{e} \quad \mu_{(n)} = n\mu.$$

Note que, quando $\alpha = 1$, a igualdade $\mu_{(n)} = n\mu$ segue do fato que $a_t = 1$, logo $\ln(a_t) = 0$, para todo $1 \leq t \leq n$. Além disso, definindo-se $S_n := X_1 + \cdots + X_n$, segue que,

- se $\alpha \neq 1$, como $X \sim S_\alpha(\beta, \sigma, \mu)$ e $S_n \sim S_\alpha(\beta, n^{1/\alpha}\sigma, n\mu)$, pela Proposição 3.1 e pela expressão (3.5), segue que
$$\frac{X - \mu}{\sigma} \sim S_\alpha(\beta, 1, 0) \quad \text{e} \quad \frac{S_n - n\mu}{n^{1/\alpha}\sigma} \sim S_\alpha(\beta, 1, 0).$$
Logo,
$$\frac{S_n - n\mu}{n^{1/\alpha}\sigma} \overset{d}{=} \frac{X - \mu}{\sigma} \iff S_n - n\mu \overset{d}{=} n^{1/\alpha}\sigma\left(\frac{X - \mu}{\sigma}\right) \iff S_n \overset{d}{=} n^{1/\alpha}(X - \mu) + n\mu$$
$$\iff S_n \overset{d}{=} n^{1/\alpha}X + \mu(n - n^{1/\alpha}). \tag{3.9}$$

- se $\alpha = 1$ então, $X \sim S_1(\beta, \sigma, \mu)$, $S_n \sim S_1(\beta, n\sigma, n\mu)$ mas, pela expressão (3.5), a equivalência dada na expressão (3.9) não é satisfeita. Entretanto, pela Proposição 3.1, temos que
$$X \sim S_1(\beta, \sigma, \mu) \iff nX \sim S_1\Big(\beta, n\sigma, n\mu - \frac{2}{\pi} n \ln(n)\sigma\beta\Big) \iff nX + \frac{2}{\pi} n \ln(n)\sigma\beta \sim S_1\Big(\beta, n\sigma, n\mu\Big)$$
$$\iff S_n \overset{d}{=} nX + \frac{2}{\pi} n \ln(n)\sigma\beta. \tag{3.10}$$

Portanto, das expressões (3.9) e (3.10), segue que,

$$S_n = X_1 + \cdots + X_n \overset{d}{=} \begin{cases} n^{1/\alpha}X + \mu(n - n^{1/\alpha}), & \text{se} \quad \alpha \neq 1, \\ nX + \frac{2}{\pi} n \ln(n)\sigma\beta, & \text{se} \quad \alpha = 1, \end{cases}$$

e o resultado enunciado segue. ∎

Por meio de um argumento semelhante ao do teorema central do limite prova-se o seguinte resultado.

Teorema 3.2. *Seja $X \sim S_\alpha(\beta, \sigma, \mu)$, com $\alpha \in (0, 2)$. Então,*

$$\lim_{x \to \infty} x^\alpha \mathbb{P}(X > x) = C_\alpha(1 + \beta)\sigma^\alpha \quad e \quad \lim_{x \to \infty} x^\alpha \mathbb{P}(X < -x) = C_\alpha(1 - \beta)\sigma^\alpha,$$

onde

$$C_\alpha = \left(\int_0^\infty x^{-\alpha} \sin(x) dx\right)^{-1} = \begin{cases} \left[\Gamma(1 - \alpha)\cos\left(\frac{\pi\alpha}{2}\right)\right]^{-1}, & se \quad \alpha \neq 1, \\ \frac{2}{\pi}, & se \quad \alpha = 1. \end{cases}$$

Prova: Veja Samorodnitsky e Taqqu (1994, página 16). ∎

O Teorema 3.2 é de grande importância pois determina o comportamento assintótico das caudas da distribuição α-estável. A proposição que segue é decorrente do Teorema 3.2 e nos fornece informações sobre a existência dos momentos de order r, para todo $r > 0$, de uma variável aleatória estável.

Proposição 3.2. *Seja $X \sim S_\alpha(\beta, \sigma, \mu)$, com $\alpha \in (0, 2)$. Então,*

$$\mathbb{E}\big(|X|^r\big) < \infty, \quad \textit{para todo} \quad 0 < r < \alpha,$$
$$\mathbb{E}\big(|X|^r\big) = \infty, \quad \textit{para todo} \quad r \geq \alpha.$$

Prova: Veja Samorodnitsky e Taqqu (1994). ∎

Da Proposição 3.2 concluimos que, se $\alpha \leq 1$, todos os momentos da variável aleatória X são infinitos. São fatos conhecidos na literatura que, se $\alpha > 1$ a média da distribuição existe e é igual a μ; se $\beta = 0$, a distribuição é simétrica em torno de μ; se $\beta > 0$, a distribuição tem a cauda direita mais pesada e, se $\beta < 0$, a distribuição tem a cauda esquerda mais pesada.

3.1.1 Processos Estocásticos Estáveis

Antes de definir um processo estocástico estável apresentamos algumas definições relacionadas a estabilidade em $\mathbb{R}^m$. Tais definições são importantes pois são utilizadas nas provas dos teoremas apresentados nesta subseção.

Definição 3.2. (Vetor Aleatório Estável). Um vetor aleatório $\mathbf{X} = (X_1, \cdots, X_m)'$ em $\mathbb{R}^m$ é dito estável se, para quaisquer A e B positivos, existem $C \in \mathbb{R}$, $C > 0$, e um vetor $\mathbf{D} \in \mathbb{R}^m$ tais que

$$A\mathbf{X}^{(1)} + B\mathbf{X}^{(2)} \stackrel{d}{=} C\mathbf{X} + \mathbf{D}, \tag{3.11}$$

onde $\mathbf{X}^{(1)}$ e $\mathbf{X}^{(2)}$ são cópias independentes de $\mathbf{X}$.

Seja $\mathbf{X} = (X_1, \cdots, X_m)'$ um vetor aleatório estável em $\mathbb{R}^m$. Então,

- as seguintes expressões são comuns na literatura: "$X_1, \cdots, X_m$ são conjuntamente estáveis", "$\mathbf{X}$ tem distribuição estável em $\mathbb{R}^m$", ou ainda, "a distribuição de $\mathbf{X}$ é estável multivariada".
- dizemos que $\mathbf{X}$ é estritamente estável se $\mathbf{D} = \mathbf{0}$ na expressão (3.11).
- $\mathbf{X}$ é dito estável simétrico se ele é estável, de acordo com a Definição 3.2, e satisfaz a igualdade $\mathbb{P}(\mathbf{X} \in A) = \mathbb{P}(-\mathbf{X} \in A)$, para todo boreliano $A \subseteq \mathbb{R}^m$.
- $\mathbf{X}$ é estável se, e somente se, para todo $n \geq 2$, existem $\alpha \in (0, 2]$ e um vetor $\mathbf{D}_n \in \mathbb{R}^m$ tais que

  $$\mathbf{X}^{(1)} + \cdots + \mathbf{X}^{(n)} \stackrel{d}{=} n^{1/\alpha}\mathbf{X} + \mathbf{D}_n,$$

  onde $\mathbf{X}^{(i)}$ é uma cópia independente de $\mathbf{X}$, para todo $i \in \{1, \cdots, n\}$ (veja Samorodnitsky e Taqqu, 1994).

Os Teoremas 3.3 e 3.4 tratam da estabilidade das variáveis aleatórias (quando vistas individualmente) de um vetor aleatório $\mathbf{X} \in \mathbb{R}^m$ estável. Tal resultado é de grande importância para o estudo de processos estocásticos estacionários.

Teorema 3.3. *Seja $\mathbf{X} = (X_1, \cdots, X_m)'$ um vetor aleatório estável (estritamente estável ou estável simétrico) em $\mathbb{R}^m$. Então, existe uma constante $\alpha \in (0, 2]$ tal que, as constantes A, B e C, dadas na expressão (3.11), satisfazem $C = (A^\alpha + B^\alpha)^{1/\alpha}$. Além disso, toda combinação linear, $Y = \sum_{k=1}^m b_k X_k$, de componentes de $\mathbf{X}$, é uma variável aleatória α-estável (respectivamente, estritamente α-estável ou α-estável simétrica).*

Prova: Veja Samorodnitsky e Taqqu (1994, página 58). ∎

Teorema 3.4. *Seja $\mathbf{X} = (X_1, \cdots, X_m)'$ um vetor aleatório em $\mathbb{R}^m$.*

(i) *Se todas as combinações lineares $Y = \sum_{k=1}^m b_k X_k$ têm distribuição estritamente estável, então $\mathbf{X}$ é um vetor aleatório estritamente estável.*

(ii) *Se todas as combinações lineares tem distribuição estável simétrica, então $\mathbf{X}$ é um vetor aleatório estável simétrico.*

(iii) *Se todas as combinações lineares são estáveis, com índice de estabilidade maior ou igual a 1, então $\mathbf{X}$ é um vetor aleatório estável em $\mathbb{R}^m$.*

Prova: Veja Samorodnitsky e Taqqu (1994, página 59). ∎

Observação 3.4. Quando $0 < \alpha < 1$, o fato de que todas as combinações lineares $Y = \sum_{k=1}^m b_k X_k$ serem estáveis não garante a estabilidade do vetor aleatório $\mathbf{X} \in \mathbb{R}^m$ (veja Samorodnitsky e Taqqu, 1994).

No que segue, definimos um processo estocástico estável $\{X_t\}_{t\in \mathrm{T}}$, onde T é um conjunto de índices qualquer, podendo ser discreto ou contínuo.

Definição 3.3. (Processo Estocástico Estável). Um processo estocástico $\{X_t\}_{t\in \mathrm{T}}$ é *estável* se todas as suas distribuições finito-dimensionais são estáveis. Além disso, ele é dito ser *estritamente estável*, ou *estável simétrico*, se todas as suas distribuições finito-dimensionais são, respectivamente, estritamente estáveis ou estáveis simétricas.

Note que, como consequência do Teorema 3.3, se $\{X_t\}_{t\in \mathrm{T}}$ é um processo estocástico estável, então todas as distribuições finito-dimensionais possuem o mesmo índice de estabilidade α. Quando conveniente, diremos então que $\{X_t\}_{t\in \mathrm{T}}$ é um processo estocástico α-estável. Além disso, se $\{X_t\}_{t\in \mathrm{T}}$ é estritamente estacionário, no sentido da Definição 2.5, então as variáveis aleatórias X_t, para todo $t \in T$, são identicamente distribuídas.

O teorema que segue é uma generalização do Lema 3.1 pois trata da combinação linear de variáveis aleatórias de um processo estocástico α-estável.

Teorema 3.5. *Seja $\{X_t\}_{t\in \mathrm{T}}$ um processo estocástico. Então,*

(i) *$\{X_t\}_{t\in \mathrm{T}}$ é estritamente estável se, e somente se, todas as combinações lineares dadas por*

$$\sum_{k=1}^m b_k X_{t_k}, \quad \text{para todo} \quad m \geq 1, \quad t_1, \cdots, t_m \in T \quad e \quad b_1, \cdots, b_m \in \mathbb{R}, \tag{3.12}$$

são estritamente estáveis.

(ii) *$\{X_t\}_{t\in \mathrm{T}}$ é estável simétrico se, e somente se, todas as combinações lineares, dadas pela expressão (3.12), são estáveis simétricas.*

(iii) *Se $\alpha \geq 1$, então $\{X_t\}_{t\in \mathrm{T}}$ é α-estável se, e somente se, todas combinações lineares, definidas pela expressão (3.12), são α-estáveis.*

Prova: O resultado é consequência imediata dos Teoremas 3.3 e 3.4. ∎

3.2 Função Característica e Parametrizações

A função densidade de probabilidade de uma variável estável é conhecida em forma fechada apenas para os três casos mencionados na Observação 3.2. Sendo assim, tornou-se comum utilizar a função característica, dada pela expressão (3.3), para descrever distribuições α-estáveis. Entretanto, na literatura encontramos diferentes parametrizações para distribuições α-estáveis e muita confusão é causada devido à elas. A grande variedade de parametrizações é causada por uma combinação de evolução histórica e inúmeros problemas que tem sido analisados e requerem a utilização de representações mais apropriadas da distribuição estável. No que segue apresentamos as parametrizações mais utilizadas na literatura.

Observação 3.5. Assume-se conhecido os seguintes fatos:

(1) $t = |t|\,\mathrm{sgn}(t)$ e $|t|^{\alpha}\mathrm{sgn}(t) = |t|\,\mathrm{sgn}(t)|t|^{\alpha-1} = t|t|^{\alpha-1}$, para todo $t \in \mathbb{R}$;

(2) $\mathrm{sgn}(t) = t/|t|$ e $\ln(\sigma) + \ln(|t|) = \ln(\sigma|t|)$, para todo $t \neq 0$;

(3) por definição, $\mathrm{sgn}(t)\ln(|t|) := 0$, para $t = 0$.

Pela expressão (3.3), para todo $t \in \mathbb{R}$ e $\alpha \neq 1$,

$$\begin{aligned}\varphi_X(t) &= \exp\left\{\mathrm{i}\mu t - \sigma^{\alpha}|t|^{\alpha}\left[1 - \mathrm{i}\beta\,\mathrm{sgn}(t)\tan\left(\frac{\pi\alpha}{2}\right)\right]\right\}\\ &= \exp\left\{\mathrm{i}\mu t + \sigma^{\alpha}\left[-|t|^{\alpha} + \mathrm{i}\beta|t|^{\alpha}\mathrm{sgn}(t)\tan\left(\frac{\pi\alpha}{2}\right)\right]\right\}\\ &= \exp\left\{\mathrm{i}\mu t + \sigma^{\alpha}\left[-|t|^{\alpha} + \mathrm{i}t\left(\beta|t|^{\alpha-1}\tan\left(\frac{\pi\alpha}{2}\right)\right)\right]\right\};\end{aligned}$$

e, para $\alpha = 1$,

$$\begin{aligned}\varphi_X(t) &= \exp\left\{\mathrm{i}\mu t - \sigma|t|\left[1 + \mathrm{i}\beta\,\frac{2}{\pi}\,\mathrm{sgn}(t)\ln(|t|)\right]\right\}\\ &= \exp\left\{\mathrm{i}\mu t + \sigma\left[-|t| - \mathrm{i}\beta\,\frac{2}{\pi}|t|\,\mathrm{sgn}(t)\ln(|t|)\right]\right\}\\ &= \exp\left\{\mathrm{i}\mu t + \sigma\left[-|t| + \mathrm{i}t\left(-\beta\,\frac{2}{\pi}\ln(|t|)\right)\right]\right\}.\end{aligned}$$

Ou seja, a expressão (3.3) pode ser reescrita como

$$\varphi_X(t) = \exp\left\{\mathrm{i}\mu t + \sigma^{\alpha}\left[-|t|^{\alpha} + \mathrm{i}t\omega(t,\alpha,\beta)\right]\right\}, \quad \text{para todo } t \in \mathbb{R}, \tag{3.13}$$

onde

$$\omega(t,\alpha,\beta) = \begin{cases} \beta|t|^{\alpha-1}\tan\left(\frac{\pi\alpha}{2}\right), & \text{se} \quad \alpha \neq 1,\\ -\beta\frac{2}{\pi}\ln(|t|), & \text{se} \quad \alpha = 1.\end{cases}$$

A parametrização dada pela expressão (3.13) será referenciada no que segue como parametrização canônica. A desvantagem da parametrização canônica é que a função $\omega(t,\alpha,\beta)$ é descontínua em $\alpha = 1$, para todo $\beta \neq 0$.

Observação 3.6. A expressão (3.13) é uma pequena variante da parametrização (A) de Zolotarev (veja Zolotarev, 1986, página 9), na qual $\varphi_X(\cdot)$ é dada por

$$\varphi_X(t) = \exp\big\{\lambda_A\big[\mathrm{i}t\gamma_A - |t|^\alpha + \mathrm{i}t\omega_A(t,\alpha,\beta)\big]\big\}, \quad \text{para todo } t\in\mathbb{R}, \tag{3.14}$$

onde $\omega_A(t,\alpha,\beta) = \omega(t,\alpha,\beta)$. Note que, partindo de (3.14), obtemos a parametrização (3.13) tomando $\mu = \lambda_A\gamma_A$ e $\sigma = \lambda_A^{1/\alpha}$.

Partindo da expressão (3.3), com $\alpha\neq 1$, temos

$$\begin{aligned}
\varphi_X(t) &= \exp\left\{\mathrm{i}\mu t - \sigma^\alpha|t|^\alpha\left[1 - \mathrm{i}\beta\,\mathrm{sgn}(t)\tan\left(\frac{\pi\alpha}{2}\right)\right]\right\}\\
&= \exp\left\{\mathrm{i}\mu t + \sigma^\alpha\left[-|t|^\alpha + \mathrm{i}t\beta|t|^{\alpha-1}\tan\left(\frac{\pi\alpha}{2}\right) - \mathrm{i}t\beta\tan\left(\frac{\pi\alpha}{2}\right) + \mathrm{i}t\beta\tan\left(\frac{\pi\alpha}{2}\right)\right]\right\}\\
&= \exp\left\{\mathrm{i}t\left[\mu + \beta\sigma^\alpha\tan\left(\frac{\pi\alpha}{2}\right)\right] + \sigma^\alpha\left[-|t|^\alpha + \mathrm{i}t\beta\big(|t|^{\alpha-1}-1\big)\tan\left(\frac{\pi\alpha}{2}\right)\right]\right\},
\end{aligned} \tag{3.15}$$

para todo $t\in\mathbb{R}$. Segue, pela expressão (3.15), que $\varphi_X(\cdot)$ pode ser reescrita como

$$\varphi_X(t) = \exp\big\{\mathrm{i}\mu_M t + \sigma^\alpha\big[-|t|^\alpha + \mathrm{i}t\omega_M(t,\alpha,\beta)\big]\big\}, \quad \text{para todo } t\in\mathbb{R}, \tag{3.16}$$

com

$$\omega_M(t,\alpha,\beta) := \begin{cases} \beta(|t|^{\alpha-1}-1)\tan\left(\frac{\pi\alpha}{2}\right), & \text{se } \alpha\neq 1,\\ -\beta\frac{2}{\pi}\ln\big(|t|\big), & \text{se } \alpha = 1\end{cases} \quad \text{e} \quad \mu_M := \begin{cases} \mu + \beta\sigma^\alpha\tan\left(\frac{\pi\alpha}{2}\right), & \text{se } \alpha\neq 1,\\ \mu, & \text{se } \alpha = 1.\end{cases}$$

Substituindo-se $\mu_M = \lambda_M\gamma_M$ e $\sigma = \lambda_M^{1/\alpha}$ na expressão (3.16), obtemos a parametrização (M) de Zolotarev (veja Zolotarev, 1986, página 11). Nessa parametrização a função $\omega_M(t,\alpha,\beta)$ é juntamente contínua em α e β.

De maneira semelhante, partindo novamente da expressão (3.3), com $\alpha\neq 1$, temos

$$\begin{aligned}
\varphi_X(t) &= \exp\left\{\mathrm{i}\mu t - \sigma^\alpha|t|^\alpha\left[1 - \mathrm{i}\beta\,\mathrm{sgn}(t)\tan\left(\frac{\pi\alpha}{2}\right)\right]\right\}\\
&= \exp\left\{\mathrm{i}\mu t - \sigma^\alpha|t|^\alpha\left[1 - \mathrm{i}\beta\mathrm{sgn}(t)\tan\left(\frac{\pi\alpha}{2}\right) + \mathrm{i}\beta\mathrm{sgn}(t)(\sigma|t|)^{1-\alpha}\tan\left(\frac{\pi\alpha}{2}\right)\right.\right.\\
&\qquad \left.\left. - \mathrm{i}\beta\mathrm{sgn}(t)(\sigma|t|)^{1-\alpha}\tan\left(\frac{\pi\alpha}{2}\right)\right]\right\}\\
&= \exp\left\{\mathrm{i}\left[\mu + \beta\sigma\tan\left(\frac{\pi\alpha}{2}\right)\right]t - \sigma^\alpha|t|^\alpha\left[1 + \mathrm{i}\,\mathrm{sgn}(t)\beta\big((\sigma|t|)^{1-\alpha}-1\big)\tan\left(\frac{\pi\alpha}{2}\right)\right]\right\}
\end{aligned} \tag{3.17}$$

e ainda, para $\alpha = 1$

$$\begin{aligned}
\varphi_X(t) &= \exp\left\{\mathrm{i}\mu t - \sigma|t|\left[1 + \mathrm{i}\beta\frac{2}{\pi}\mathrm{sgn}(t)\ln(|t|)\right]\right\}\\
&= \exp\left\{\mathrm{i}\mu t - \sigma|t|\left[1 + \mathrm{i}\beta\frac{2}{\pi}\mathrm{sgn}(t)\ln(|t|) + \mathrm{i}\beta\frac{\pi}{2}\mathrm{sgn}(t)\ln(\sigma) - \mathrm{i}\beta\frac{2}{\pi}\mathrm{sgn}(t)\ln(\sigma)\right]\right\}\\
&= \exp\left\{\mathrm{i}\left[\mu + \beta\frac{2}{\pi}\sigma\ln(\sigma)\right]t - \sigma|t|\left[1 + \mathrm{i}\,\mathrm{sgn}(t)\beta\frac{2}{\pi}\ln(\sigma|t|)\right]\right\},
\end{aligned} \tag{3.18}$$

para todo $t\in\mathbb{R}$. Das expressões (3.17) e (3.18) conclui-se que $\varphi_X(\cdot)$ pode ser reescrita como

$$\varphi_X(t) = \exp\big\{\mathrm{i}\mu_0 t - \sigma^\alpha|t|^\alpha\big[1 + \mathrm{i}\,\mathrm{sgn}(t)\,\omega_0(t,\alpha,\beta)\big]\big\}, \quad \text{para todo } t\in\mathbb{R}, \tag{3.19}$$

com

$$\omega_0(t,\alpha,\beta) = \begin{cases} \beta\big((\sigma|t|)^{1-\alpha}-1\big)\tan\left(\frac{\pi\alpha}{2}\right), & \text{se } \alpha\neq 1,\\ \beta\frac{2}{\pi}\ln(\sigma|t|), & \text{se } \alpha = 1\end{cases} \quad \text{e} \quad \mu_0 = \begin{cases} \mu + \beta\sigma\tan\left(\frac{\pi\alpha}{2}\right), & \text{se } \alpha\neq 1,\\ \mu + \beta\sigma\frac{2}{\pi}\ln(\sigma), & \text{se } \alpha = 1.\end{cases}$$

Proposta por Nolan (1997) e referenciada na literatura como parametrização S^0, a expressão (3.19) é uma das parametrizações mais utilizadas atualmente. Na parametrização S^0 ambas, a função característica e a função densidade de probabilidade, são juntamente contínuas em α e β.

Observação 3.7. Note que,

(1) quando $\beta = 0$, segue que $\mu = \mu_M = \mu_0$, para todo $\alpha \in (0, 2]$, $\sigma \geq 0$ e $\mu \in \mathbb{R}$. Nesse caso, as três parametrizações (3.13), (3.16) e (3.19) coincidem e $\varphi_X(\cdot)$ é dada por

$$\varphi_X(t) = e^{\mathrm{i}\mu t - \sigma^\alpha |t|^\alpha}, \quad \text{para todo } t \in \mathbb{R};$$

(2) quando $\beta \neq 0$, temos os seguintes casos,

- se $\alpha \neq 1$ e $\sigma \neq 1$, todas as parametrizações diferem pois $\mu_M = \mu + \beta\sigma^\alpha \tan\left(\frac{\pi\alpha}{2}\right)$ e $\mu_0 = \mu + \beta\sigma \tan\left(\frac{\pi\alpha}{2}\right)$, para qualquer $\mu \in \mathbb{R}$;
- se $\alpha \neq 1$ e $\sigma = 1$, temos $\mu_M = \mu_0 \neq \mu$. Nesse caso, apenas as parametrizações (3.16) e (3.19) coincidem;
- se $\alpha = 1$ e $\sigma \neq 1$, temos $\mu = \mu_M \neq \mu_0$. Nesse caso, apenas as parametrizações (3.13) e (3.16) coincidem;
- quando $\alpha = 1$ e $\sigma = 1$, temos $\mu = \mu_M = \mu_0$. Nesse caso, as três parametrizações (3.13), (3.16) e (3.19) coincidem e $\varphi_X(\cdot)$ é dada por

$$\varphi_X(t) = \exp\left\{\mathrm{i}\mu t - \left[|t| + \mathrm{i}\, t\, \beta \frac{2}{\pi} \ln(|t|)\right]\right\}, \quad \text{para todo } t \in \mathbb{R}.$$

Observação 3.8. Seja X uma variável aleatória $S_\alpha(\beta, \sigma, \mu)$. Então, utilizamos a notação $X \sim S^M_\alpha(\beta, \sigma, \mu_M)$ para indicar que estamos considerando a parametrização dada pela expressão (3.16). De forma análoga, utilizamos a notação $X \sim S^0_\alpha(\beta, \sigma, \mu_0)$ para indicar que estamos considerando a parametrização (3.19).

Pela expressão (3.5) concluimos que a relação $X \sim S_\alpha(\beta, \sigma, \mu) \iff \frac{X-\mu}{\sigma} \sim S_\alpha(\beta, 1, 0)$ é válida somente para o caso $\alpha \neq 1$. De forma análoga, na parametrização (3.19), quando $\alpha = 1$, temos $\mu_M = \mu$. Sendo assim, $X \sim S^M_\alpha(\beta, \sigma, \mu_M)$ pode não implicar $\frac{X-\mu_M}{\sigma} \sim S^M_\alpha(\beta, 1, 0)$. No lema que segue mostramos que, na parametrização S^0, a relação de equivalência é válida para todo $\alpha \in (0, 2]$.

Lema 3.2. *Sejam X uma variável aleatória, $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\sigma \geq 0$ e $\mu_0 \in \mathbb{R}$. Então,*

$$X \sim S^0_\alpha(\beta, \sigma, \mu_0) \iff \frac{X - \mu_0}{\sigma} \sim S^0_\alpha(\beta, 1, 0).$$

Prova: Para provar essa equivalência, é suficiente mostrar que, se $X \sim S^0_\alpha(\beta, \sigma, \mu_0)$, então

(a) $X + a \sim S^0_\alpha(\beta, \sigma, \mu_0 + a)$, para qualquer $a \in \mathbb{R}$,

(b) $aX \sim S^0_\alpha(\mathrm{sgn}(a)\beta, |a|\sigma, a\mu_0)$, para qualquer $a \in \mathbb{R} \setminus \{0\}$.

para todo $\alpha \in (0, 2]$. Para provar (a), note que $\varphi_{X+a}(t) = \mathbb{E}(e^{\mathrm{i}t(X+a)}) = e^{\mathrm{i}ta}\mathbb{E}(e^{\mathrm{i}tX}) = e^{\mathrm{i}ta}\varphi_X(t)$. Logo, pela expressão (3.19), segue que

$$\begin{aligned}\varphi_{X+a}(t) &= e^{\mathrm{i}ta} \exp\left\{\mathrm{i}\mu_0 t - \sigma^\alpha |t|^\alpha \left(1 + \mathrm{i}\,\mathrm{sgn}(t)\omega_0(t, \alpha, \beta)\right)\right\} \\ &= \exp\left\{\mathrm{i}(\mu_0 + a)t - \sigma^\alpha |t|^\alpha \left(1 + \mathrm{i}\,\mathrm{sgn}(t)\omega_0(t, \alpha, \beta)\right)\right\}, \quad \text{para todo } \alpha \in (0, 2].\end{aligned}$$

Pela unicidade da função característica, segue que $X + a \sim S^0_\alpha(\beta, \sigma, \mu_0 + a)$. Para provar (b), note que $\varphi_{aX}(t) = \mathbb{E}(e^{\mathrm{i}t(aX)}) = \mathbb{E}(e^{\mathrm{i}(at)X}) = \varphi_X(at)$. Logo, pela expressão (3.19), segue que

$$\begin{aligned}\varphi_{aX}(t) = \varphi_X(at) &= \exp\left\{\mathrm{i}\mu_0 at - \sigma^\alpha |at|^\alpha \left[1 + \mathrm{i}\,\mathrm{sgn}(at)\beta\left((\sigma|at|)^{1-\alpha} - 1\right)\tan\left(\frac{\pi\alpha}{2}\right)\right]\right\} \\ &= \exp\left\{\mathrm{i}a\mu_0 t - (|a|\sigma)^\alpha |t|^\alpha \left[1 + \mathrm{i}\,\mathrm{sgn}(t)\mathrm{sgn}(a)\beta\left((|a|\sigma|t|)^{1-\alpha} - 1\right)\tan\left(\frac{\pi\alpha}{2}\right)\right]\right\}, \quad \text{se } \alpha \neq 1,\end{aligned}$$

e

$$\varphi_{aX}(t) = \varphi_X(at) = \exp\left\{\mathrm{i}\mu_0 at - \sigma|at|\left[1 + \mathrm{i}\,\mathrm{sgn}(at)\,\beta\frac{2}{\pi}\ln\left(\sigma|at|\right)\right]\right\}$$
$$= \exp\left\{\mathrm{i}a\mu_0 t - |a|\sigma|t|\left[1 + \mathrm{i}\,\mathrm{sgn}(t)\mathrm{sgn}(a)\,\beta\frac{2}{\pi}\ln\left(|a|\sigma|t|\right)\right]\right\}, \quad \text{se } \alpha = 1.$$

Pela unicidade da função característica, $aX \sim S^0_\alpha(\mathrm{sgn}(a)\beta, |a|\sigma, a\mu_0)$ e o resultado segue. ∎

3.3 Cálculo da Função Densidade de Probabilidade Estável

Um dos maiores problemas encontrados quando trabalhamos com variáveis aleatórias estáveis é que, com excessão dos exemplos apresentados na Figura 3.1, a função densidade de probabilidade não possui uma forma fechada conhecida. Devido a esse fato, aproximações ou integração numérica direta devem ser empregadas. Tais procedimentos aumentam consideravelmente o tempo computacional e acarretam perda de precisão nas estimações.

O teorema que segue nos fornece uma expressão para a função densidade de probabilidade de uma variável aleatória α-estável padrão ($\mu = 0, \sigma = 1$), na parametrização canônica. Dado que $|\varphi_X(\cdot)|$ é integrável em toda reta, tal expressão é obtida através da fórmula da inversão, dada pela expressão (2.5).

Teorema 3.6. *Seja $X \sim S_\alpha(\beta, 1, 0)$, com $\alpha \in (0, 2]$. Então, a função densidade de probabilidade de X é dada por*

$$p_X(x;\alpha,\beta) = \frac{1}{\pi}\int_0^\infty \cos\big(h(x,t;\alpha,\beta)\big)e^{-t^\alpha}dt, \tag{3.20}$$

onde

$$h(x,t;\alpha,\beta) = \begin{cases} xt - t^\alpha\beta\tan\left(\frac{\pi\alpha}{2}\right), & se \quad \alpha \neq 1, \\ xt + t\beta\frac{2}{\pi}\ln(t), & se \quad \alpha = 1. \end{cases}$$

Prova: Note que, pela expressão (3.13) segue que $|\varphi_X(t)| = e^{-|t|^\alpha}$. Portanto, $|\varphi_X(\cdot)|$ é integrável em toda reta. Pelo Corolário 2.1, a função densidade de probabilidade de X é dada por

$$p_X(x;\alpha,\beta) = \frac{1}{2\pi}\int_{-\infty}^\infty e^{-\mathrm{i}xt}\varphi_X(t)dt. \tag{3.21}$$

Pela Proposição 2.1, $\varphi_X(0) = 0$ e $\varphi_X(-t) = \overline{\varphi_X(t)}$, para todo $t \in \mathbb{R}$. Segue que $e^{\mathrm{i}xt}\varphi_X(-t) = \overline{e^{-\mathrm{i}xt}\varphi_X(t)}$, para todo $t > 0$ e a expressão (3.21) pode ser reescrita como

$$p_X(x;\alpha,\beta) = \frac{1}{2\pi}\int_0^\infty e^{\mathrm{i}xt}\varphi_X(-t)dt + \frac{1}{2\pi}\int_0^\infty e^{-\mathrm{i}xt}\varphi_X(t)dt$$
$$= \frac{1}{2\pi}\int_0^\infty \overline{e^{-\mathrm{i}xt}\varphi_X(t)}dt + \frac{1}{2\pi}\int_0^\infty e^{-\mathrm{i}xt}\varphi_X(t)dt$$
$$= \frac{1}{\pi}\mathfrak{Re}\left(\int_0^\infty e^{-\mathrm{i}xt}\varphi_X(t)dt\right). \tag{3.22}$$

Pela expressão (3.13), se $\sigma = 1, \mu = 0$ e $t \geq 0$, então,

$$\varphi_X(t) = \begin{cases} \exp\left\{-t^\alpha + \mathrm{i}t^\alpha\beta\tan\left(\frac{\pi\alpha}{2}\right)\right\}, & \text{se} \quad \alpha \neq 1, \\ \exp\left\{-t - \mathrm{i}t\beta\frac{2}{\pi}\ln(t)\right\}, & \text{se} \quad \alpha = 1. \end{cases}$$

Segue que, para todo $t \geq 0$,

$$e^{-itx}\varphi_X(t) = \begin{cases} e^{-t^\alpha}\exp\left\{-\mathrm{i}[xt - t^\alpha\beta\tan(\frac{\pi\alpha}{2})]\right\}, & \text{se} \quad \alpha \neq 1, \\ e^{-t}\exp\left\{-\mathrm{i}[xt + t\beta\frac{2}{\pi}\ln(t)]\right\}, & \text{se} \quad \alpha = 1. \end{cases} \tag{3.23}$$

Além disso, $e^{-\mathrm{i}\theta} = \cos(\theta) - \mathrm{i}\sin(\theta)$, para todo $\theta \in \mathbb{R}$. Portanto, pelas expressões (3.22) e (3.23) a igualdade dada na expressão (3.20) segue. ∎

Observação 3.9. O Teorema 3.6 pode também ser enunciado assumindo-se $X \sim S^M_\alpha(\beta, 1, 0)$ ou $X \sim S^0_\alpha(\beta, 1, 0)$. Note que, como estamos tomando $\sigma = 1$, segue da Observação 3.7 que as parametrizações (3.16) e (3.19) coincidem, para todo $\alpha \in (0, 2]$ e $\beta \in [-1, 1]$. Aém disso, as três parametrizações (3.13), (3.16) e (3.19) coincidem quando $\alpha = 1$ e $\beta \in [-1, 1]$. Sendo assim, ao considerar a parametrização (3.16) ou a (3.19), a função $h(x, t; \alpha, \beta)$ dada na expressão (3.20) deve ser substituída por

$$h(x, t; \alpha, \beta) = \begin{cases} xt + \beta\tan\left(\frac{\pi\alpha}{2}\right)(t - t^\alpha), & \alpha \neq 1, \\ xt + \beta t\frac{2}{\pi}\ln(t), & \alpha = 1. \end{cases}$$

Nesse caso, a prova do resultado segue os mesmos passos daquela apresentada para o Teorema 3.6.

Note que, dizer que $X \sim S_\alpha(\beta, \sigma, \mu)$ é equivalente a dizer que $X \sim S^M_\alpha(\beta, \sigma, \mu_M)$, onde μ_0 é definido na expressão (3.16), ou ainda, que $X \sim S^0_\alpha(\beta, \sigma, \mu_0)$, onde μ_0 é definido na expressão (3.19). Portanto, a seguinte igualdade é sempre válida

$$p_X(x; \alpha, \beta, \sigma, \mu) = p^M_X(x; \alpha, \beta, \sigma, \mu_M) = p^0_X(x; \alpha, \beta, \sigma, \mu_0), \quad \text{para todo } x \in \mathbb{R}, \tag{3.24}$$

onde $p_X(\cdot; \alpha, \beta, \sigma, \mu)$, $p^M_X(\cdot; \alpha, \beta, \sigma, \mu_M)$ e $p^0_X(x; \alpha, \beta, \sigma, \mu_0)$ representam a função densidade de probabilidade de X quando consideramos, respectivamente, a parametrização (3.13), (3.16) e (3.19).

Observação 3.10. Observamos que, pela Proposição 3.1 e pelo Lemma 3.2, somente na parametrização S^0 a equivalência $X \sim S^0_\alpha(\beta, \sigma, \mu_0) \iff Z = \frac{X - \mu_0}{\sigma} \sim S^0_\alpha(\beta, 1, 0)$ é válida para todo $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\sigma \geq 0$ e $\mu \in \mathbb{R}$. Sendo assim, do ponto de vista computacional, a parametrização S^0 torna-se mais conveniente pois $p^0_X(\cdot; \alpha, \beta, \sigma, \mu_0)$ pode sempre ser reescrita como

$$p^0_X(x; \alpha, \beta, \sigma, \mu_0) = \frac{1}{\sigma}p^0_Z\Big(\frac{x - \mu_0}{\sigma}; \alpha, \beta, 1, 0\Big), \quad \text{para todo } x \in \mathbb{R} \ \text{ e } \ \alpha \in (0, 2],$$

onde $p^0_Z(\cdot; \alpha, \beta, 1, 0) := p^0_Z(\cdot; \alpha, \beta)$ é a função densidade de probabilidade de uma variável aleatória $Z \sim S^0_\alpha(\beta, 1, 0)$.

Zolotarev (1986) apresenta uma forma alternativa para representar a função densidade de probabilidade de uma variável aleatória α-estável. As expressões derivadas por Zolotarev (1986), página 74, foram obtidas a partir da expressão (3.20) utilizando a parametrização (B) (veja Zolotarev, 1986, página 12). Entretanto, na literatura encontramos expressões equivalentes para outras parametrizações. Independente da parametrização utilizada para derivá-las, tais expressões são referenciadas como "equações integrais de Zolotarev". No teorema que segue é apresentada tal representação para a função densidade de probabilidade de uma variável aleatória α-estável padronizada, na representação S^0, isto é, $X \sim S^0_\alpha(\beta, 1, 0)$.

Teorema 3.7. *Seja $X \sim S^0_\alpha(\beta, 1, 0)$ e $p^0_X(x; \alpha, \beta)$ a sua função densidade de probabilidade. Defina*

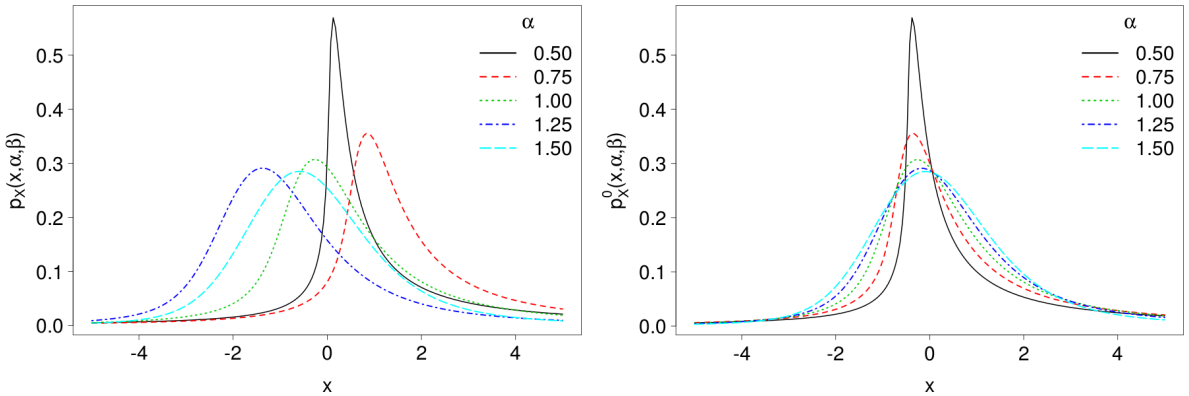
$$\zeta = -\beta\tan\left(\frac{\pi\alpha}{2}\right), \quad \xi = \begin{cases} \frac{1}{\alpha}\arctan(-\zeta), & se \quad \alpha \neq 1, \\ \frac{\pi}{2}, & se \quad \alpha = 1, \end{cases} \quad e$$

$$V(t;\alpha,\beta)=\begin{cases}\cos(\alpha\xi)^{\frac{1}{\alpha-1}}\left[\dfrac{\cos(t)}{\sin\left(\alpha(\xi+t)\right)}\right]^{\frac{\alpha}{\alpha-1}}\dfrac{\cos\left(\alpha\xi+(\alpha-1)t\right)}{\cos(t)}, & se \quad \alpha\neq 1,\\ \dfrac{2}{\pi}\left(\dfrac{\frac{\pi}{2}+\beta t}{\cos(t)}\right)\exp\left\{\dfrac{1}{\beta}\left(\dfrac{\pi}{2}+\beta t\right)\tan(t)\right\}, & se \quad \alpha=1 \ e \ \beta\neq 0.\end{cases}$$

Então, $p_X^0(x;\alpha,\beta)$ *pode ser expressa como*

$$p_X^0(x;\alpha,\beta)=\begin{cases}\dfrac{\alpha(x-\zeta)^{\frac{1}{\alpha-1}}}{\pi|\alpha-1|}\displaystyle\int_{-\xi}^{\frac{\pi}{2}}V(t;\alpha,\beta)\exp\Big\{-(x-\zeta)^{\frac{\alpha}{\alpha-1}}V(t;\alpha,\beta)\Big\}dt, & se \quad \alpha\neq 1 \ e \ x>\zeta;\\ \dfrac{\Gamma\left(1+\frac{1}{\alpha}\right)\cos(\xi)}{\pi(1+\zeta^2)^{\frac{1}{2\alpha}}}, & se \quad \alpha\neq 1 \ e \ x=\zeta;\\ p_X^0(-x;\alpha,-\beta), & se \quad \alpha\neq 1 \ e \ x<\zeta;\\ \dfrac{1}{2|\beta|}e^{-\frac{\pi x}{2\beta}}\displaystyle\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}V(t;1,\beta)\exp\Big\{-e^{-\frac{\pi x}{2\beta}}V(t;1,\beta)\Big\}dt, & se \quad \alpha=1 \ e \ \beta\neq 0;\\ \dfrac{1}{\pi(1+x^2)}, & se \quad \alpha=1 \ e \ \beta=0.\end{cases}$$

Prova: Veja Nolan (1997, teorema 1). ∎



(a) $X\sim S_\alpha(\beta,1,0)$ **(b)** $X\sim S_\alpha^0(\beta,1,0)$

Figura 3.2: Função densidade de probabilidade estável em diferentes parametrizações. (a) Parametrização Canônica, com $\mu=0$; (b) Parametrização S_0, com $\mu_0=0$. Em ambos os casos $\alpha\in\{0.5,\ 0.75,\ 1,\ 1.25,\ 1.5\}$, $\beta=0.5$ e $\sigma=1$.

A Figura 3.2 apresenta o gráfico da função densidade de probabilidade nas parametrizações canônica e S^0. Na Figura 3.2(a) tomamos $\mu=0$. Na Figura 3.2(b) assumimos $\mu_0=0$. Em ambos os casos considera-se $\beta=0.5$, $\sigma=1$ e diferentes valores de α. Note que, pelas Observações 3.7 e 3.9 conclui-se que, quando $\sigma=1$, as parametrizações (3.16) e S^0 coincidem para todo $\alpha\in(0,2]$, $\beta\in[-1,1]$ e $\mu\in\mathbb{R}$. Em particular, dizer que $X\sim S_\alpha^M(\beta,1,0)$ é equivalente a afirmar que $X\sim S_\alpha^0(\beta,1,0)$ e, portanto, $p_X^M(x;\alpha,\beta)=p_X^0(x;\alpha,\beta)$, para todo $x\in\mathbb{R}$, onde $p_X^M(\cdot;\alpha,\beta):=p_X^M(\cdot;\alpha,\beta,1,0)$. Sendo assim, a Figura 3.2(b) também representa a função densidade de probabilidade na parametrização (3.16), com $\mu_M=0$.

Lema 3.3. *O suporte da função densidade de probabilidade de uma variável aleatória estável,*

nas parametrizações canônica e S^0, é dado, respectivamente, por

$$\text{supp}\{p_X(x;\alpha,\beta,\sigma,\mu)\} = \begin{cases} [\mu,\infty), & se \quad \alpha<1 \quad e \quad \beta=1, \\ (-\infty,\mu], & se \quad \alpha<1 \quad e \quad \beta=-1, \\ (-\infty,\infty), & caso\ contrário \end{cases} \tag{3.25}$$

e

$$\text{supp}\{p^0_X(x;\alpha,\beta,\sigma,\mu_0)\} = \begin{cases} [\mu_0-\sigma\tan(\frac{\pi\alpha}{2}),\infty), & se \quad \alpha<1 \quad e \quad \beta=1, \\ (-\infty,\mu_0+\sigma\tan(\frac{\pi\alpha}{2})], & se \quad \alpha<1 \quad e \quad \beta=-1, \\ (-\infty,\infty), & caso\ contrário. \end{cases} \tag{3.26}$$

Prova: Para a prova da igualdade dada na expressão (3.25), veja Zolotarev (1986), página 80. A igualdade dada na expressão (3.26) segue de imediato da igualdade (3.24) e do fato que $\mu_0 = \mu+\beta\sigma\tan(\frac{\pi\alpha}{2})$, se $\alpha\neq 1$. ∎

O teorema que segue apresenta uma expressão para $\mathbb{E}(|X|^r)$, onde X é uma variável aleatória estável simétrica, com $\alpha\in(0,2]$ e $0<r<\alpha$. Como consequência desse teorema, o Corolário 3.1 apresenta a expressão para $\mathbb{E}(|X|)$, onde X é uma variável aleatória estável simétrica, com $\alpha\in(1,2]$. O Corolário 3.2 apresenta uma representação integral para um número real qualquer $r\geq 2$.

Teorema 3.8. *Seja X uma variável aleatória α-estável simétrica padrão, com $\alpha\in(0,2]$. Então, para todo $0<r<\alpha$,*

$$\mathbb{E}\big(|X|^r\big) = \frac{2^{r-1}}{r\int_0^\infty u^{-r-1}\sin^2(u)du}\Gamma\Big(1-\frac{r}{\alpha}\Big).$$

Prova: Veja Samorodnitsky e Taqqu (1994, página 18). ∎

Corolário 3.1. *Seja X uma variável aleatória α-estável simétrica padrão, com $\alpha\in(1,2]$. Então*

$$\mathbb{E}\big(|X|\big) = \frac{2}{\pi}\Gamma\Big(1-\frac{1}{\alpha}\Big).$$

Prova: A prova é consequência imediata do Teorema 3.8, basta notar que $\int_0^\infty \frac{\sin^2(u)}{u^2}du = \frac{\pi}{2}$ (veja Gradshteyn e Ryzhik, 2000, equação 3.821, item 9). ∎

Corolário 3.2. *Para todo $r\in\mathbb{R}$, tal que $r\geq 2$,*

$$\int_0^{\frac{\pi}{2}} \sin\Big(\frac{r}{r-1}t\Big)\big[\cos(t)\big]^{-\frac{r-1}{r}}\Big[\cos\Big(\frac{1}{r-1}\Big)\Big]^{-\frac{1}{r}}dt = r. \tag{3.27}$$

Prova: A prova da igualdade dada na expressão (3.27) consiste em mostrar que, para uma variável $X\sim S_\alpha(\beta,\sigma,\mu)$, com $\beta=0,\sigma=1,\mu=0$ e $\alpha\in(1,2]$, $\mathbb{E}(|X|)$ é dada por

$$\mathbb{E}\big(|X|\big) = \frac{2}{\pi}\left(\frac{\alpha-1}{\alpha}\right)\Gamma\left(1-\frac{1}{\alpha}\right)\int_0^{\frac{\pi}{2}}\sin(\alpha t)\big[\cos(t)\big]^{-\frac{1}{\alpha}}\big[\cos(\alpha t-t)\big]^{-\frac{\alpha-1}{\alpha}}dt.$$

Então, comparando com a expressão dada no Teorema 3.8 encontramos a expressão dada em (3.27). No que segue, apresentamos os detalhes da prova.

Seja X uma variável aleatória tal que $X\sim S_\alpha(\beta,\sigma,\mu)$, com $\beta=0,\sigma=1,\mu=0$ e $\alpha\in(1,2]$. Da expressão (3.19) segue que, através de reparametrização, $X\sim S^0_\alpha(\beta,\sigma,\mu_0)$, com $\beta=0,\sigma=1$ e $\mu_0=0$. Pelo Teorema 3.7, segue que

$$\zeta=0,\quad \xi=0,\quad V(t;\alpha,0)=\left(\frac{\cos(t)}{\sin(\alpha t)}\right)^{\frac{\alpha}{\alpha-1}}\frac{\cos(\alpha t-t)}{\cos(t)},\quad p^0_X(0;\alpha,0)=\frac{1}{\pi}\Gamma\left(1+\frac{1}{\alpha}\right)$$

e, para todo $x > 0$,

$$p_X^0(-x;\alpha,0) = p_X^0(x;\alpha,0) = \frac{\alpha x^{\frac{1}{\alpha-1}}}{\pi(\alpha-1)} \int_0^{\frac{\pi}{2}} V(t;\alpha,0) \exp\Big\{ - x^{\frac{\alpha}{\alpha-1}} V(t;\alpha,0)\Big\} dt.$$

Como X é uma variável aleatória simétrica temos

$$\mathbb{E}\big(|X|\big) = \int_{-\infty}^{\infty} |x| p_X^0(x;\alpha,0) dx \;=\; 2\int_0^{\infty} x p_X^0(x;\alpha,0) dx.$$

Segue que,

$$\begin{aligned}
\mathbb{E}\big(|X|\big) &= 2\int_0^{\infty} x\left[\frac{\alpha x^{\frac{1}{\alpha-1}}}{\pi(\alpha-1)} \int_0^{\frac{\pi}{2}} V(t;\alpha,0) \exp\Big\{ - x^{\frac{\alpha}{\alpha-1}} V(t;\alpha,0)\Big\} dt\right] dx \\
&= \frac{2}{\pi}\frac{\alpha}{(\alpha-1)} \int_0^{\infty}\int_0^{\frac{\pi}{2}} x^{\frac{\alpha}{\alpha-1}} V(t;\alpha,0) \exp\Big\{ - x^{\frac{\alpha}{\alpha-1}} V(t;\alpha,0)\Big\} dt dx \\
&= \frac{2}{\pi}\frac{\alpha}{(\alpha-1)} \int_0^{\frac{\pi}{2}}\int_0^{\infty} x^{\frac{\alpha}{\alpha-1}} V(t;\alpha,0) \exp\Big\{ - x^{\frac{\alpha}{\alpha-1}} V(t;\alpha,0)\Big\} dx dt. \qquad (3.28)
\end{aligned}$$

Defina $b := V(t;\alpha,0)$ e $\nu - 1 := a := \frac{\alpha}{\alpha-1}$. Basta mostrar que $b > 0$ e a fórmula (Gradshteyn e Ryzhik, 2000, página 364)

$$\int_0^{\infty} x^{\nu-1} e^{-bx^a} dx = \frac{1}{a} b^{-\frac{\nu}{a}} \Gamma\left(\frac{\nu}{a}\right), \quad \text{se } \mathfrak{Re}(\nu) >, \mathfrak{Re}(b) > 0 \text{ e } a > 0, \qquad (3.29)$$

pode ser aplicada, para resolver a equação (3.28). Para isso, note que

$$b = V(t;\alpha,0) = \left(\frac{\cos(t)}{\sin(\alpha t)}\right)^{\frac{\alpha}{\alpha-1}} \frac{\cos(\alpha t - t)}{\cos(t)} = \frac{\big(\cos(t)\big)^{\frac{1}{\alpha-1}}}{\big(\sin(\alpha t)\big)^{\frac{\alpha}{\alpha-1}}} \cos(\alpha t - t).$$

Como $\alpha \in (0,2]$, temos $\sin(\alpha t) > 0$, $\cos(t) > 0$ e $\cos(\alpha t - t) > 0$, para todo $0 < t < \frac{\pi}{2}$. Logo, $b = V(t;\alpha,0) > 0$, para todo $0 < t < \frac{\pi}{2}$. Portanto, como $\nu - 1 = a$ e $a = \frac{\alpha}{\alpha-1}$ da expressão (3.29), segue que

$$\begin{aligned}
\int_0^{\infty} x^{\frac{\alpha}{\alpha-1}} V(t;\alpha,0) e^{-x^{\frac{\alpha}{\alpha-1}} V(t;\alpha,0)} dx &= \int_0^{\infty} b x^a e^{-bx^a} dx = \frac{1}{a} b^{-\frac{1}{a}} \Gamma\left(1 + \frac{1}{a}\right) \\
&= \Gamma\left(1 + \frac{\alpha-1}{\alpha}\right) \frac{\alpha-1}{\alpha} \left[V(t;\alpha,0)\right]^{-\frac{\alpha-1}{\alpha}}, \qquad (3.30)
\end{aligned}$$

onde

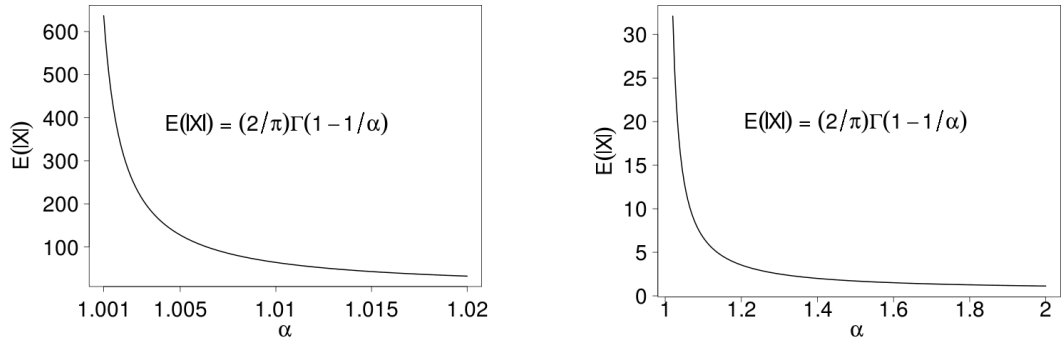
$$\left[V(t;\alpha,0)\right]^{-\frac{\alpha-1}{\alpha}} = \frac{\sin(\alpha t)}{\cos(t)} \left[\frac{\cos(\alpha t - t)}{\cos(t)}\right]^{-\frac{\alpha-1}{\alpha}} = \sin(\alpha t)\big[\cos(t)\big]^{-\frac{1}{\alpha}} \big[\cos(\alpha t - t)\big]^{-\frac{\alpha-1}{\alpha}}. \qquad (3.31)$$

Substituindo as expressões (3.30) e (3.31), na expressão (3.28), temos

$$\begin{aligned}
\mathbb{E}\big(|X|\big) &= \frac{2}{\pi}\frac{\alpha}{(\alpha-1)} \int_0^{\frac{\pi}{2}}\int_0^{\infty} x^{\frac{\alpha}{\alpha-1}} V(t;\alpha,0) \exp\Big\{ - x^{\frac{\alpha}{\alpha-1}} V(t;\alpha,0)\Big\} dx dt \\
&= \frac{2}{\pi}\Gamma\left(1 + \frac{\alpha-1}{\alpha}\right) \int_0^{\frac{\pi}{2}} \sin(\alpha t)\big[\cos(t)\big]^{-\frac{1}{\alpha}} \big[\cos(\alpha t - t)\big]^{-\frac{\alpha-1}{\alpha}} dt.
\end{aligned}$$

Observando que $\left(1 - \frac{1}{\alpha}\right)\Gamma\left(1 - \frac{1}{\alpha}\right) = \Gamma\left(2 - \frac{1}{\alpha}\right) = \Gamma\left(1 + \frac{\alpha-1}{\alpha}\right)$, temos

$$\mathbb{E}\big(|X|\big) = \frac{2}{\pi}\left(\frac{\alpha-1}{\alpha}\right)\Gamma\left(1 - \frac{1}{\alpha}\right) \int_0^{\frac{\pi}{2}} \sin(\alpha t)\big[\cos(t)\big]^{-\frac{1}{\alpha}} \big[\cos(\alpha t - t)\big]^{-\frac{\alpha-1}{\alpha}} dt.$$



(a) $1.001 \leq \alpha \leq 1.02$ **(b)** $1.02 \leq \alpha \leq 2$

Figura 3.3: Valores de $\mathbb{E}(|X|)$, onde $X \sim S_\alpha(0,1,0)$, para $1 < \alpha \leq 2$: (a) $1.001 \leq \alpha \leq 1.02$; (b) $1.02 \leq \alpha \leq 2$.

Por outro lado, pelo Corolário 3.1, $\mathbb{E}(|X|) = \frac{2}{\pi}\Gamma\left(1 - \frac{1}{\alpha}\right)$. Portanto,

$$\int_0^{\frac{\pi}{2}} \sin(\alpha t)\big[\cos(t)\big]^{-\frac{1}{\alpha}}\big[\cos(\alpha t - t)\big]^{-\frac{\alpha-1}{\alpha}} dt = \frac{\alpha}{\alpha - 1}. \tag{3.32}$$

Tomando $\alpha = \frac{r}{r-1}$, temos $1 < \alpha \leq 2$ e $r = \frac{\alpha}{\alpha-1}$. Substituindo α na expressão (3.32), o resultado segue. ∎

A Figura 3.3 apresenta os valores de $\mathbb{E}(|X|)$, onde $X \sim S_\alpha(0,1,0)$, para $1 < \alpha \leq 2$. Tais valores foram obtidos avaliando-se a expressão apresentada no Corolário 3.1. Observando as Figura 3.3(a) e 3.3(b), fica claro que $\mathbb{E}(|X|) \to \infty$, quando $\alpha \to 1$.

No que segue, apresentamos alguns dos métodos encontrados na literatura para avaliar a função densidade de probabilidade para distribuições estáveis.

3.3.1 Métodos de Integração Numérica Direta

Nolan (1999) propõe um método para calcular a função densidade de probabilidade de uma variável aleatória α-estável. Tal método consiste em avaliar numericamente a integral dada pela expressão (3.20). Devido a não continuidade da função $\omega(\cdot, \alpha, \beta)$, na parametrização canônica, a parametrização utilizada por Nolan é aquela dada pela expressão (3.16), que coincide com a parametrização (3.19) pois os cálculos são feitos para $\sigma = 1$. O método proposto consiste em dividir a região de integração em intervalos onde o cosseno muda de sinal. Quando $\beta = 0$, os pontos extremos dos intervalos são encontrados analiticamente. Nos outros casos, os extremos dos intervalos são calculados numericamente e a integral é aproximada em cada intervalo.

Os seguintes resultados foram reportados por Nolan (1999):

- Se $\alpha < 1$, o expoente e^{-t^α} decresce lentamente e a região de integração deve ser maior para obter maior precisão, o que implica maior tempo computacional.
- Se $\alpha < 1$, o número de subintervalos obtidos é em geral muito grande. Isso faz com que os erros de arredondamento aumentem significativamente e a precisão de seis casas decimais não pode ser obtida.
- Se $\beta \neq 0$ e $0 < |\alpha - 1| < 0,001$, o cálculo do termo $\tan\left(\frac{\pi\alpha}{2}\right)(t - t^\alpha)$ é numericamente instável.
- Para valores grandes de x o integrando da expressão (3.20) oscila rapidamente. Tal comportamento pode ser observado na Figura 3.4(a) que ilustra o comportamento da função

$f(t) = e^{-t^\alpha}\cos\left(xt + \beta\tan(\frac{\pi\alpha}{2})(t - t^\alpha)\right)$, o integrando da expressão (3.20) na parametrização S^0, para $x = 7$ e $\alpha = \beta = 0.5$.

Devido às dificuldades encontradas ao calcular numericamente a integral em (3.20), Nolan (1997) propõe outro método de integração direta que consiste em avaliar as equações integrais apresentadas no Teorema 3.7. Atualmente, o método proposto por Nolan (1997) é o mais utilizado pois é computacionalmente mais simples. Isso se deve ao fato que o integrando é uma função contínua, limitada, não oscilante e a região de integração é um intervalo fechado (conforme Nolan, 1997). Mais especificamente, o integrando a ser avaliado é uma função que assume o valor 0, quando $t = -\xi = -\frac{1}{\alpha}\arctan\left(\beta\tan(\frac{\pi\alpha}{2})\right)$ (limite inferior de integração), cresce monotonamente até um valor máximo $(1/e)$, que é atingido para um único valor de $t_0 \in [-\xi, \pi/2]$, e então decresce monotonamente até atingir zero quando $t = \pi/2$ (limite superior de integração). Sendo assim, Nolan (1997) sugere avaliar a integral dada no Teorema 3.7 em duas partes, de $-\xi$ a t_0 e de t_0 a $\frac{\pi}{2}$.

A Figura 3.4(b) ilustra o comportamento da função (integrando definido no Teorema 3.7)

$$f(t) = \frac{\alpha(x-\zeta)^{\frac{1}{\alpha-1}}}{\pi|\alpha-1|}V(t;\alpha,\beta)\exp\Big\{-(x-\zeta)^{\frac{\alpha}{\alpha-1}}V(t;\alpha,\beta)\Big\},$$

com $x = 7$ e $\alpha = 0.5 = \beta$, para $-\xi < t \leq \frac{\pi}{2}$, onde $V(\cdot;\alpha,\beta)$ é a função definida no Teorema 3.7. Note que, nesse caso, o número de pontos de inflexão da função no integrando, para os mesmos valores de x, α e β, é muito menor que no caso da Figura 3.4(a).

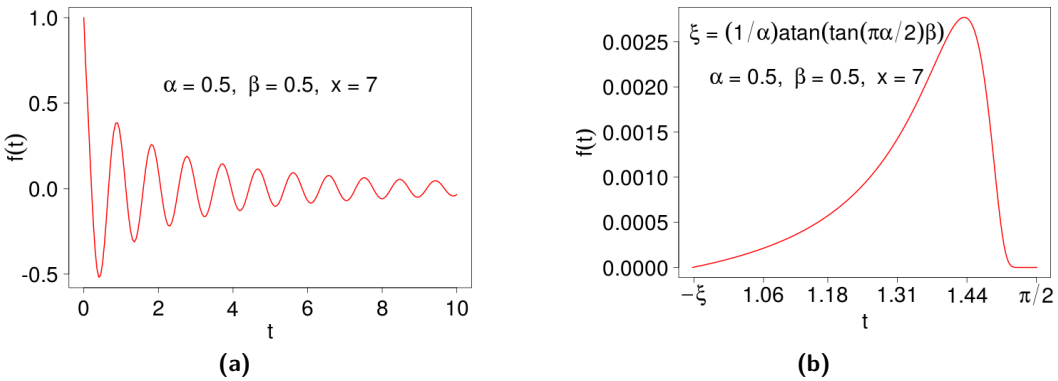


Figura 3.4: Comportamento do integrando definido: (a) no Teorema 3.6, ou seja, na parametrização canônica; (b) no Teorema 3.7, isto é, na representação de Zolotarev. Em ambos os casos, $x = 7$ e $\alpha = \beta = 0.5$.

Quando $\alpha > 1$, outro método de integração numérica direta pode ser utilizado. Nesse caso, a integral dada em (3.20) é substituída pela aproximação

$$p_X(x;\alpha,\beta) \approx \frac{1}{\pi}\int_0^{\Delta}\cos\left(xt + \beta\tan\left(\frac{\pi\alpha}{2}\right)(t - t^\alpha)\right)e^{-t^\alpha}dt, \tag{3.33}$$

onde $\Delta = \Delta(\alpha,\beta)$ é a raiz da equação $\Gamma\left(\frac{1}{\alpha}, \Delta^\alpha\right) = \alpha\pi\varepsilon$ e ε é tal que o erro da expressão (3.33) não é maior do que ε. A função $\Gamma(\cdot,\cdot)$ é a Gama incompleta, definida na expressão (2.2). A expressão (3.33) é então avaliada através de algum método de integração adaptativa com precisão $\tilde{\varepsilon}$. Segundo Belov (2005) esse método mostra-se realmente efetivo em termos de aspectos computacionais se a quadratura Gaussiana de 96 pontos é aplicada com $\tilde{\varepsilon} < 10^{-10}$.

3.3.2 Método da Transformada Rápida de Fourier

Mittnik et al. (1999) apresenta um algoritmo para calcular a função densidade de probabilidade de uma variável aleatória α-estável utilizando a transformada rápida de Fourier (FFT). Tal método

baseia-se no fato que a função densidade de probabilidade de uma variável aleatória α-estável pode ser reescrita em termos da função característica, de acordo com a expressão (3.21). A integral (3.21) é então calculada para N pontos distintos, igualmente espaçados, com distância h, dados por $x_k = \left(k - 1 - \frac{N}{2}\right)h$, para $k = 1, \cdots, N$. Tomando $t = 2\pi\omega$, a expressão (3.21) pode ser reescrita como

$$p_X(x_k;\alpha,\beta) = \int_{-\infty}^{\infty} \varphi_X(2\pi\omega) e^{-\mathrm{i}x_k 2\pi\omega} d\omega, \quad \text{para } k = 1, \cdots, N. \tag{3.34}$$

A integral (3.34) pode então ser aproximada utilizando-se a regra do retângulo para N pontos com espaçamento s, isto é,

$$p_X(x_k;\alpha,\beta) \approx s \sum_{n=1}^{N} \varphi_X\big(2\pi s\big[n - 1 - N/2\big]\big) e^{-\mathrm{i}x_k 2\pi s(n-1-N/2)}, \quad \text{para } k = 1, \cdots, N.$$

Tomando $s = \frac{1}{hN}$, obtém-se a aproximação

$$p_X(x_k;\alpha,\beta) \approx s(-1)^{k-1-\frac{N}{2}} \sum_{n=1}^{N} (-1)^{n-1} \varphi_X\big(2\pi s\big[n - 1 - N/2\big]\big) e^{\frac{-\mathrm{i}2\pi(n-1)(k-1)}{N}}, \tag{3.35}$$

para $k = 1, \cdots, N$. Note que a expressão (3.35) é obtida aplicando-se a FFT à sequência

$$(-1)^{n-1} \varphi_X\big(2\pi s\big[n - 1 - N/2\big]\big), \quad \text{para } n = 1, \cdots, N$$

e, em seguida, dividindo-se o k-ésimo termo por $s(-1)^{k-1-\frac{N}{2}}$, para $k = 1, \cdots, N$.

O método consiste então de duas etapas. A primeira delas consiste da especificação dos pontos igualmente espaçados para os quais a função densidade de probabilidade será calculada e o cálculo do valor da função para esses pontos. A FFT torna-se mais eficiente quando o número N de pontos é uma potência de 2, isto é, $N = 2^K$, para algum $K \in \mathbb{N}$. Em geral, na literatura utiliza-se $K = 13$ e $h = 0.01$. A segunda etapa consiste em aplicar uma interpolação linear entre os pontos igualmente espaçados para encontrar o valor da função para os demais valores de x.

3.3.3 *Método das Duas Quadraturas*

No que segue apresentamos o método proposto por Belov (2005), para avaliar a função densidade de probabilidade de uma variável aleatória α-estável, utilizando as quadraturas Gaussiana e de Laguerre (veja o Apêndice A).

O método proposto por Belov (2005), denominado método das duas quadraturas, consiste em avaliar a integral dada pela expressão (3.20) subdividindo-se a região de integração em duas outras regiões como segue

$$p_X(x;\alpha,\beta) = \int_0^T f(x,t;\alpha,\beta)dt + \int_0^{\infty} f(x,u+T;\alpha,\beta)du,$$

onde $f(x,t;\alpha,\beta) = \frac{1}{\pi}\cos\big(h(x,t;\alpha,\beta)\big)e^{-t^{\alpha}}$ e $h(x,t;\alpha,\beta)$ é a função definida em (3.20). A primeira integral é avaliada utilizando-se a regra da quadratura Gaussiana com 96 pontos e a segunda integral é avaliada utilizando-se a regra da quadratura de Laguerre também com 96 pontos. Para valores grandes de $|x|$ utiliza-se a expansão assintótica de Bergström (Bergström, 1952) que, na representação canônica, é dada por

$$p_X(x;\alpha,\beta) = \frac{1}{\pi x}\sum_{k=1}^{N} \frac{(-1)^{k-1}}{k!}\frac{\Gamma(\alpha k+1)}{|x|^{\alpha k}} \sin\left(k\left[\frac{\pi\alpha}{2} + \beta - \alpha\arg(x)\right]\right) + O(|x|^{-\alpha(N+1)-1}),$$

onde $\Gamma(\cdot) = \Gamma(\cdot, 0)$ é definida pela expressão (2.2) e $\arg(\cdot)$ é a função $\arg : \mathbb{C} \setminus \{0\} \to (-\pi, \pi]$ definida por

$$x + \mathrm{i}y \mapsto \begin{cases} 2\arctan\left(\frac{y}{\sqrt{x^2+y^2}+x}\right), & \text{se} \quad x > 0 \ \text{ ou } \ y \neq 0, \\ \pi, & \text{se} \quad x < 0 \ \text{ e } \ y = 0, \\ \text{indefinida}, & \text{se} \quad x = 0 \ \text{ e } \ y = 0. \end{cases}$$

No Apêndice B apresentamos uma comparação da precisão da FFT e da dupla quadratura versus a integração numérica direta para diferentes valores de α e β.

3.4 Simulações

Em muitos estudos práticos torna-se necessário gerar variáveis aleatórias cuja distribuição é estável. O resultado que segue, apresentado por Chambers et al. (1976), fornece uma expressão simples para gerar variáveis estáveis a partir de variáveis aleatórias com distribuição uniforme.

Teorema 3.9. *Sejam U e W variáveis aleatórias independentes, com U uniformemente distribuída em $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ e W exponencialmente distribuída com média 1. Para qualquer $\alpha \in (0, 2]$ e $\beta \in [-1, 1]$, defina*

$$\zeta := -\beta \tan\left(\frac{\pi\alpha}{2}\right) \quad e \quad \xi := \begin{cases} \frac{1}{\alpha}\arctan(-\zeta), & \alpha \neq 1, \\ \frac{\pi}{2}, & \alpha = 1. \end{cases}$$

Então, a variável aleatória X, definida por

$$X := \begin{cases} (1+\zeta^2)^{\frac{1}{2\alpha}} \dfrac{\sin\left(\alpha[U+\xi]\right)}{\cos(U)^{\frac{1}{\alpha}}} \left[\dfrac{\cos\left(U - \alpha[U+\xi]\right)}{W}\right]^{\frac{1-\alpha}{\alpha}}, & se \quad \alpha \neq 1, \\ \dfrac{1}{\xi}\left\{\left(\dfrac{\pi}{2} + \beta U\right)\tan(U) - \beta \ln\left(\dfrac{\frac{\pi}{2} W \cos(U)}{\frac{\pi}{2} + \beta U}\right)\right\}, & se \quad \alpha = 1, \end{cases}$$

tem distribuição $X \sim S_\alpha(\beta, 1, 0)$ e, para todo $\mu \in \mathbb{R}$ e $\sigma \geq 0$, a variável Y, definida por

$$Y := \begin{cases} \sigma X + \mu, & se \quad \alpha \neq 1, \\ \sigma X + \mu + \frac{2}{\pi}\beta\sigma\ln(\sigma), & se \quad \alpha = 1, \end{cases}$$

tem distribuição $S_\alpha(\beta, \sigma, \mu)$.

Prova: Veja Chambers et al. (1976). ∎

Observação 3.11. As variáveis aleatórias U e W podem ser facilmente obtidas, a partir de variáveis aleatórias Θ_1 e Θ_2, independentes e identicamente distribuídas, com função de distribuição $\mathcal{U}(0, 1)$, utilizando-se as definições

$$U := \pi\left(\Theta_1 - \frac{1}{2}\right) \quad \text{e} \quad W := -\ln(\Theta_2).$$

Capítulo 4

Processos SFIEGARCH

Neste capítulo tratamos dos processos FIEGARCH com sazonalidade, denotados por SFIEGARCH. Essa classe de modelos é uma generalização dos processos FIEGARCH introduzidos por Bollerslev e Mikkelsen (1996) e detalhadamente estudados em Prass (2008) e Lopes e Prass (2013a). Dois casos são estudados. No primeiro caso consideramos os processos para os quais as inovações possuem variância finita. Tais processos são apresentados na Seção 4.1 e discutidos com detalhes em Lopes e Prass (2013b) (veja o Apêndice D). No segundo caso, estendemos a definição do processo SFIEGARCH para inovações com distribuição α-estável quando $0 < \alpha < 2$, portanto, quando a variância é infinita. Os processos definidos dessa maneira são denotados por α-SFIEGARCH para deixar implícita a relação com distribuições α-estáveis. Tais processos são apresentados na Seção 4.3 e discutidos com detalhes em Prass, Lopes e Crato (2013) (veja o Apêndice I).

Em ambos os casos, variância finita ou infinita, apresentamos condições necessárias e suficientes para que os processos SFIEGARCH estejam bem definidos. Mais especificamente, mostramos que, se a inovação possui variância finita então, o processo SFIEGARCH está bem definido para $d < 0.5$ e, no caso de variância infinita, para $(1-d)\alpha > 1$. Também tratamos da invertibilidade, estacionariedade (fraca e estrita), ergodicidade e representação espectral (no caso de variância finita) destes processos. Mais detalhes são dados no que segue.

Observação 4.1. Neste capítulo, sempre que utilizarmos *itálico* nas palavras Lema, Teorema, Proposição ou Observação, estamos nos referindo ao resultado apresentado em Lopes e Prass (2013b), no caso da Seção 4.1, ou em Prass, Lopes e Crato (2013), no caso da Seção 4.3.

4.1 Inovações com Variância Finita

Modelos da família ARCH (Engle, 1982) (no inglês, ARCH-*type models*) e modelos de volatilidade estocástica (Breidt et al., 1998) (no inglês, *stochastic volatility models*) constituem as principais classes utilizadas na modelagem da volatilidade de séries temporais financeiras. Entretanto, modelos de volatilidade estocástica possuem uma desvantagem em relação aos demais. Para essa classe, a volatilidade é definida como uma variável latente que não é diretamente observável, o que torna o processo de estimação mais complicado.

Quando nos referimos aos modelos da família ARCH, não estamos tratando apenas do ARCH(p) (Engle, 1982), mas também de suas generalizações. Bollerslev (2010) apresenta uma revisão dos modelos da família ARCH conhecidos na literatura até a data em que o autor preparou o manuscrito[1]. Dentre eles, destacamos os modelos GARCH(p,q) (Bollerslev, 1986), EGARCH(p,q) (Nelson, 1991), FIGARCH(p,d,q) (Baillie et al., 1996) e FIEGARCH(p,d,q) (Bollerslev e Mikkelsen, 1996). A vantagem do modelo FIEGARCH sobre os demais está no fato que tais modelos,

[1] A versão de setembro de 2008 está disponível em SSRN: http://ssrn.com/abstract=1263250.

- análogo aos modelos ARCH e GARCH, são capazes de descrever cluster de volatilidade;
- análogo aos modelos EGARCH, capturam a assimetria dos retornos (a volatilidade reage de forma assimétrica aos retornos, tendendo a ser maior para retornos negativos);
- análogo aos modelos FIGARCH, levam em conta a longa dependência na volatilidade, com a vantagem de ser fracamente estacionário quando $d < 0.5$ (Lopes e Prass, 2013a).

Um estudo extensivo das propriedades teóricas dos processos FIEGARCH pode ser encontrado em Prass (2008) e em Lopes e Prass (2013a). Tais trabalhos também apresentam resultados relacionados a previsão da volatilidade em processos FIEGARCH, análise do desempenho do estimador de quase-verossimilhança em amostras finitas, bem como análise de dados reais e comparação com outros modelos da família ARCH.

Com os avanços da informática, especialmente em termos de armazenamento, transmissão e manipulação de dados, a análise de dados financeiros coletados em intervalos de tempo cada vez menores, como por exemplo retornos intra-diários, tornou-se rotina. O estudo desses dados revelou, por exemplo, a existência tanto de longa dependência, quanto de comportamento cíclico na volatilidade de séries temporais financeiras de alta frequência. No caso dos retornos (log-retornos) da taxa de câmbio, essa característica é geralmente atribuída ao fato de que os horários de negociação nos mercados europeus, asiáticos e norte-americanos sobrepõem-se uns aos outros (conforme Bordignon et al., 2007). No caso de retornos (log-retornos) de preços de ações, os horários de abertura e fechamento do mercado, bem como o horário de almoço são fatores que influenciam no comportamento cíclico da volatilidade.

Embora os modelos FIEGARCH descrevam a longa dependência na volatilidade, tais modelos não são apropriados para séries financeiras com longa dependência do tipo sazonal. Sendo assim, Bordignon et al. (2007, 2009) propõem os modelos G-GARCH (Gegenbauer-GARCH), que incluem polinômios de Gegenbauer na equação que descreve a volatilidade; PLM-GARCH (*periodic long-memory* GARCH), uma extensão natural do modelo FIGARCH; PLM-LGARCH (*periodic long-memory* log-GARCH), que generaliza o modelo log-GARCH (Ding et al., 1993) e PLM-EGARCH (*periodic long-memory* EGARCH), uma generalização do modelo FIEGARCH.

Segundo Bordignon et al. (2007), resultados sobre a estacionariedade fraca dos processos G-GARCH são desconhecidos. Bordignon et al. (2009) afirma, sem apresentar uma prova formal, que o processo PLM-GARCH não é fracamente estacionário mas, sob certas condições, obtém-se estacionariedade estrita. Propriedades dos processos PLM-LGARCH e PLM-EGARCH não são discutidas por estes autores.

No artigo Lopes e Prass (2013b) (veja Apêndice D) apresentamos a definição formal (veja Definição 4.1 abaixo) e propriedades teóricas dos processos FIEGARCH com sazonalidade (SFIEGARCH) que, assim como os modelos PLM-EGARCH (Bordignon et al., 2009), são uma generalização do modelo EGARCH. Apresentamos no restante desta seção um resumo dos resultados obtidos.

No que segue, $\mathcal{B}$ é o operador defasagem definido por $\mathcal{B}^k(X_t) = X_{t-k}$, para todo $k \in \mathbb{N}$, e $(1-\mathcal{B}^s)^{-d}$ é o operador definido em termos de sua expansão em série de Maclaurin por

$$(1-\mathcal{B}^s)^{-d} = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} (\mathcal{B}^s)^j := \sum_{j=0}^{\infty} \delta_{-d,j} \mathcal{B}^{sj} := \sum_{k=0}^{\infty} \pi_{d,k} \mathcal{B}^k, \tag{4.1}$$

onde $\Gamma(\cdot)$ é a função Gama, dada na expressão (2.2),

$$\pi_{d,k} := 0, \text{ se } \frac{k}{s} \notin \mathbb{N}, \quad \text{e} \quad \pi_{d,sj} := \delta_{-d,j} := \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}, \text{ se } \frac{k}{s} = j, \quad \text{para todo} \quad k, j \in \mathbb{N}.$$

Além disso, $a(\cdot)$ e $b(\cdot)$ são polinômios de ordem p e q, respectivamente, definidos por

$$a(z) = \sum_{i=0}^{p}(-a_i)z^i = 1 - \sum_{i=1}^{p} a_i z^i \quad \text{e} \quad b(z) = \sum_{j=0}^{q}(-b_j)z^j = 1 - \sum_{j=1}^{q} b_j z^j, \tag{4.2}$$

com $a_0 = b_0 = -1$. Para que o operador $\frac{a(\mathcal{B})}{b(\mathcal{B})}$ esteja bem definido, assumimos ainda que $b(\cdot)$ não possui raízes no disco unitário $\{z : |z| \leq 1\}$ e que $a(\cdot)$ e $b(\cdot)$ não possuem raízes em comum.

Definição 4.1. (Processos SFIEGARCH). Seja $\{X_t\}_{t\in\mathbb{Z}}$ o processo estocástico definido por

$$X_t = h_t Z_t, \tag{4.3}$$

$$\ln(h_t^2) = \omega + \frac{a(\mathcal{B})}{b(\mathcal{B})}(1-\mathcal{B}^s)^{-d} g(Z_{t-1}), \quad \text{para todo } t \in \mathbb{Z}, \tag{4.4}$$

onde $\omega \in \mathbb{R}$; $s \in \mathbb{N} \setminus \{0\}$; $d < 0.5$; $\{Z_t\}_{t\in\mathbb{Z}}$ é uma sequência de variáveis aleatórias i.i.d., com média zero e variância um e

$$g(Z_t) = \theta Z_t + \gamma\big[|Z_t| - \mathbb{E}\big(|Z_t|\big)\big], \quad \text{com} \quad \theta, \gamma \in \mathbb{R}, \quad \text{para todo } t \in \mathbb{Z}. \tag{4.5}$$

Então, $\{X_t\}_{t\in\mathbb{Z}}$ é um processo FIEGARCH com sazonalidade, denotado por SFIEGARCH$(p, d, q)_s$.

Observação 4.2. Em Prass (2008) e Lopes e Prass (2013a,b) utiliza-se a notação $\alpha(\cdot)$ e $\beta(\cdot)$, ao invés de $a(\cdot)$ e $b(\cdot)$. Nesse trabalho optamos por utilizar a última notação para evitar que os coeficientes α_i e β_j (aqui denotados a_i e b_j, respectivamente), para $i \in \{0, \cdots, p\}$ e $j \in \{0, \cdots, q\}$, sejam confundidos com os parâmetros α e β da distribuição α-estável. Além disso, para evitar confusão com o parâmetro escala σ da distribuição α-estável, neste trabalho, a variável aleatória σ_t, que representa a volatilidade (veja Prass, 2008; Lopes e Prass, 2013a,b), passa a ser denotada por h_t, para todo $t \in \mathbb{Z}$. Tal notação é muito utilizada para o caso em que as inovações possuem variância infinita pois, nesse caso, associar σ_t^2 (ou, equivalentemente, h_t^2) à variância condicional de X_t, para todo $t \in \mathbb{Z}$, deixa de fazer sentido.

Observação 4.3. Para os processos PLM-EGARCH(p, m, d, q, s), a variável aleatória h_t^2 é definida através da expressão

$$(1-\mathcal{B}^s)^d\phi(\mathcal{B})\big(\ln(h_t^2) - \omega\big) = a(\mathcal{B})Z_t + c(\mathcal{B})\big(|Z_t| - \mathbb{E}(|Z_t|)\big), \quad \text{para todo } t \in \mathbb{Z},$$

onde $a(z) = \sum_{k=1}^{p} a_k z^k$ e $c(z) = \sum_{l=1}^{m} c_l z^l$ são polinômios, respectivamente, de grau p e m, $\phi(z) = \sum_{j=0}^{q-s}\phi_j z^j$ é um polinômio de grau $q-s$ que satisfaz $(1-\mathcal{B}^s)^d\phi(\mathcal{B}) = 1 - b(\mathcal{B})$, onde $b(z) = \sum_{i=1}^{q} b_i z^i$ é um polinômio de ordem q. Entretanto, dado que até o presente momento, não encontramos $d \neq 0$ fracionário, $q \in \mathbb{N}$, $s \in \mathbb{N}^*$ e polinômios $\phi(\cdot)$ e $b(\cdot)$ que satisfaçam as condições impostas por Bordignon et al. (2009), processos PLM-EGARCH não são estudados nesse trabalho.

Uma representação alternativa para o processo $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$, dado em (4.4), é obtida se definirmos $\lambda(\cdot)$ como

$$\lambda(z) := \frac{a(z)}{b(z)}(1-z^s)^{-d} = \sum_{k=0}^{\infty}\lambda_{d,k}z^k, \quad \text{para todo } |z| < 1, \tag{4.6}$$

onde $\{\lambda_{d,k}\}_{k\in\mathbb{Z}}$ são os coeficientes da expansão por série infinita do polinômio $\frac{a(z)}{b(z)}(1-z^s)^{-d}$. Assim, a expressão (4.4) pode ser reescrita como

$$\ln(h_t^2) - \omega = \sum_{k=0}^{\infty}\lambda_{d,k}g(Z_{t-1-k}), \quad \text{para todo } t \in \mathbb{Z}. \tag{4.7}$$

O comportamento assintótico dos coeficientes $\lambda_{d,k}$, quando $k \to \infty$, é apresentado no *Teorema* 1.1. Tal resultado nos permite concluir que, dado $\mathsf{r} > 0$, $\sum_{k=0}^{\infty}|\lambda_{d,k}|^{\mathsf{r}} < \infty$ se, e somente se, $\sum_{k=0}^{\infty}|\pi_{d,k}|^{\mathsf{r}} < \infty$, onde $\{\pi_{d,k}\}_{k\in\mathbb{N}}$ é a sequência definida na expressão (4.1).

O *Teorema* 1.2 fornece uma representação assintótica alternativa para esses coeficientes, que nos permite concluir que (veja a *Observação* 2.7)

$$\lambda_{d,k} \approx \frac{s^{1-d}}{\Gamma(d)k^{1-d}}\frac{a(1)}{b(1)}, \quad \text{quando} \quad k \to \infty. \tag{4.8}$$

A expressão (4.8) nos fornece uma estimativa para o ponto de truncamento do polinômio $\lambda(\cdot)$, necessário para as simulações de Monte Carlo. Além disso, na *Proposição* 1.1 apresentamos a seguinte fórmula de recorrência para o cálculo dos coeficientes $\lambda_{d,k}$, para todo $k \in \mathbb{N}$

$$\lambda_{d,k} = \begin{cases} 1, & \text{se} \quad k = 0, \\ -a_k + \sum_{i=0}^{k-1} \lambda_{d,i}\Big(\sum_{j=0}^{(k-i)\wedge q} \delta^*_{d,\frac{k-i-j}{s}}\, b_j\Big), & \text{se} \quad k \leq p; \\ \sum_{i=0}^{k-1} \lambda_{d,i}\Big(\sum_{j=0}^{(k-i)\wedge q} \delta^*_{d,\frac{k-i-j}{s}}\, b_j\Big), & \text{se} \quad k > p, \end{cases} \tag{4.9}$$

onde $(k-i)\wedge q = \min\{k-i, q\}$ e $\delta^*_{d,m} = \delta_{d,m}$, se $m \in \mathbb{N}$, e $\delta^*_{d,m} = 0$, se $m \notin \mathbb{N}$, com $\delta_{d,m}$, para todo $m \in \mathbb{N}$, definido em (4.1). Tal resultado é fundamental para gerar amostras de processos SFIEGARCH. Na Seção 4.2.1 apresentamos simulações envolvendo o cálculo dos coeficientes do polinômio $\lambda(\cdot)$ para diferentes valores de p, d e q. Tal estudo é importante, do ponto de vista prático, pois precisamos estabelecer o ponto de truncamento da série $\sum_{k=0}^{\infty} \lambda_{d,k} z^k$ a fim de gerar amostras de processos SFIEGARCH.

De (Prass, 2008, proposição 2.2 e lema 2.3) sabemos que $\{g(Z_t)\}_{t\in\mathbb{Z}}$ é um processo ruído branco e sua variância σ_g^2 é dada por

$$\sigma_g^2 = \theta^2 + \gamma^2 - [\gamma\mathbb{E}(|Z_0|)]^2 + 2\theta\gamma\mathbb{E}(Z_0|Z_0|). \tag{4.10}$$

Consequentemente, o processo estocástico $\{g(Z_t)\}_{t\in\mathbb{Z}}$ é (fraca e estritamente) estacionário e ergódico. Na *Observação* 1.7 apresentamos a expressão de σ_g^2 para o caso em que $Z_0 \sim \text{GED}(\nu)$, isto é, distribuição erro generalizada, com parâmetro ν (Nelson, 1991). Tal resultado é importante para as simulações de Monte Carlo.

A representação assintótica dos coeficientes $\lambda_{d,k}$, quando $k \to \infty$, bem como a estacionariedade e ergodicidade de $\{g(Z_t)\}_{t\in\mathbb{Z}}$, são propriedades importantes pois permitem provar que, se $d \leq 0.5$, as variáveis aleatórias $\ln(h_t^2) - \omega$ são finitas quase certamente, para todo $t \in \mathbb{Z}$ (*Lema* 1.1). Mais especificamente, se $d < 0.5$, o processo $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$ é um SARFIMA$(p,0,q)\times(0,d,0)_s$ causal (*Corolário* 1.1). Prova-se ainda que, se $d < 0.5$, as variáveis aleatórias X_t, definidas pela expressão (4.3), são finitas com probabilidade 1, para todo $t \in \mathbb{Z}$ (*Corolário* 1.2). Condições para a invertibilidade, estacionariedade (fraca e estrita) e ergodicidade do processo $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$ são discutidas no *Teorema* 1.3 e no *Lema* 2.1. Ressaltamos que, no *Teorema* 1.3 mostramos que um processo SARFIMA$(p,0,q)\times(0,D,0)_s$ é inversível para $-1 < D < 0.5$, extendendo assim a amplitude do intervalo para o parâmetro D que, na literatura, aparece como $|D| < 0.5$.

Nesse trabalho, nossa atenção é voltada especialmente para os processos $\{X_t\}_{t\in\mathbb{Z}}$ e $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ pois, na prática, $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$ não é observável. Entretanto, as propriedades de $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$ são importantes para que possamos estudar os demais processos de interesse. Prova-se que os processos estocásticos $\{h_t^2\}_{t\in\mathbb{Z}}$ (*Corolário* 2.1), $\{X_t\}_{t\in\mathbb{Z}}$ e, sob certas condições, $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ (*Teorema* 2.1) são estritamente estacionários e ergódicos. Isto porém, não implica estacionariedade fraca pois, para alguma distribuição das variáveis aleatórias $\{Z_t\}_{t\in\mathbb{Z}}$, a média ou a variância destes processos pode não ser finita (veja Nelson, 1991). Uma condição suficiente para que $\{h_t^2\}_{t\in\mathbb{Z}}$ e $\{X_t^2\}_{t\in\mathbb{Z}}$ sejam fracamente estacionários é apresentada no *Teorema* 2.2. Tal condição é satisfeita, por exemplo, quando

$Z_0 \sim \text{GED}(\nu)$, com $\nu > 1$ (*Corolário* 2.2). Apresentamos ainda as expressões para as medidas de assimetria e curtose de um processo SFIEGARCH$(p, d, q)_s$ estacionário, com $\mathbb{E}(|X_0|^4) < \infty$ (*Proposição* 2.1).

Assumindo que $\{X_t\}_{t\in\mathbb{Z}}$ é um processo SFIEGARCH$(p, d, q)_s$ fracamente estacionário (portanto, $d < 0.5$), apresentamos a expressão da função de autocovariância do processo estocástico $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$ (*Lema* 2.2) e, a partir dela, derivamos a expressão da função de autocovariância de $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ (*Teorema* 2.3). O comportamento assintótico dessas funções é dado nos *Corolários* 2.3 e 2.4. Observamos que a representação assintótica apresentada nesse trabalho é diferente daquela dada em Bisognin e Lopes (2009). No *Teorema* 3.1 apresentamos a função densidade espectral do processo $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, que é dada por

$$f_{\ln(X_t^2)}(\lambda) = f_{\ln(h_t^2)}(\lambda) + \frac{C_1}{\pi}\mathfrak{Re}\big(e^{-\mathrm{i}\lambda}\Lambda(\lambda)\big) + f_{\ln(Z_t^2)}(\lambda), \quad \text{para todo } \lambda \in [0, \pi], \tag{4.11}$$

onde

$$C_1 := \text{Cov}\big(\ln(Z_0^2), g(Z_0)\big) = \theta\,\mathbb{E}\big(Z_0 \ln(Z_0^2)\big) + \gamma\mathbb{E}\big(|Z_0| \ln(Z_0^2)\big) - \gamma\mathbb{E}\big(|Z_0|\big)\mathbb{E}\big(\ln(Z_0^2)\big), \tag{4.12}$$

$\Lambda(z) := \lambda(e^{-\mathrm{i}z})$, com $\lambda(\cdot)$ definido em (4.6), e

$$f_{\ln(h_t^2)}(\lambda) = \frac{\sigma_g^2}{2\pi}\frac{|a(e^{-i\lambda})|^2}{|b(e^{-i\lambda})|^2}|1 - e^{-is\lambda}|^{-2d} = \frac{\sigma_g^2}{2\pi}\frac{|a(e^{-i\lambda})|^2}{|b(e^{-i\lambda})|^2}\left[2\sin\left(\frac{s\lambda}{2}\right)\right]^{-2d} \text{ e } \quad f_{\ln(Z_t^2)}(\lambda) = \frac{\sigma^2_{\ln(Z_t^2)}}{2\pi}, \tag{4.13}$$

para todo $\lambda \in [0, \pi]$, são as funções densidade espectral de $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$ e $\{\ln(Z_t^2)\}_{t\in\mathbb{Z}}$, respectivamente.

Exemplos gráficos do comportamento das funções densidade espectral dos processos $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$ e $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ são apresentados no Apêndice F.

Lopes e Prass (2013b) trata ainda da previsão para processos SFIEGARCH e apresenta uma aplicação dos modelos aos log-retornos intradiários do índice S&P500. Mais detalhes sobre previsão em processos SFIEGARCH, bem como a descrição dos resultados apresentados em Lopes e Prass (2013b) sobre esse assunto, são discutidos no Capítulo 6.

4.2 *Processos SFIEGARCH: Estudo Preliminar*

Nesta seção apresentamos simulações envolvendo o cálculo dos coeficientes do polinômio $\lambda(\cdot) = \sum_{k=0}^{\infty}\lambda_{d,k}z^k$, definido em (4.6). Também apresentamos uma comparação entre amostras de processos SFIEGARCH gerados a partir dos mesmos ruídos para diferentes pontos de truncamento.

4.2.1 *Ponto de Truncamento do Polinômio $\lambda(\cdot)$*

No estudo que segue, utilizamos a aproximação dada pela expressão (4.8), ou seja,

$$\lambda_{d,k} \approx \frac{s^{1-d}}{\Gamma(d)k^{1-d}}\frac{\alpha(1)}{b(1)} := \hat{\lambda}_{d,k}. \tag{4.14}$$

para obter uma estimativa do ponto de truncamento m para o polinômio $\lambda(\cdot)$. Adota-se o seguinte critério: dado $\varepsilon > 0$, tomamos o menor inteiro m para o qual $|\hat{\lambda}_{d,m}| \leq \varepsilon$.

Obviamente, a estimativa para o valor de m poderia ser obtida utilizando-se o valor exato de $\lambda_{d,k}$, dado por (4.9), isto é,

$$\lambda_{d,k} = \begin{cases} 1, & \text{se} \quad k = 0; \\ -\alpha_k^* + \displaystyle\sum_{i=0}^{k-1}\lambda_{d,i}\Big(\sum_{j=0}^{(k-i)\wedge q}\delta^*_{d,\frac{k-i-j}{s}}b_j\Big), & \text{se} \quad k \geq 1; \end{cases} \tag{4.15}$$

onde $(k-i)\wedge q = \min\{k-i, q\}$, $\alpha_k^* = \alpha_k$, se $k \leq p$, $\alpha_k^* = 0$, se $k > p$, $\delta_{d,m}^* = \delta_{d,m}$, se $m \in \mathbb{N}$, e $\delta_{d,m}^* = 0$, se $m \in \mathbb{N}$, com $\delta_{d,m}$, para todo $m \in \mathbb{N}$, definido em (4.1). Entretanto, nesse caso, para encontrar $\lambda_{d,m}$ tal que $|\lambda_{d,m}| < \varepsilon$, teríamos que calcular todos os valores de $\lambda_{d,k}$, para $k \leq m$. Por outro lado, da expressão (4.14), é fácil ver que

- se $\frac{a(1)}{b(1)} > 0$, segue que $\hat{\lambda}_{d,k} < 0$, se $-0.45 \leq d < 0$, e $\hat{\lambda}_{d,k} > 0$, se $0 < d \leq 0.45$, para todo $k \in \mathbb{N}$;
- dado $\varepsilon > 0$,
$$\left|\frac{s^{1-d}}{\Gamma(d)m^{1-d}}\frac{a(1)}{b(1)}\right| \leq \varepsilon \quad \Longleftrightarrow \quad m \geq \left(\left|\frac{s^{1-d}}{\varepsilon\Gamma(d)}\frac{a(1)}{b(1)}\right|\right)^{\frac{1}{1-d}}; \tag{4.16}$$
- o valor de m não depende dos valores p e q, mas sim do quociente $\left(\left|\frac{a(1)}{b(1)}\right|\right)^{\frac{1}{1-d}}$;
- pela expressão (4.16), o valor do ponto de truncamento para qualquer $s > 1$ é dado por $\mathfrak{m}_s = s \times m$, onde m é o valor obtido quando $s = 1$.

Observamos que a aproximação dada em (4.14) é válida quando $k \to \infty$. Sendo assim, nesse estudo também analisamos se, para o valor de m obtido com tal aproximação, $|\lambda_{d,m}|$ é de fato menor do que ε. Como veremos a seguir, em alguns casos, o valor de m estimado pela expressão (4.16) é muito grande, mesmo quando $s = 1$. Sendo assim, investigamos a possibilidade de utilizar o valor $m = 50000$, para $s \in \{1, 2\}$. Obviamente, para outros valores de s esse valor deve ser proporcionalmente maior.

Para essa simulação fixamos:

- $s \in \{1, 2\}$, p e $q \in \{0, 1, 2\}$. Os coeficientes a_i e b_j dos polinômios $a(\cdot)$ e $b(\cdot)$, respectivamente, foram escolhidos aleatoriamente. Além disso, dois casos são considerados,
  - Caso 1 (C1): consideramos $a(z) := f_p(z)$ e $b(z) := g_q(z)$, para $p, q \in \{0, 1, 2\}$;
  - Caso 2 (C2): consideramos $a(\cdot) := g_p(\cdot)$ e $b(\cdot) := f_q(\cdot)$, para $p, q \in \{0, 1, 2\}$;

  onde $f_p(\cdot)$ e $g_q(\cdot)$, para $p, q \in \{0, 1, 2\}$ são os polinômios dados por

$$\begin{array}{llll} f_0(z) = 1, & f_1(z) = 1 - 0.4z & \text{e} & f_2(z) = 1 - 0.2z + 0.3z^2; \\ g_0(z) = 1, & g_1(z) = 1 - 0.8z & \text{e} & g_2(z) = 1 - 0.2z + 0.5z^2. \end{array}$$

  Com isso, é possível analisar a influência da razão $\frac{a(1)}{b(1)}$, no cálculo dos coeficientes $\lambda_{d,k}$ e do ponto de truncamento m. Os valores desse quociente, para ambos os casos considerados nesta simulação, são dados na Tabela 4.1;
- $d \in \{-0.45, -0.4, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3, 0.4, 0.45\}$. O caso $p = q = d = 0$ foi excluído dessa simulação;
- o critério de convergência considerado é $\varepsilon \in \{10^{-3}, 10^{-4}\}$;
- por ser mais compacta, em tabelas utilizamos a notação científica e-b ao invés de $\times 10^{-b}$.

Na Tabela E.1 do Apêndice E apresentamos os valores do ponto de truncamento m, estimados a partir da expressão (4.16) sob a restrição $|\hat{\lambda}_{d,m}| \leq \varepsilon$, para valores de p, q, d e ε descritos acima. Como mencionamos anteriormente, $\mathfrak{m}_2 = 2\mathfrak{m}_1$, onde $\mathfrak{m}_s$ é o valor de m estimado quando $s \in \{1, 2\}$. Portanto, somente os valores para $s = 1$ são reportados. Apresentamos também o valor da diferença $diff = \lambda_{d,m} - \hat{\lambda}_{d,m}$, para o valor de m estimado pela expressão (4.16), $\hat{\lambda}_{d,m}$ obtido pela aproximação (4.14) e $\lambda_{d,m}$ dado pela expressão (4.9). O valor dessa diferença permite-nos analisar quão acurada

Tabela 4.1: Valores da razão $\frac{a(1)}{b(1)}$ para os polinômios $a(\cdot)$ e $b(\cdot)$ utilizados na simulação.

Grau	Caso 1 (C1)			Caso 2 (C2)		
	$q=0$	$q=1$	$q=2$	$q=0$	$q=1$	$q=2$
$p=0$	1.0000	5.0000	0.7692	1.0000	1.6667	0.9091
$p=1$	0.6000	3.0000	0.4615	0.2000	0.3333	0.1818
$p=2$	1.1000	5.5000	0.8462	1.3000	2.1667	1.1818

é a aproximação proposta e ainda, se $|\lambda_{d,m}| \leq \varepsilon$ pois $\lambda_{d,m} = \hat{\lambda}_{d,m} + \mathit{diff}$. Nessa tabela também apresentamos o valor do coeficiente $\lambda_{d,m}$ e a diferença $\mathit{diff} = \lambda_{d,m} - \hat{\lambda}_{d,m}$, para $m = 50000$. Tal valor foi escolhido a fim de analisarmos a possibilidade de utilizá-lo como valor padrão nas próximas simulações (veja a Seção 4.2.2).

Pelos resultados apresentados nas Tabelas E.1 - E.7 (veja o Apêndice E) conclui-se que,

- para os valores de m obtidos a partir da expressão (4.16), a diferença entre $\lambda_{d,m}$ e $\hat{\lambda}_{d,m}$ é menor ou igual a 4.08e-4 (valor absoluto), o que mostra que a expressão (4.14) fornece uma boa aproximação para o valor real do parâmetro;

- em vários casos obtivemos $|\hat{\lambda}_{d,m}| \leq \varepsilon$, mas $|\lambda_{d,m}| > \varepsilon$, para os valores de m obtidos a partir da expressão (4.16). Para destacar esses casos na Tabela E.1, utilizamos o símbolo "†" ao lado do valor de diff. Observando o sinal de diff conclui-se que, em tais casos, $\lambda_{d,m} < -\varepsilon \leq \hat{\lambda}_{d,m} < 0$ (se $d < 0$) ou $0 < \hat{\lambda}_{d,m} \leq \varepsilon < \lambda_{d,m}$ (se $d > 0$). Como $\lambda_{d,m} = \hat{\lambda}_{d,m} + \mathit{diff}$ e $|\mathit{diff}| \leq 5.54\text{e-}4$, se $|\lambda_{d,m}| > \varepsilon$, conclui-se que $|\lambda_{d,m}|$ será, no máximo, igual a 1.0554×10^{3}, quando $\varepsilon = 10^{-3}$, e 1.554×10^{-4}, quando $\varepsilon = 10^{-4}$.

- para $m = 50000$, $s = 1$ e todos os valores de d, p e q, conclui-se que $\hat{\lambda}_{d,m}$ é uma boa aproximação pois o valor de $|\mathit{diff}| = |\lambda_{d,m} - \hat{\lambda}_{d,m}|$ é muito pequeno se comparado ao valor de $\lambda_{d,m}$.

- para $m = 50000$, $s = 2$, $p = 0$, $q = 1$ (no caso C1) e todos os valores de d (com excessão de $d = -0.3$), os valores de diff mostram que $|\hat{\lambda}_{d,m}|$ é quase duas vezes maior do que $|\lambda_{d,m}|$. Comportamento semelhante é observado nos seguintes casos
  $p = 0$, $q = 1$ e $d \in \{-0.30, 0.30\}$ (C2)
  $p = 0$, $q = 2$ e $d \in \{-0.45, -0.10, 0.10\}$ (C1), $d \in \{-0.45, -0.20, 0.10\}$ (C2)
  $p = 1$, $q = 0$ e $d \in \{-0.30, -0.20, 0.20, 0.30\}$ (C1), $d \notin \{-0.45, 0.10, 0.40\}$ (C2)
  $p = 1$, $q = 1$ e $d \neq 0.20$ (C1), $d \notin \{-0.40, -0.10, 0.45\}$ (C2)
  $p = 1$, $q = 2$ e $d \notin \{-0.40, -0.30, 0.30, 0.45\}$ (C1), $d \notin \{-0.40, 0.45\}$ (C2)
  $p = 2$, $q = 0$ e $d = 0.20$ (C1 e C2)
  $p = 2$, $q = 1$ e todos os valores de d (C1), $d \in \{-0.30, 0.30\}$ (C2).

  Nos demais casos, o valor de $|\mathit{diff}| = |\lambda_{d,m} - \hat{\lambda}_{d,m}|$ é pequeno se comparado ao valor de $\lambda_{d,m}$.

- para $m = 50000$ e $s \in \{1, 2\}$ obtivemos: 42 casos para os quais $10^{-3} \leq |\lambda_{d,m}| \leq 7.28 \times 10^{-3}$ (em todos $d \geq 0.3$) e 60 casos para os quais $10^{-4} < |\lambda_{d,m}| \leq 10^{-3}$ (em todos $d \geq 0.2$). Tais situações foram destacadas na Tabela E.1, utilizando-se, respectivamente, os símbolos "†" e "⋆", ao lado do valor de diff. Em todos os outros casos $|\lambda_{d,m}| \leq 10^{-4}$.

Nesta seção analisamos apenas a relação entre os valores aproximados e reais do coeficiente $\lambda_{d,m}$ para um dado ponto de truncamento m. Na próxima seção estudamos com mais detalhes a influência do ponto de truncamento m na geração de amostras de processos SFIEGARCH. A principal questão a ser analisada é quão "diferentes" são duas amostras $\{X_t(m_1)\}_{t=1}^{n}$ e $\{X_t(m_2)\}_{t=1}^{n}$ obtidas considerando-se, respectivamente, dois pontos de truncamento m_1 e m_2.

4.2.2 *Simulações Envolvendo Processos SFIEGARCH*

No que segue apresentamos resultados referentes à simulação de processos SFIEGARCH, para diferentes valores de p, d, q e s. O objetivo dessa simulação é comparar os valores de $\{X_t\}_{t=1}^n$, obtidos a partir de valores de $\{h_t\}_{t=1}^n$ gerados utilizando-se diferentes pontos de truncamento para o polinômio $\lambda(\cdot)$, definido em (4.6). O procedimento é explicado em detalhes no que segue.

- Fixamos o tamanho amostral $n = 1000$ e $s \in \{1, 12\}$.
- Consideramos p e $q \in \{0, 1, 2\}$ e os coeficientes a_i e b_j dos polinômios $a(\cdot)$ e $b(\cdot)$ são os mesmos utilizados na Seção 4.2.1 (caso C1).
- $d \in \{0.1, 0.2, 0.3, 0.4\}$. Além disso, assumimos $\omega = 0$, $\theta = -0.15$ e $\gamma = 0.25$.
- Para cada valor de p, d, q, s, consideramos dois pontos de truncamento: o valor dado na Tabela E.1, para $\varepsilon = 10^{-3}$, que denotamos por m_1, e os valores padrão propostos $m_2 = 50000$ (para valores pequenos de s) e $m_3 = 100000$ (para valores grandes de s).
- Para gerar a série temporal $\{Z_t\}_{t=-\mathfrak{N}}^n$, onde $\mathfrak{N} = \max\{m_1, m_2\}$, assumimos $Z_t \sim \mathcal{N}(0, 1)$, para todo $t \in \mathbb{Z}$. Sendo assim, $\mathbb{E}(|Z_t|) = \sqrt{2/\pi}$, e a expressão da função $g(\cdot)$ torna-se

$$g(Z_t) = \theta Z_t + \gamma\big[|Z_t| - \sqrt{2/\pi}\big], \quad \text{para todo } t \in \mathbb{Z}.$$

- Denotamos por $\{h_t(m)\}_{t=1}^n$, $\{\ln(h_t^2(m))\}_{t=1}^n$ e $\{X_t(m)\}_{t=1}^n$ as séries temporais obtidas utilizando-se o ponto de truncamento $m \in \{m_1, m_2, m_3\}$.

Fixados esses valores, seguimos os passos descritos abaixo:

Passo 1. obtemos $\{\lambda_{d,k}\}_{k=0}^m$ através da expressão (4.9), para cada valor de $m \in \{m_1, m_2, m_3\}$;

Passo 2. para cada valor de $m \in \{m_1, m_2, m_3\}$, geramos a série temporal $\{\ln(h_t^2(m))\}_{t=1}^n$, utilizando a expressão (4.4) e a versão truncada de (4.6). Note que apenas os valores $\{Z_t\}_{t=-m}^n$, da série temporal $\{Z_t\}_{t=-\mathfrak{N}}^n$, são utilizados para gerar $\{\ln(h_t^2(m))\}_{t=1}^n$;

Passo 3. as séries temporais $\{h_t(m)\}_{t=1}^n$ e $\{X_t(m)\}_{t=1}^n$ são obtidas, respectivamente, através das igualdades $h_t(m) = 0.5\exp\{\ln(h_t^2(m))\}$ e (4.3). Observamos que, neste passo, apenas os valores $\{Z_t\}_{t=1}^n$ são utilizados para gerar $\{X_t(m)\}_{t=1}^n$.

Medidas para Comparação

No que segue, Y_t representa qualquer uma das duas variáveis aleatórias $\ln(h_t^2)$ ou X_t, para todo $t \in \mathbb{Z}$. As medidas utilizadas para comparar duas amostras $\{Y_t(m_1)\}_{t=1}^n$ e $\{Y_t(m_j)\}_{t=1}^n$, para $j \in \{2, 3\}$, são: a diferença mínima (MIN); a diferença máxima (MAX); a diferença média absoluta (MAE) e o desvio padrão das diferenças (SD), definidos por

$$\text{MIN} = \min_{t=1,\cdots,n}\{|e_t|\}, \quad \text{MAX} = \max_{t=1,\cdots,n}\{|e_t|\}, \quad \text{MAE} = \frac{1}{n}\sum_{t=1}^n |e_t| \quad \text{e} \quad \text{SD} = \sqrt{\frac{1}{n}\sum_{t=1}^n (e_t - \bar{e})^2},$$

onde $\bar{e} = \frac{1}{n}\sum_{t=1}^n e_t$ e $e_t := Y_t(m_1) - Y_t(m_j)$, para cada $t = 1, \cdots, n$ e $j \in \{2, 3\}$.

Nas Tabelas E.8 e E.9 (veja o Apêndice E) apresentamos os resultados da simulação para $s = 1$. As Tabelas E.10 e E.11 (veja o Apêndice E) referem-se ao caso $s = 12$. Os valores reportados nas

tabelas correspondem às estatísticas MIN, MAX, MAE, SD e $\bar{e}$ definidas acima. Além disso, apresentamos também os valores mínimo e máximo das variáveis aleatórias $\{X_t(m_1)\}_{t=1}^n$ e $\{\ln(h_t^2(m_1))\}_{t=1}^n$, representados em forma de intervalo nas colunas denotadas por RANGE(X_t) e RANGE($\ln(h_t^2)$). Tais valores no permitem ter uma idéia da magnitude dos valores das séries temporais e nos permitem concluir se a diferença entre os valores X_t (ou $\ln(h_t^2)$) obtidos com diferentes valores de m é significativa ou não.

4.3 *Inovações com Distribuição α-Estável*

Na Seção 4.1 introduzimos os processos SFIEGARCH$(p,d,q)_s$ sob a hipótese que a sequência de variáveis aleatórias $\{Z_t\}_{t\in\mathbb{Z}}$, dadas na Definição 4.1, possui variância finita. Nesta seção discutimos o caso em que Z_t, para todo $t\in\mathbb{Z}$, é uma variável aleatória estável.

A classe das distribuições estáveis tornou-se popular na modelagem financeira por duas razões principais. Em primeiro lugar, distribuições estáveis possuem a propriedade de estabilidade e domínios da atração (veja o Capítulo 3). Sendo assim, fornecem boas aproximações para uma ampla variedade de dados que são observados na prática (Kozubowski et al., 2003). Em segundo lugar, distribuições da família α-estável permitem a modelagem de caudas pesadas e da assimetria, características que são frequentemente observadas em dados financeiros.

Como apontado por Panorska et al. (1995), modelos clássicos que consideram a distribuição t-Student como a distribuição condicional das inovações, permitem a modelagem de caudas mais pesadas do que os modelos Gaussianos. Entretanto, em contraste com a distribuição Gaussiana, a distribuição t-Student não apresenta a propriedade de estabilidade, tão desejada na prática.

No artigo Prass, Lopes e Crato (2013) consideramos apenas distribuições α-estáveis simétricas ($\mu=\beta=0$), denotadas por $S\alpha S$, e introduzimos os processos α-SFIEGARCH$(p,d,q)_s$, para $\alpha\in(1,2]$. Observamos que, quando $\alpha=2$, obtemos o processo SFIEGARCH$(p,d,q)_s$ com inovações Gaussianas, estudados em Lopes e Prass (2013b). O resumo dos resultados apresentados em Prass, Lopes e Crato (2013) é dado a seguir.

Definição 4.2. (Processos α-SFIEGARCH). Seja $\{X_t\}_{t\in\mathbb{Z}}$ o processo estocástico definido por

$$X_t = h_t Z_t \quad \text{e} \quad \ln(h_t^2) = \omega + \frac{a(\mathcal{B})}{b(\mathcal{B})}(1-\mathcal{B}^s)^{-d} g(Z_{t-1}), \quad \text{para todo } t\in\mathbb{Z}, \tag{4.17}$$

onde $\omega\in\mathbb{R}$; $s\in\mathbb{N}\setminus\{0,1\}$; $\{Z_t\}_{t\in\mathbb{Z}}$ é uma sequência de variáveis aleatórias i.i.d. $S\alpha S$ com parâmetro de escala $\sigma=1$; $a(\cdot)$ e $b(\cdot)$ são os polinômios definidos em (4.2), $(1-\mathcal{B}^s)^{-d}$ é o operador definido em (4.1) e $g(\cdot)$ é definida pela expressão (4.5). Então, $\{X_t\}_{t\in\mathbb{Z}}$ é um processo FIEGARCH sazonal com inovações α-estáveis, denotado por α-SFIEGARCH$(p,d,q)_s$.

Observação 4.4. Sem perda de generalidade, o valor $\sigma=1$, dado na Definição 4.2, pode ser substituído por qualquer $\sigma>0$. Entretanto, para que o processo seja unicamente determinado, no momento da estimação dos parâmetros, deve-se assumir que esse valor é conhecido e fixo (veja Prass, Lopes e Crato, 2013, Observação 1).

Pela expressão (4.17) conclui-se que, se as variáveis aleatórias $\{h_t\}_{t\in\mathbb{Z}}$ são finitas com probabilidade 1, então h_t é $\mathcal{F}_{t-1}$-mensurável, onde $\mathcal{F}_t$ denota a σ-álgebra gerada por $\{Z_k\}_{k\leq t}$. Entretanto, dado que Z_0 não possui variância finita, as variáveis aleatórias h_t não correspondem ao desvio padrão de $X_t|\mathcal{F}_{t-1}$, para $t\in\mathbb{Z}$ (veja a Seção 4.1 para o caso de variância finita). Por outro lado, utilizando-se os resultados do Capítulo 3, mostra-se que, se $Z_t\sim S_\alpha(0,\sigma,0)$, com $\alpha\in(1,2]$, então $X_t|\mathcal{F}_{t-1}\sim S_\alpha(0,\sigma h_t,0)$, para todo $t\in\mathbb{Z}$. Em particular, quando $\sigma=1$, h_t é o parâmetro de escala

da distribuição de $X_t|\mathcal{F}_{t-1}$.

Dado que h_t é finita com probabilidade 1 se, e somente se, a variável aleatória $\ln(h_t^2)$ é finita com probabilidade 1, seguimos os mesmo passos da Seção 4.1 e reescrevemos $\ln(h_t^2)$ como

$$\ln(h_t^2) - \omega = \frac{a(\mathcal{B})}{b(\mathcal{B})}(1-\mathcal{B}^s)^{-d} g(Z_{t-1}) = \sum_{k=0}^{\infty} \lambda_{d,k} g(Z_{t-1-k}), \quad \text{para todo } t \in \mathbb{Z}, \tag{4.18}$$

com $\lambda_{d,k}$, para todo $k \in \mathbb{N}$, definido através da expressão (4.6). Dessa forma, mostra-se que (4.18) converge quase certamente se, e somente se, $(1-d)\alpha > 1$ (veja o *Teorema* 2). No *Teorema* 2 mostra-se ainda que, se $(1-d)\alpha > 1$, a série (4.18) converge em norma em $\mathcal{L}^r$, para todo $0 < r < \alpha$. Além disso, se $d \leq 0$, a série (4.18) converge absolutamente com probabilidade 1.

Observamos que, para provar o *Teorema* 2, primeiramente mostramos que o processo estocástico $\{g(Z_t)\}_{t\in\mathbb{Z}}$ possui momento de ordem r finito, para todo $0 < r < \alpha$ (veja o *Lema* 1). Além disso, no *Teorema* 1, mostramos que

$$\mathbb{P}(|g(Z_0)| > x) \sim C_{\alpha,\sigma}\left(|\theta+\gamma|^\alpha + |\theta-\gamma|^\alpha\right) x^{-\alpha}, \quad \text{quando } x \to \infty,$$

onde

$$C_{\alpha,\sigma} = \sigma^\alpha \left(\int_0^\infty x^{-\alpha}\sin(x)dx\right)^{-1} = \frac{(1-\alpha)\sigma^\alpha}{\Gamma(2-\alpha)\cos\left(\frac{\pi\alpha}{2}\right)}, \quad \text{se } \alpha \neq 1,$$

ou seja, o comportamento assintótico das caudas da função de distribuição da variável aleatória $g(Z_t)$ é semelhante ao apresentado pela variável aleatória Z_t, para todo $t \in \mathbb{Z}$ (Teorema 3.2 do Capítulo 3).

A estacionariedade estrita e ergodicidade dos processos α-SFIEGARCH são discutidas no *Teorema* 3. Em particular, mostra-se que, se $d < 1 - \frac{1}{\alpha}$, então os processos estocásticos $\{X_t\}_{t\in\mathbb{Z}}$, $\{h_t\}_{t\in\mathbb{Z}}$ e $\{\ln(h_t^2) - \omega\}_{t\in\mathbb{Z}}$, são estritamente estacionários e ergódicos. O *Teorema* 3 também trata da invertibilidade do processo $\{\ln(h_t^2) - \omega\}_{t\in\mathbb{Z}}$, isto é, mostra-se que , se $|d| < 1 - \frac{1}{\alpha}$ e $a(z) \neq 0$ no disco fechado $\{z : |z| \leq 1\}$, então

$$\lim_{m\to\infty} \mathbb{E}\left(\left|\sum_{k=0}^{m} \tilde{\lambda}_{d,k}(\ln(h_t^2) - \omega) - g(Z_{t-1})\right|^r\right) = 0, \quad \text{para todo} \quad 0 < r < \alpha,$$

onde os coeficientes $\tilde{\lambda}_{d,k}$, para todo $k \in \mathbb{N}$, são definidos por

$$\sum_{k=0}^{\infty} \tilde{\lambda}_{d,k} z^k := \frac{b(z)}{a(z)}(1-z^s)^d = \lambda^{-1}(z), \quad |z| < 1.$$

O artigo Prass, Lopes e Crato (2013) também trata da estimação dos parâmetros do modelo α-SFIEGARCH utilizando-se inferência Bayesiana. Esse assunto é discutido com mais detalhes no Capítulo 5.

Capítulo 5

Estimação dos Parâmetros do Modelo

O fenômeno de longa dependência em séries temporais foi observado primeiramente por Hurst (1951, 1957) quando analisava a série temporal de níveis do rio Nilo. Mais tarde, esse conceito aparece nos trabalhos de Mandelbrot e Wallis (1968) e McLeod e Hipel (1978), também na área de hidrologia. Desde então, tal fenômeno tem sido tema de estudo e vários métodos destinados a identificar e estimar o parâmetro de longa dependência H, denominado $coeficiente\ de\ Hurst$, aparecem na literatura.

Os primeiros modelos para séries temporais com longa dependência foram introduzidos por Mandelbrot (1965), Mandelbrot e Van Ness (1968) e Mandelbrot e Wallis (1968, 1969), no caso de tempo contínuo, e por Granger e Joyeux (1980) e Hosking (1981), no caso de tempo discreto. Mais recentemente, para modelar a longa dependência na volatilidade, foram introduzidos modelos não-lineares tais como o FIGARCH (Baillie et al., 1996), o FIEGARCH (Bollerslev e Mikkelsen, 1996; Lopes e Prass, 2013a) e o modelo LMSV (Breidt et al., 1998). Para alguns desses modelos, generalizações que também levam em conta a sazonalidade foram propostas. Alguns exemplos são os modelos SARFIMA (Porter-Hudak, 1990; Bisognin, 2007; Bisognin e Lopes, 2007, 2009), k-factor Gegenbauer (Giraitis e Leipus, 1995) e k-factor GARMA (Woodward et al., 1998), que generalizam o modelo ARFIMA, e os modelos PLM-EGARCH (Bordignon et al., 2009) e SFIEGARCH (Lopes e Prass, 2013b), que generalizam o modelo FIEGARCH.

No caso da variância finita, a estimação dos parâmetros de processos FIEGARCH/SFIEGARCH é feita através do método da quase-máxima verossimilhança, que estima todos os parâmetros do modelo com exceção daqueles associados às inovações. No caso dos modelos como ARFIMA/SARFIMA ou FIEGARCH/SFIEGARCH com inovações α-estáveis, ainda não há consenso na literatura sobre qual o melhor caminho a seguir: se a estimação de todos os parâmetros de uma única vez, pelo método da máxima verossimilhança e/ou pseudo-máxima verossimilhança, ou se a estimação em duas etapas, sendo uma para a estimação dos parâmetros da distribuição e outra para os demais parâmetros do modelo.

Neste capítulo descrevemos os principais métodos para estimação, tanto do parâmetro de longa dependência d, quanto dos demais parâmetros dos modelos SFIEGARCH descritos no Capítulo 4. Dentre os métodos heurísticos mais utilizados na estimação do parâmetro H destacamos as estatísticas R/S, KPSS e V/S. Ressaltamos que os métodos heurísticos são em geral utilizados como uma primeira estimativa do parâmetro H e em testes de hipótese.

Além dos métodos heurísticos, discutimos também métodos paraméricos e semi-paramétricos para estimação dos parâmetros do modelo. No domínio do tempo, apresentamos os estimadores da máxima (pseudo-máxima ou quase-máxima) verossimilhança (MLE) e, no domínio da frequência, o estimador de Whittle. O MLE nos permite estimar todos os parâmetros do modelo de uma só vez,

inclusive os parâmetros relacionados à função de distribuição das inovações. Entretanto, para tais estimadores, é necessário impormos hipóteses quanto a função de distribuição dos dados. O estimador de Whittle é mais flexível no sentido de que não são necessárias hipóteses quanto à distribuição dos dados. Tal estimador também possui uma versão modificada que pode ser aplicada para processos com distribuição α-estável.

5.1 Estatísticas Baseadas em Somas Parciais

Considere um processo estocástico $\{X_t\}_{t\in\mathbb{Z}}$ e uma série temporal $\{X_t\}_{t=1}^n$, obtida a partir desse processo. No que segue, apresentamos as principais estatísticas de teste, baseadas nas somas parciais de $\{X_t\}_{t=1}^n$, utilizadas para testar a existência de longa dependência.

Nesta seção, adotamos a seguinte notação:

- $\{W(t)\}_{t\in[0,1]}$ é o Movimento Browniano padrão.
- $\{W^0(t)\}_{t\in[0,1]}$, onde $W^0(t) = W(t) - tW(1)$, é uma ponte Browniana.
- Dada uma série temporal $\{X_t\}_{t=1}^n$, com tamanho amostral n, $\bar{X}$ e S_n^2 são, respectivamente, a média e a variância amostral de $\{X_t\}_{t=1}^n$, isto é,

$$\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t, \quad \text{e} \quad S_n^2 = \frac{1}{n}\sum_{t=1}^{n}(X_t - \bar{X})^2.$$

5.1.1 Estatística R/S

A estatística R/S é definida por

$$Q_n := \frac{1}{S_n}\left[\max_{1\le k\le n}\sum_{j=1}^{k}(X_j - \bar{X}) - \min_{1\le k\le n}\sum_{j=1}^{k}(X_j - \bar{X})\right], \tag{5.1}$$

onde $S_n = \sqrt{S_n^2}$ é o desvio padrão amostral. Tal estatística foi introduzida por Hurst (1951) com o nome *rescaled adjusted range,* com o propósito de testar a existência de longa dependência em séries temporais, bem como estimar o parâmetro H, denominado *coeficiente de Hurst*, relacionado à essa característica.

Se as variáveis aleatórias $\{X_t\}_{t\in\mathbb{Z}}$ são i.i.d., com distribuição comum $\mathcal{N}(\mu_X, \sigma_X^2)$, então $Q_n/\sqrt{n}$ converge fracamente para uma variável aleatória que está no domínio de atração de uma ponte Browniana no intervalo $[0,1]$. Isto é,

$$\frac{Q_n}{\sqrt{n}} \Longrightarrow U_{R/S} \quad \text{onde} \quad U_{R/S} = \max_{0\le t\le 1} W^0(t) - \min_{0\le t\le 1} W^0(t). \tag{5.2}$$

Hurst (1951) observa que, para muitas séries temporais observadas, é válida a relação

$$\mathbb{E}(Q_n) \sim C_H n^H, \quad \text{quando } n\to\infty,$$

onde $C_H > 0$, com $H \in (0.5, 1)$. Feller (1951) e Annis e Lloyd (1976), mostram que quando as observações $\{X_t\}_{t=1}^n$ são provenientes de um processo estocástico com curta dependência

$$\mathbb{E}(Q_n) \sim C n^{0.5}, \quad \text{quando } n\to\infty,$$

onde $C > 0$.

5.1.2 *Estimação do Coeficiente de Hurst H*

No que segue, descrevemos os passos para a estimação do coeficiente de Hurst H, dada uma série temporal $\{X_t\}_{t=1}^n$, de tamanho amostral n.

1. Subdivide-se as n observações da série temporal $\{X_t\}_{t=1}^n$ em K blocos adjacentes, cada um de tamanho $\lfloor \frac{n}{K} \rfloor$, onde $\lfloor \cdot \rfloor$ denota a função maior inteiro. Sem perda de generalidade, no que segue, assumimos que $\frac{n}{K} \in \mathbb{N}$. Obtém-se assim a sequência dos pontos iniciais do intervalo $\{t_i\}_{i=1}^K$, onde $t_i = (i-1)\frac{n}{K} + 1$, para todo $i = 1, \cdots, K$.

2. Seja $N \in \{N_1, N_2, \cdots, N_\ell\}$, com $N_1 \leq \cdots \leq N_\ell \leq n$, onde $N_1 := N_{min}$ é o tamanho amostral mínimo a ser utilizado. O número $N \in \{N_1, N_2, \cdots, N_\ell\}$ representa o tamanho das amostras da série temporal $\{X_t\}_{t=1}^n$ para as quais calculamos a estatística R/S. Para cada valor de N, calcula-se a estatística $Q_{N,i}$, dada na expressão (5.1), iniciando no ponto t_i, para os valores de $i \in 1, \cdots, \ell$, tais que $t_i + N \leq n$. Observe que, para valores de $N \leq \frac{n}{K}$, obtemos K diferentes valores da estatística. Por outro lado, para valores de N aproximando-se de n, obtemos menos valores e, se $N > n - \frac{n}{K}$, obtemos apenas um valor.

3. Escolhendo pontos logaritmicamente espaçados para os valores de N, constrói-se o gráfico de $\log(Q_{N,i})$ versus $\log(N)$. Ajustando-se uma reta a esses pontos, encontra-se o estimador de H. Para uma série temporal com característica de longa dependência a reta deve ter inclinação $H > 0.5$.

4. Taqqu et al. (1995) sugerem utilizar apenas os valores da estatística Q_N para valores de N dentro de um certo intervalo. Nas rotinas encontradas na página oficial do autor Murad S. Taqqu encontramos como padrão o intervalo $10^{0.7} \leq N \leq 10^{2.5}$. Obviamente, tais valores dependem do tamanho amostral. Segundo os autores, valores muito pequenos de N influenciam na estimação de H devido a efeitos transientes (para mais detalhes, veja Taqqu et al., 1995), enquanto que valores muito altos de N geram poucos valores da estatística R/S.

5.1.3 *Estatística R/S Modificada*

Lo (1991) observa que a estatística R/S não é robusta para curta dependência e propõe uma versão modificada dessa estatística. Tal modificação é conhecida na literatura como *estatística R/S modificada* e, neste trabalho, é denotada por R/S$_m$. A estatística R/S$_m$ é dada por

$$\tilde{Q}_{n,q} = \frac{1}{\hat{\sigma}_n(q)} \left[\max_{1 \leq k \leq n} \sum_{j=1}^{k} (X_j - \bar{X}) - \min_{1 \leq k \leq n} \sum_{j=1}^{k} (X_j - \bar{X}) \right],$$

onde

$$\hat{\sigma}_n^2(q) = \hat{\gamma}_X(0) + 2 \sum_{j=1}^{q} \omega_j(q) \hat{\gamma}_X(j), \quad \omega_j(q) = 1 - \frac{j}{q+1}, \quad q < n, \tag{5.3}$$

e $\hat{\gamma}_X(\cdot)$ é a função de autocovariância amostral do processo $\{X_t\}_{t \in \mathbb{Z}}$. Os pesos $\omega_j(q)$ são os mesmos utilizados por Newey e West (1987), que sugerem escolher $q = 4\left(\frac{n}{100}\right)^{2/9}$. Zivot e Wang (2005) utilizam $q = 4\left(\frac{n}{100}\right)^{1/4}$.

Lo (1991) apresenta a distribuição limite da estatística R/S modificada sob ambas as hipóteses, de longa e curta dependência. Além disso, Lo (1991) afirma que a estatística R/S modificada é robusta para curta dependência. Lo (1991) mostra que, na ausência de longa dependência e com a

escolha correta de q, a variável aleatória $\frac{\tilde{Q}_{n,q}}{\sqrt{n}}$ também converge fracamente para a variável aleatória $U_{R/S}$, dada pela expressão (5.2).

A distribuição da variável aleatória $U_{R/S}$ é conhecida (veja Kennedy, 1976) e é dada por

$$\mathbb{P}(U_{R/S} \leq x) = 1 - 2\sum_{n=1}^{\infty}(4x^2n^2 - 1)e^{-2x^2n^2}.$$

Segue que $\mathbb{P}(0.809 \leq U_{R/S} \leq 1.862) = 0.95$. Lo (1991) considera o intervalo $[0.809, 1.862]$ como região de aceitação assintótica, à 95% de confiança, para testar a hipótese nula

$$H_0 : \text{não existe longa dependência, isto é, } H = 0.5$$

versus a hipótese alternativa,

$$H_1 : \text{existe longa dependência, isto é, } 0.5 < H < 1.$$

Ao contrário da estatística R/S, a estatística R/S modificada não apresenta um método gráfico de estimação do coeficiente de Hurst. O método proposto por Lo (1991) apenas indica se existe longa dependência ou não. Por outro lado, estudos empíricos apresentados por Teverovsky et al. (1999) mostram que, se $\{X_t\}_{t=1}^n$ provém de um ruído Gaussiano fracionário (FGN), com $H > 0.5$, então, $\frac{\tilde{Q}_{n,q}}{\sqrt{n}}$ é estritamente decrescente, com respeito a q. Além disso, o gráfico de $\log\left(\frac{\tilde{Q}_{n,q}}{\sqrt{n}}\right)$ versus $\log(q)$ é uma reta com inclinação aproximadamente igual a $0.5 - H$. Além disso, no caso de dados i.i.d. (isto é, $H = 0.5$), $\frac{\tilde{Q}_{n,q}}{\sqrt{n}}$ é aproximadamente constante.

Embora no caso da estatística R/S modificada, o método gráfico descrito na Subseção 5.1.2 não seja acurado, ele nos fornece uma idéia do valor do parâmetro H.

5.1.4 *Estatística KPSS*

Introduzido por Kwiatkowski et al. (1992), o KPSS é um teste que, assim como a estatística R/S, baseia-se em somas parciais. Ele foi inicialmente proposto para testar a hipótese nula de estacionariedade contra a hipótese de existência de uma raíz unitária. Lee e Schmidt (1996) utilizam a estatística KPSS para testar a presença de longa dependência em séries temporais estacionárias.

Dada uma série temporal $\{X_t\}_{t=1}^n$, a estatística KPSS é dada por

$$T_n = \frac{1}{\hat{\sigma}_n^2(q)n^2}\sum_{k=1}^{n}\left(\sum_{j=1}^{k}(X_j - \bar{X})\right)^2,$$

onde $\hat{\sigma}_n^2(q)$ é definida pela expressão (5.3). Sob as hipóteses de estacionariedade e ausência de longa dependência

$$T_n \xRightarrow{d} \int_0^1 (W^0(t))^2 dt,$$

onde $\{W^0(t)\}_{t\in[0,1]}$ é uma ponte Browniana.

5.1.5 *Estatística V/S*

Giraitis et al. (2003) propõem uma modificação para a estatística KPSS. A estatística modificada é denominada *rescaled variance*, ou simplesmente, V/S. Tal estatística apresenta uma correção para a média e é mais sensível para mudanças na variância do que a estatística KPSS.

Dada uma série temporal $\{X_t\}_{t=1}^n$, a estatística V/S é dada por

$$M_n = \frac{1}{\hat{\sigma}_n^2(q)n^2}\left[\sum_{k=1}^{n}\left(\sum_{j=1}^{k}(X_j-\bar{X})\right)^2 - \frac{1}{n}\left(\sum_{k=1}^{n}\sum_{j=1}^{k}(X_j-\bar{X})\right)^2\right],$$

onde $\hat{\sigma}_n^2(q)$ é definida pela expressão (5.3). Sob as hipóteses de estacionariedade e ausência de longa dependência

$$M_n \overset{d}{\Longrightarrow} \int_0^1 (W^0(t))^2 dt - \left(\int_0^1 W^0(t)dt\right)^2,$$

onde $\{W^0(t)\}_{t\in[0,1]}$ é uma ponte Browniana.

Observação 5.1.

1. A estimação do coeficiente de Hurst, para os casos das estatísticas KPSS e V/S, segue os mesmos passos descritos na Subseção 5.1.2. Entretanto, no caso das estatísticas KPSS e V/S, a inclinação da reta obtida é $b > 1$.

2. O coeficiente de Hurst H satisfaz $H = d + 0.5$ e o coeficiente b satisfaz $b = 2H - 1 = 2d$.

3. Giraitis et al. (2003) sugerem utilizar $N = 40$ como tamanho mínimo amostral para cada bloco e $N < \frac{n}{K}$ como tamanho máximo. Além disso, para tamanho amostral $n \in \{3000, 5000\}$, os autores utilizam $\frac{n}{K} = 40$, em ambos os casos. Note que, da forma como os autores definem N_{min} e K, não é necessário utilizar pontos de truncamento pois não utilizam-se valores pequenos de N e, mesmo para valores grandes de N obtém-se várias amostras da estatística em questão.

5.2 *Método de Estimação da Máxima Verossimilhança* (MLE)

Nesta seção relembramos o método de estimação da máxima verossimilhança e definimos os estimadores de pseudo-máxima verossimilhança (PMLE) e quase-máxima verossimilhança (QMLE). Os estimadores de pseudo-máxima verossimilhança e quase-máxima verossimilhança são empregados quando a verdadeira distribuição dos dados não é conhecida. No que segue, dado um processo estocástico $\{X_t\}_{t\in\mathbb{Z}}$, denotamos por $\mathcal{F}_t$ a σ-álgebra gerada por $\{X_k\}_{k\leq t}$.

Seja X uma variável aleatória com função densidade de probabilidade $p_X(\cdot\,;\boldsymbol{\eta})$, onde $\boldsymbol{\eta}$ é o vetor de parâmetros desconhecidos que pertence ao espaço de parâmetros $\boldsymbol{\Theta} \subseteq \mathbb{R}^k$. Sabe-se que a *função de verossimilhança* de $\boldsymbol{\eta}$ é dada por

$$L(\boldsymbol{\eta}; x) := p_X(x; \boldsymbol{\eta}), \quad \text{para cada } x \in \mathbb{R},$$

e o *estimador de máxima verossimilhança* de $\boldsymbol{\eta}$ é, por definição,

$$\hat{\boldsymbol{\eta}}_{\text{MLE}} := \underset{\boldsymbol{\eta}\in\boldsymbol{\Theta}}{\operatorname{argmax}} \{L(\boldsymbol{\eta}; x)\}.$$

Note que, maximizar a função $L(\cdot; x)$ é equivalente a maximizar $\ln\big(L(\boldsymbol{\eta}; x)\big)$ pois $\ln(\cdot)$ é uma função estritamente crescente. Segue que, $\hat{\boldsymbol{\eta}}$ pode ser definido, equivalentemente, por

$$\hat{\boldsymbol{\eta}}_{\text{MLE}} := \underset{\boldsymbol{\eta}\in\boldsymbol{\Theta}}{\operatorname{argmax}} \big\{\ln\big(L(\boldsymbol{\eta}; x)\big)\big\}.$$

A função $\mathscr{L}(\cdot; x) := \ln\big(L(\cdot; x)\big)$ é denominada *função de log-verossimilhança*. Além disso, se $G(\cdot)$ é uma função mensurável de $\boldsymbol{\eta}$ então, $\hat{\upsilon} = G(\hat{\boldsymbol{\eta}})$ é o estimador de máxima verossimilhança de $\upsilon = G(\boldsymbol{\eta})$.

De forma análoga, dada uma sequência de variáveis aleatórias $\{X_t\}_{t=1}^n$, as funções de verossimilhança e log-verossimilhança de $\boldsymbol{\eta}$ são dadas, respectivamente, por

$$L(\boldsymbol{\eta};\boldsymbol{x}) := p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\eta}) \quad \text{e} \quad \mathscr{L}(\boldsymbol{\eta};\boldsymbol{x}) := \ln\big(L(\boldsymbol{x};\boldsymbol{\eta})\big)$$

onde $p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\eta})$ é a função densidade de probabilidade de $\boldsymbol{X} = (X_1, \cdots, X_n)'$ e $\boldsymbol{x} = (x_1, \cdots, x_n)'$.

Note que, se as variáveis aleatórias X_t, para todo $t = 1, \cdots, n$, são independentes, a função de log-verossimilhança pode ser reescrita como

$$\mathscr{L}(\boldsymbol{\eta};\boldsymbol{x}) = \sum_{t=1}^{n} \ln\big(L_t(x_t;\boldsymbol{\eta})\big), \text{ onde } L_t(x_t;\boldsymbol{\eta}) = p_{X_t}(x_t;\boldsymbol{\eta}), \text{ para todo } t = 1, \cdots, n.$$

O estimador de máxima verossimilhança é obtido por

$$\hat{\boldsymbol{\eta}}_{\text{MLE}} := \underset{\boldsymbol{\eta}\in\boldsymbol{\Theta}}{\operatorname{argmax}} \{L(\boldsymbol{\eta};\boldsymbol{x})\} \quad \text{ou, equivalentemente,} \quad \hat{\boldsymbol{\eta}}_{\text{MLE}} := \underset{\boldsymbol{\eta}\in\boldsymbol{\Theta}}{\operatorname{argmax}} \{\mathscr{L}(\boldsymbol{\eta};\boldsymbol{x})\}.$$

Observamos que, na maioria das aplicações, $p_{X_t}(\cdot;\boldsymbol{\eta})$, $p_{X_t|\mathcal{F}_{t-1}}(\cdot;\boldsymbol{\eta})$ ou $p_{\boldsymbol{X}}(\cdot;\boldsymbol{\eta})$ são desconhecidas, impossibilitando assim a obtenção do estimador de máxima verossimilhança. Nesse caso, um estimador para $\boldsymbol{\eta}$ é obtido considerando-se o método de *pseudo-máxima verossimilhança*, proposto por Besag (1975). O método da pseudo-máxima verossimilhança consiste em substituir nas equações acima a função densidade de probabilidade de X_t (ou $X_t|\mathcal{F}_{t-1}$), para cada $t \in \mathbb{Z}$, por uma função densidade de probabilidade qualquer. Sob esse cenário, as funções de verossimilhança e log-verossimilhança passam a ser denotadas *função de pseudo-verossimilhança* e *função de log-pseudo-verossimilhança*, respectivamente, e o estimador de máxima verossimilhança é denominado *estimador de pseudo-máxima verossimilhança*.

Observação 5.2. A função de pseudo-verossimilhança, construída sob a hipótese de normalidade da variável aleatória $X_t|\mathcal{F}_{t-1}$, é também conhecida como *função de quase-verossimilhança* e o estimador obtido a partir de tal função é denominado *estimador de quase-máxima verossimilhança*.

5.2.1 QMLE para os Processos SFIEGARCH

Seja $\{X_t\}_{t\in\mathbb{Z}}$ um processo estocástico SFIEGARCH, dado na Definição 4.1. Seja $\boldsymbol{\eta}$ o vetor de parâmetros a ser estimado, isto é,

$$\boldsymbol{\eta} = (d, \theta, \gamma, \omega, a_1, \cdots, a_p, b_1, \cdots, b_q)'.$$

Denote por $\mathcal{F}_t$ a σ-álgebra gerada por $\{Z_k\}_{k\leq t}$, para cada $t \in \mathbb{Z}$. Sob a hipótese de normalidade das variáveis aleatórias Z_t, para todo $t \in \mathbb{Z}$, e do fato que h_t é $\mathcal{F}_{t-1}$ mensurável, obtemos $X_t|\mathcal{F}_{t-1} \sim \mathcal{N}(0, h_t^2)$, para cada $t \in \mathbb{Z}$. Sendo assim, pela independência de $\{Z_t\}_{t\in\mathbb{Z}}$ (e, consequentemente, de $X_t|\mathcal{F}_{t-1}$, para todo $t \in \mathbb{Z}$), o estimador de quase-verossimilhança $\hat{\boldsymbol{\eta}}_{QMLE}$ de $\boldsymbol{\eta}$ é obtido maximizando-se a expressão

$$\mathscr{L}(\boldsymbol{\eta};\boldsymbol{x}) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\left[\ln(h_t^2) + \frac{x_t^2}{h_t^2}\right],$$

onde $\boldsymbol{x} = (x_1, \cdots, x_n)$ denota parte de uma realização do processo $\{X_t\}_{t\in\mathbb{Z}}$ e a sequência $\{h_t\}_{t=1}^n$ é obtida recursivamente, através da relação $h_1 = e^{0.5\omega}$, $z_1 = x_1 h_1^{-1}$,

$$h_t = \exp\left\{\frac{\omega}{2} + \frac{1}{2}\sum_{k=0}^{n-1}\lambda_{d,k}\Big[\theta z_{t-1-k} + \gamma\Big(|z_{t-1-k}| - \mu_{|Z|}\Big)\Big]\right\} \quad \text{e} \quad z_t = \frac{x_t}{h_t}, \quad \text{para todo } t > 1,$$

onde $\mu_{|Z|} = \sqrt{\frac{2}{\pi}}$ e $g(Z_t) = 0$ para $t \leq 0$.

5.2.2 PMLE para os Processos SFIEGARCH

De forma análoga ao caso da Seção 5.2.1, se assumirmos que as variáveis aleatórias Z_t, para todo $t \in \mathbb{Z}$, possuem distribuição t-Student padronizada com ν graus de liberdade, a função de pseudo-verossimilhança da variável aleatória $X_t|\mathcal{F}_{t-1}$ é dada por

$$\ell_t(\boldsymbol{\eta}; x_t) = \frac{1}{h_t} \frac{\Gamma\big((\nu+1)/2\big)}{\Gamma(\nu/2)\sqrt{(\nu-2)\pi}} \left(1 + \frac{x_t^2}{h_t^2(\nu-2)}\right)^{-\frac{\nu+1}{2}}, \quad \text{para todo} \quad t \in \mathbb{Z},$$

onde

$$\boldsymbol{\eta} = (\nu, d, \theta, \gamma, \omega, a_1, \cdots, a_p, b_1, \cdots, b_q)',$$

e o estimador $\hat{\boldsymbol{\eta}}_{PMLE}$ de $\boldsymbol{\eta}$ é obtido maximizando-se a expressão

$$\mathscr{L}(\boldsymbol{\eta}; \boldsymbol{x}) = \frac{1}{2}\sum_{t=1}^{n} \ln(h_t^2) + n \ln\left(\frac{\Gamma\big((\nu+1)/2\big)}{\Gamma(\nu/2)\sqrt{(\nu-2)\pi}}\right) - \sum_{t=1}^{n}\left[\frac{\nu+1}{2}\ln\left(1 + \frac{x_t^2}{h_t^2(\nu-2)}\right)\right],$$

onde $\boldsymbol{x} = (x_1, \cdots, x_n)'$ denota parte de uma realização do processo $\{X_t\}_{t\in\mathbb{Z}}$ e a sequência $\{h_t\}_{t=1}^n$ é obtida recursivamente analogamente ao QMLE, com $\mu_{|Z|} = \mathbb{E}(|Z_0|)$.

Observação 5.3. De forma análoga constrói-se o PMLE no caso da distribuição GED (veja Prass, Lopes e Achcar, 2013).

5.2.3 MLE para os Processos α-SFIEGARCH

Seja $\{X_t\}_{t\in\mathbb{Z}}$ um processo estocástico α-SFIEGARCH, dado na Definição 4.2. Sem perda de generalidade assumimos $\sigma = 1$. Seja $\boldsymbol{\eta}$ o vetor de parâmetros a ser estimado, isto é

$$\boldsymbol{\eta} = (\alpha, d, \theta, \gamma, \omega, a_1, \cdots, a_p, b_1, \cdots, b_q)'.$$

Por definição, Z_t são variáveis aleatórias i.i.d. $S\alpha S$, para todo $t \in \mathbb{Z}$. Do Teorema 3.6 segue que a função densidade de probabilidade de Z_t, para todo $t \in \mathbb{Z}$, é dada por

$$p_Z(x; \alpha, 0) = \frac{1}{\pi}\int_0^\infty \cos(xt) e^{-t^\alpha} dt, \quad \text{para todo } x \in \mathbb{R},$$

pois $h(x, t; \alpha, \beta) = xt$, para todo $\alpha \in (0, 2]$. Ou ainda, pela expressão (3.19) segue que a parametrização S_0 coincide com a parametrização canônica e assim, pelo Teorema 3.7, a função densidade de probabilidade de Z_t é dada, alternativamente, por $p_Z(0; \alpha, 0) = \frac{1}{\pi}\Gamma\left(1 + \frac{1}{\alpha}\right)$ e, para todo $x > 0$,

$$p_Z(-x; \alpha, 0) = p_Z(x; \alpha, 0) = \frac{\alpha x^{\frac{1}{\alpha-1}}}{\pi(\alpha-1)} \int_0^{\frac{\pi}{2}} V(t; \alpha, 0) \exp\Big\{ - x^{\frac{\alpha}{\alpha-1}} V(t; \alpha, 0)\Big\} dt,$$

onde $V(t; \alpha, 0) = \left(\frac{\cos(t)}{\sin(\alpha t)}\right)^{\frac{\alpha}{\alpha-1}} \frac{\cos(\alpha t - t)}{\cos(t)}$.

Note ainda que, dada $\mathcal{F}_{t-1}$, a σ-álgebra gerada pela informação obtida até o instante $t-1$, h_t é uma constante conhecida. Segue, pela Proposição 3.1, que a variável aleatória $X_t|\mathcal{F}_{t-1}$, para todo $t \in \mathbb{Z}$, tem distribuição $S_\alpha(0, h_t, 0)$ e sua função densidade de probabilidade é dada por

$$p_{X_t|\mathcal{F}_{t-1}}(x_t; \alpha, 0) = \frac{1}{h_t} p_Z(x_t h_t^{-1}; \alpha, 0), \quad \text{para todo } t \in \mathbb{Z} \text{ e } x_t \in \mathbb{R}.$$

Pela independência de $\{Z_t\}_{t\in\mathbb{Z}}$ segue que as variáveis aleatórias $X_t|\mathcal{F}_{t-1}$, para todo $t \in \mathbb{Z}$, são independentes. Portanto, dada uma amostra $\{x_t\}_{t=1}^n$ do processo $\{X_t\}_{t\in\mathbb{Z}}$, temos $L_t(x_t; \boldsymbol{\eta}) = p_{X_t|\mathcal{F}_{t-1}}(x_t; \alpha, 0)$, para todo $t = 1, \cdots, n$, e assim, a função de log-verossimilhança de $\boldsymbol{\eta}$ é dada por

$$\mathscr{L}(\boldsymbol{\eta}; \boldsymbol{x}) = \sum_{t=1}^{n} \ln\big(L_t(x_t; \boldsymbol{\eta})\big) = -\sum_{t=1}^{n} \ln(h_t) + \sum_{t=1}^{n} \ln\left(p_Z(x_t h_t^{-1}; \alpha, 0)\right),$$

onde a sequência $\{h_t\}_{t=1}^n$ é obtida recursivamente analogamente aos casos anteriores, com $\mu_{|Z|} = 2/\pi\Gamma(1-1/\alpha)$.

As propriedades do estimador de pseudo-máxima verossimilhança dependem tanto da distribuição verdadeira como da distribuição utilizada para construir o estimador. Resultados assintóticos para este estimador, no caso dos processos SFIEGARCH aqui estudados, ainda permanecem um problema em aberto.

Um estudo analisando as propriedades do estimador de quase-verossimilhança para amostras finitas é apresentado em Prass e Lopes (2012) (veja o Apêndice G). O estudo envolve a geração de amostras de processos SFIEGARCH$(0, d, 0)_s$ para os quais $\{Z_t\}_{t\in\mathbb{Z}}$ possui distribuição t-Student ou GED, com diferentes valores do parâmetro ν associado a cada uma das distribuições. Os resultados em Prass e Lopes (2012) indicam que o desempenho do método melhora com o aumento do tamanho amostral da série temporal observada. Além disso, como era esperado, para a distribuição t-Student, a estimação dos parâmetros é mais acurada para valores grandes de ν. E ainda, como era também esperado, para a distribuição GED o método de quase-verossimilhança apresenta melhores resultados quando ν está próximo de 2 (que corresponde ao caso Gaussiano).

5.3 *Inferência Bayesiana via MCMC*

Métodos Bayesianos têm sido utilizados com frequência na análise de dados financeiros, principalmente devido à complexidade da função de verossimilhança para modelos de volatilidade estocástica (veja, por exemplo, Meyer e Yu, 2000). A principal diferença do método Bayesiano para a análise clássica é que, no primerio caso, os parâmetros do modelo são tratados como variáveis aleatórias enquanto que, no segundo caso, são tratados como constantes reais (ou complexas, de acordo com o problema).

Nos artigos Prass, Lopes e Achcar (2013) (veja o Apêndice H) e Prass, Lopes e Crato (2013) (veja o Apêndice I) descrevemos a metodologia de estimação Bayesiana utilizando métodos de Monte Carlo via Cadeias de Markov (*Monte Carlo Markov Chain* ou MCMC) para gerar amostras das distribuições à posteriori dos parâmetros de modelos FIEGARCH, sob a hipótese de que $Z_0 \sim$ GED(ν) e α-SFIEGARCH, respectivamente. Obviamente, o mesmo método se aplica a qualquer outro modelo SFIEGARCH.

Prass, Lopes e Achcar (2013) apresenta um estudo das propriedades do método Bayesiano para amostras finitas considerando diferentes valores para o parâmetro ν associado à distribuição GED. O trabalho também analisa a sensibilidade do método em relação à escolha das funções densidade à priori e dos hiperparâmetros associados a elas. Prass, Lopes e Crato (2013) analisa o desempenho do método para amostras finitas na presença de longa dependência sazonal (ou não) na volatilidade. O artigo também apresenta uma aplicacão aos dados analizados em Lopes e Prass (2013a,b) a fim de comparar os resultados obtidos através dos métodos clássico e Bayesiano.

5.4 *Estimadores Baseados nas Funções Periodograma e Densidade Espectral*

Nesta seção apresentamos a aproximação para o MLE Gaussiano, proposta por Whittle (1953), e a correspondente versão do estimador para o caso discreto. Tais estimadores pertencem à classe dos estimadores paramétricos. Enquanto que a versão originalmente proposto por Whittle (1953) considera a matriz $A(\boldsymbol{\eta})$, cujas entradas são dadas pela expressão (5.5), na versão discreta tal matriz é substituída pela função periodograma. Além destes, encontramos ainda na literatura o estimador conhecido como *local-Whittle* (ou ainda, estimador Gaussiano semiparamétrico), inicialmente

sugerido por Künsch (1987) e mais tarde estudado por Robinson (1995). O local-Whittle pertence à classe semiparamétrica e assume que a função densidade espectral $f_X(\lambda;\boldsymbol{\eta})$ do processo $\{X_t\}_{t\in\mathbb{Z}}$ pode ser aproximada por $c\lambda^{1-2H}$ (taxa de decaimento desta função), em uma vizinhança da origem.

5.4.1 *MLE Gaussiano e a Aproximação de Whittle*

Sejam $\{X_t\}_{t\in\mathbb{Z}}$ um processo estocástico Gaussiano estacionário com função densidade espectral $f_X(\cdot;\boldsymbol{\eta}_0)$ e $\{X_t\}_{t=1}^n$ uma série temporal obtida a partir de $\{X_t\}_{t\in\mathbb{Z}}$. No que segue, $\boldsymbol{X}=(X_1,\cdots,X_n)'$, $\Sigma_n(\boldsymbol{\eta}_0)=[\gamma_X(j-l)]_{j,l=1,\cdots,n}$ é a matriz de covariâncias de $\boldsymbol{X}$, $|\Sigma_n|$ é o determinante de $\Sigma_n(\cdot)$ e $\boldsymbol{\eta}_0$ é o vetor de parâmetros associado ao modelo descrito.

Sabe-se (veja, por exemplo, Beran, 1994) que a função densidade de probabilidade conjunta de $\boldsymbol{X}$ é dada por

$$p(\boldsymbol{x};\boldsymbol{\eta}_0)=(2\pi)^{-\frac{n}{2}}|\Sigma_n(\boldsymbol{\eta}_0)|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\boldsymbol{x}'\Sigma_n(\boldsymbol{\eta}_0)^{-1}\boldsymbol{x}\right\},\ \text{para todo } \boldsymbol{x}=(x_1,\cdots,x_n)'\in\mathbb{R}^n,$$

e a função de log-verossimilhança é dada por

$$\mathscr{L}(p(\boldsymbol{x};\boldsymbol{\eta}))=-\frac{n}{2}\ln(2\pi)-\frac{1}{2}\ln(|\Sigma_n(\boldsymbol{\eta})|)-\frac{1}{2}\boldsymbol{x}'\Sigma_n(\boldsymbol{\eta})^{-1}\boldsymbol{x}. \tag{5.4}$$

Devido à complexidade do problema de maximização da expressão (5.4), Whittle (1953) propõe uma aproximação para essa função. Tal aproximação é conhecida na literatura como MLE *aproximado de Whittle*. A aproximação de Whittle baseia-se no fato que os únicos termos da expressão (5.4) que dependem de $\boldsymbol{\eta}$ são $\ln(|\Sigma_n(\boldsymbol{\eta})|)$ e $\boldsymbol{x}'\Sigma_n(\boldsymbol{\eta})^{-1}\boldsymbol{x}$. Na aproximação de Whittle (veja Beran, 1994, e referências ali contidas) temos,

(a) o termo $\ln(|\Sigma_n(\boldsymbol{\eta})|)$ é substituído por $\frac{n}{2\pi}\int_{-\pi}^{\pi}\ln\big(f_X(\lambda;\boldsymbol{\eta})\big)d\lambda$ pois

$$\lim_{n\to\infty}\frac{1}{n}\ln(|\Sigma_n(\boldsymbol{\eta})|)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln\big(f_X(\lambda;\boldsymbol{\eta})\big)d\lambda.$$

(b) quanto ao termo $\boldsymbol{x}'\Sigma_n(\boldsymbol{\eta})^{-1}\boldsymbol{x}$, a matriz $\Sigma_n(\boldsymbol{\eta})^{-1}$ é substituída por $A(\boldsymbol{\eta})=[a(j-l)]_{j,l=1,\cdots,n}$, onde

$$a(j-l)=\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\frac{1}{f_X(\lambda;\boldsymbol{\eta})}e^{\mathrm{i}(j-l)\lambda}d\lambda,\ \text{para todo } j,l=1,\cdots,n. \tag{5.5}$$

Note que o termo $\frac{n}{2}\ln(2\pi)$, na expressão (5.4), não depende de $\boldsymbol{\eta}$. Assim, Whittle (1953) propõe o MLE aproximado para $\boldsymbol{\eta}_0$, obtido através da minimização, com respeito a $\boldsymbol{\eta}$, da função

$$\mathscr{L}_W(p(\boldsymbol{x};\boldsymbol{\eta}))=\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln\big(f_X(\lambda;\boldsymbol{\eta})\big)d\lambda+\frac{1}{n}\boldsymbol{x}'A(\boldsymbol{\eta})^{-1}\boldsymbol{x}. \tag{5.6}$$

5.4.2 *Versão Discreta do Estimador de Whittle*

Embora de fácil implementação, o estimador aproximado de Whittle, apresentado na Seção 5.4.1, requer a construção da matriz $A(\boldsymbol{\eta})$ e, consequentemente, a avaliação das integrais dadas na expressão (5.5), para cada valor de $\boldsymbol{\eta}$, no momento da maximização. Tal tarefa pode tornar-se computacionalmente demorada para grandes tamanhos amostrais ou quando o parâmetro $\boldsymbol{\eta}$ possui dimensão elevada. Sendo assim, a construção descrita a seguir é proposta na literatura (veja, por exemplo, Beran, 1994).

Lembramos que, dada uma série temporal $\{X_t\}_{t=1}^n$, a função periodograma desta série temporal é definida como

$$I_n(\lambda) = \frac{1}{2\pi n}\Big|\sum_{t=1}^{n} X_t e^{-i\lambda t}\Big|^2, \text{ para todo } \lambda \in [0,\pi].$$

Logo,

$$\begin{aligned}\frac{1}{n}\boldsymbol{X}' A(\boldsymbol{\eta})^{-1}\boldsymbol{X} &= \frac{1}{n}\sum_{j=1}^{n}\sum_{l=1}^{n} X_j a(j-l) X_l = \frac{1}{n}\sum_{j=1}^{n}\sum_{l=1}^{n} X_j X_l \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\frac{1}{f_X(\lambda;\boldsymbol{\eta})} e^{\mathrm{i}(j-l)\lambda} d\lambda \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{f_X(\lambda;\boldsymbol{\eta})}\left(\frac{1}{2\pi n}\sum_{j=1}^{n}\sum_{l=1}^{n} X_j e^{\mathrm{i}(j-l)\lambda} X_l\right) d\lambda \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{I_n(\lambda)}{f_X(\lambda;\boldsymbol{\eta})} d\lambda.\end{aligned}$$

Portanto, a expressão (5.6) pode ser reescrita, em termos da função periodograma, como

$$\mathscr{L}_W(p(\boldsymbol{x};\boldsymbol{\eta})) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\ln\big(f_X(\lambda;\boldsymbol{\eta})\big) d\lambda + \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{I_n(\lambda)}{f_X(\lambda;\boldsymbol{\eta})} d\lambda,$$

cuja versão discreta é dada por

$$\begin{aligned}\mathscr{L}_W(p(\boldsymbol{x};\boldsymbol{\eta})) &= \frac{2}{n}\sum_{j=1}^{\left\lfloor\frac{n-1}{2}\right\rfloor}\ln\big(f_X(\lambda_j;\boldsymbol{\eta})\big) + \frac{2}{n}\sum_{j=1}^{\left\lfloor\frac{n-1}{2}\right\rfloor}\frac{I_n(\lambda_j)}{f_X(\lambda_j;\boldsymbol{\eta})} \\ &= \frac{1}{n}\sum_{j=1}^{n-1}\ln\big(f_X(\lambda_j;\boldsymbol{\eta})\big) + \frac{1}{n}\sum_{j=1}^{n-1}\frac{I_n(\lambda_j)}{f_X(\lambda_j;\boldsymbol{\eta})},\end{aligned} \tag{5.7}$$

onde $\lambda_j = \frac{2\pi j}{n}$, para $j = 1, \cdots, n-1$. Note que a segunda igualdade dada na expressão (5.7) segue do fato que as funções periodograma e densidade espectral são ambas simétricas em torno de π.

Zaffaroni (2009) mostra que, para os processos FIEGARCH cujas inovações apresentam variância finita, o estimador de Whittle é consistente e assintóticamente normal. Embora, intuitivamente, acredita-se que o mesmo resultado vale para os processos SFIEGARCH, a prova não é análoga à apresentada em Zaffaroni (2009). Mais especificamente, a prova apresentada em Zaffaroni (2009) assume que, para todo $\boldsymbol{\eta} \in \Theta$, os coeficientes $\{\lambda_{d,k}\}_{k\in\mathbb{N}}$ dados em (4.6) satisfazem $|\lambda_{d,k}| < K|\lambda_{d,j}|$, para algum $K > 0$ e todo $j, k \in \mathbb{N}$ com $1 \leq j \leq k$, onde $\boldsymbol{\eta} = (\boldsymbol{\zeta}, \boldsymbol{\Phi})'$ é qualquer vetor admissível correspondente à função densidade espectral; $\boldsymbol{\zeta} \in \Phi_1$ são os parâmetros do modelo utilizados para obter $\{\lambda_{d,k}\}_{k\in\mathbb{N}}$; $\boldsymbol{\Phi} \in \Phi_2$ são os demais parâmetros do modelo; $\Theta = \Phi_1 \times \Phi_2$, onde Φ_1 e Φ_2 são subespaços compactos, respectivamente, de $\mathbb{R}^{\mathfrak{p}}$ e $\mathbb{R}^{\mathfrak{q}}$, onde $\mathfrak{p}+\mathfrak{q}$ é a dimensão do vetor $\boldsymbol{\eta}$. Pela expressão (4.6) é imediato que essa condição não é satisfeita para os processos SFIEGARCH quando $p = q = 0$.

Observação 5.4. Zaffaroni (2009) ressalta os seguintes fatos que são igualmente válidos para processos SFIEGARCH:

1. O parâmetro ω não é identificável através da função de autocorrelação. Portanto, não pode ser estimado através do método de Whittle. Um estimador simples para esse parâmetro é considerar a média amostral da série temporal $\{\ln(X_t^2)\}_{t=1}^n$

$$\hat{\omega} = \frac{1}{n}\sum_{t=1}^{n}\ln(X_t^2), \tag{5.8}$$

que é um estimador $\sqrt{n}$-consistente para $\mathbb{E}\big(\ln(X_t^2)\big) = \omega + \mathbb{E}\big(\ln(Z_t^2)\big)$, sob a hipótese de que $\mathbb{E}\big(\big[\ln(Z_t^2)\big]^2\big) < \infty$. Um estimador $\sqrt{n}$-consistente para ω é obtido subtraindo-se da expressão (5.8) o estimador de Whittle de $\mathbb{E}\big(\ln(Z_t^2)\big)$.

2. Em alguns casos (por exemplo, quando as inovações possuem função de distribuição simétrica) o sinal de θ não é identificável através da função densidade espectral.

3. Se a função de distribuição das inovações for especificada, o vetor de parâmetros desconhecidos $\boldsymbol{\eta}$ é dado por $\boldsymbol{\eta} = (\boldsymbol{\zeta}, \boldsymbol{\phi})'$, onde $\boldsymbol{\zeta}$ é o vetor de parâmetros desconhecidos associados ao polinômio $\lambda(\cdot)$ e $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)'$ ou $\boldsymbol{\phi} = (\phi_1, \phi_2)'$, onde $\phi_1 = \theta$, $\phi_2 = \gamma$ e ϕ_3 é um parâmetro desconhecido relacionado à função de distribuição dos dados. Por exemplo, $\phi_3 = \nu$, no caso da função de distribuição t_ν.

4. A fim de evitar a especificação da função de distribuição para as inovações, considera-se como vetor de parâmetros desconhecidos $\boldsymbol{\eta} = (\boldsymbol{\zeta}, \boldsymbol{\phi})'$, onde $\boldsymbol{\zeta}$ é o vetor de parâmetros desconhecidos associados ao polinômio $\lambda(\cdot)$ e $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)'$, onde

$$\phi_1 = \mathbb{E}([g(Z_0)]^2), \quad \phi_2 = \mathrm{Cov}\big(g(Z_0), \ln(Z_0^2)\big) \quad \text{e} \quad \phi_3 = \mathrm{Var}(\ln(Z_0^2)).$$

Sob a hipótese adicional de que a distribuição das inovações é simétrica, θ (ou θ^2) e γ podem ser associados a ϕ_1 e ϕ_2 através das relações

$$\begin{aligned}
&\phi_1 = \mathrm{Var}\big(g(Z_0)\big) = \theta^2 + \gamma^2\big(1 - [\mathbb{E}(|Z_0|)]^2\big) \quad \text{e} \\
&\phi_2 = \mathrm{Cov}\big(g(Z_0), \ln(Z_0^2)\big) = \gamma\big[\mathbb{E}\big(|Z_0|\ln(Z_0^2)\big) - \mathbb{E}\big(|Z_0|\big)\mathbb{E}\big(\ln(Z_0^2)\big)\big] = \gamma\mathrm{Cov}\big(|Z_0|, \ln(Z_0^2)\big).
\end{aligned}$$

Neste caso, $\mathbb{E}(|Z_0|)$ e $\mathrm{Cov}\big(|Z_0|, \ln(Z_0^2)\big)$ devem ser fixados à priori. Assim, ϕ_3 é o parâmetro livre e $\mathrm{Var}(\ln(Z_0^2))$, $\mathbb{E}(|Z_0|)$ e $\mathrm{Cov}\big(|Z_0|, \ln(Z_0^2)\big)$ são conjuntamente determinadas.

5.4.3 *Estimador de Whittle Modificado para Processos com Variância Infinita*

Kokoszka e Taqqu (1999) apresentam uma versão modificada do estimador de Whittle, para processos ARFIMA com inovações α-estáveis simétricas. Os autores provam que o estimador é consistente e assintóticamente normal. No que segue, apresentamos a versão modificada do estimador de Whittle para os processos α-SFIEGARCH. Observamos que tal estimador não nos fornece uma estimativa para os parâmetros relacionados à função de distribuição dos dados. Tais valores devem ser estimados em uma segunda etapa. O mesmo é válido para o parâmetro ω.

No que segue, dado um processo $\{X_t\}_{t\in\mathbb{Z}}$ α-SFIEGARCH, isto é,

$$\begin{aligned}
X_t &= h_t Z_t \\
\ln(h_t^2) &= \omega + \frac{a(\mathcal{B})}{b(\mathcal{B})}(1-\mathcal{B}^s)^{-d} g(Z_{t-1}) = \omega + \sum_{k=0}^{\infty} \lambda_{d,k} g(Z_{t-1-k}), \text{ para todo } t \in \mathbb{Z},
\end{aligned}$$

onde Z_t são variáveis aleatórias i.i.d. $S\alpha S$, para todo $t \in \mathbb{Z}$, denotamos por $\{X_t^*\}_{t\in\mathbb{Z}}$ o processo SFIEGARCH Gaussiano associado à $\{X_t\}_{t\in\mathbb{Z}}$, isto é,

$$\begin{aligned}
X_t^* &= h_t \varepsilon_t \\
\ln(h_t^{*2}) &= \omega + \frac{a(\mathcal{B})}{b(\mathcal{B})}(1-\mathcal{B}^s)^{-d} g(\varepsilon_{t-1}) = \omega + \sum_{k=0}^{\infty} \lambda_{d,k} g(\varepsilon_{t-1-k}), \text{ para todo } t \in \mathbb{Z},
\end{aligned}$$

onde ε_t são variáveis aleatórias i.i.d. $\mathcal{N}(0, \sigma^2)$, para todo $t \in \mathbb{Z}$, com $\sigma > 0$.

Note que Z_t possui variância infinita, para todo $t \in \mathbb{Z}$. Portanto, a função de densidade espectral de $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ não está bem definida. Seja $f_{X^*}(\lambda_j; \boldsymbol{\eta})$ a função densidade espectral do processo $\{\ln(X_t^{*2})\}_{t\in\mathbb{Z}}$. Defina

$$\tilde{\mathscr{L}}_W(p(\boldsymbol{x}; \boldsymbol{\eta})) := \frac{1}{n}\sum_{j=1}^{n-1} \ln\big(f_{X^*}(\lambda_j; \boldsymbol{\eta})\big) + \frac{1}{n}\sum_{j=1}^{n-1} \frac{\tilde{I}_n(\lambda_j)}{f_{X^*}(\lambda_j; \boldsymbol{\eta})}, \tag{5.9}$$

onde $\tilde{I}_n(\lambda)$ é a função periodograma normalizada, dada por

$$\tilde{I}_n(\lambda) = \frac{1}{\sum_{j=1}^{n} \left(\ln(X_j^2)\right)^2} \left| \sum_{j=1}^{N} \ln(X_j^2) e^{-\mathrm{i}\lambda j} \right|^2, \quad \text{para todo } \lambda \in [0, \pi].$$

Um estimador do tipo Whittle para $\boldsymbol{\eta_0}$ é então dado por

$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta} \in \boldsymbol{\Theta}} \{\tilde{\mathscr{L}}_W(p(\boldsymbol{x}; \boldsymbol{\eta}))\}.$$

Observação 5.5. Observe que $\tilde{\mathscr{L}}_W(p(\boldsymbol{x}; \boldsymbol{\eta}))$ é igual a $\mathscr{L}_W(p(\boldsymbol{x}; \boldsymbol{\eta}))$, dada pela expressão (5.7), substituindo-se $I_n(\cdot)$ por sua versão normalizada $\tilde{I}_n(\lambda)$, e considerando a função densidade espectral do processo $\{\ln(X_t^*)\}_{t \in \mathbb{Z}}$.

Tanto o esimador de Whittle quanto sua versão modificada são de fácil implementação. Entretanto, no caso dos processos SFIEGARCH, maximizar a expressão (5.9) é, na prática, uma tarefa computacionalmente demorada devido à complexidade da expressão para a função densidade espectral. Ressaltamos ainda que, no caso dos processos SFIEGARCH, resultados assintóticos para o estimador descrito acima ainda são um problema em aberto. A prova apresentada em Kokoszka e Taqqu (1999) não pode ser adaptada para processos SFIEGARCH pois, ao contrário do que ocorre nos processos ARFIMA, a função densidade espectral do processo $\{\ln(X_t^*)\}_{t \in \mathbb{Z}}$ não pode ser decomposta como produto de duas funções $f_1(\cdot)$ e $f_2(\cdot)$, onde $f_1(\cdot)$ depende apenas dos parâmetros da função de distribuição e $f_2(\cdot)$ depende apenas dos demais parâmetros do modelo.

No Apêndice J apresentamos uma breve análise do comportamento do estimador descrito nesta seção para processos α-SFIEGARCH$(0, d, 0)_s$, quando $s = 1$. A simulação considera $\alpha \in \{1.25,\ 1.50,\ 1.75,\ 2.00\ \}$, $d \in \{0.1, 0.2, 0.3, 0.4\}$, $\omega = 0$ e $\theta = \gamma = 0.10$.

Capítulo 6

Previsão

Seja $\{X_t\}_{t\in\mathbb{Z}}$ um processo SFIEGARCH$(p,d,q)_s$, definido no Capítulo 4. No que segue, tratamos da previsão de X_{n+h}, X^r_{n+h}, $\ln(X^r_{n+h})$, h^r_{n+h} e $\ln(h^r_{n+h})$, para $h>0$, $n\in\mathbb{Z}$ e $r>0$ (a ser definido de acordo com as inovações do processo), dada a informação conhecida até o tempo n, para algum $n\in\mathbb{Z}$. Teoricamente, dado um $n\in\mathbb{Z}$ qualquer, a informação conhecida até o tempo n é representada por $\mathcal{F}_n=\sigma(\{Z_k\}_{k\leq n})$. Na prática, observamos apenas $\{X_t\}_{t=1}^n$ e assumimos um conjunto de condições iniciais para $\{Z_k\}_{k\leq 0}$, denotado por I_0, de forma que $\{Z_t\}_{t=1}^n$ possa ser obtido recursivamente utilizando-se as expressões dadas nas Definições 4.1 ou 4.2, dependendo se o processo tem variância finita ou infinita, respectivamente.

Dividimos nosso estudo em dois casos. Na Seção 6.1 apresentamos um resumo dos resultados apresentados em Lopes e Prass (2013b) sobre a previsão de X_{n+h}, X^2_{n+h}, $\ln(h^2_{n+h})$, h^2_{n+h} e $\ln(h^2_{n+h})$, quando $\{X_t\}_{t\in\mathbb{Z}}$ é um SFIEGARCH$(p,d,q)_s$, com $\mathbb{E}(Z_0^2)<\infty$ (veja a Seção 4.1). Na Seção 6.2 apresentamos resultados relacionados aos processos α-SFIEGARCH$(p,d,q)_s$ definidos na Seção 4.3. Observamos que todos os resultados apresentados na Seção 6.2 valem também para o caso em que $\mathbb{E}(Z_0^2)<\infty$ pois a medida utilizada para definir os preditores é a *medida de dispersão* $\mathrm{disp}(\cdot)$, definida na Seção 6.2. Por outro lado, nem todas as expressões derivadas em Lopes e Prass (2013b) valem no caso de variância infinita pois assumem a existência de momentos de ordem maior ou igual à 2, para X_t ou h_t, para todo $t\in\mathbb{Z}$.

6.1 Processos SFIEGARCH com Variância Finita

Seja $\{X_t\}_{t\in\mathbb{Z}}$ um processo SFIEGARCH$(p,d,q)_s$, definido na Seção 4.1. Utilizando-se os mesmos argumentos da prova do lema 1 em Lopes e Prass (2013a) mostra-se que X_t é um *martingale difference* com respeito a $\mathcal{F}_t=\sigma(\{Z_k\}_{k\leq t})$. Consequentemente, o melhor preditor possível (em termos da medida erro quadrático médio) para X_{n+h}, dado $\mathcal{F}_n$, coincide com o melhor preditor linear e é dado por $\mathbb{E}(X_{n+h}|\mathcal{F}_n)=0$, para todo $h>0$ e $n\in\mathbb{Z}$. Sendo assim, focamos nossa atenção nos preditores de X^2_{n+h}, $\ln(X^2_{n+h})$, h^2_{n+h} e $\ln(h^2_{n+h})$, para $h>0$ e $n\in\mathbb{Z}$.

Seja $\{Y_t\}_{t\in\mathbb{Z}}$ um processo estocástico qualquer. Lopes e Prass (2013a,b) definem a seguinte notação, adotada também nesta seção:

- utiliza-se "$\hat{}$" para denotar o preditor h passos à frente, definido em termos da espernça condicional, isto é, $\hat{Y}_{t+h}=\mathbb{E}(Y_{t+h}|\mathcal{F}_t)$;
- utiliza-se "$\tilde{}$" (por exemplo, $\tilde{Y}_{t+h}$) e "$\check{}$" (por exemplo, $\check{Y}_{t+h}$) para denotar preditores alternativos;
- $\hat{\ln}(Y_{t+h})$ representa o preditor h passos à frente da variável aleatória $\ln(Y_{t+h})$ (as notações "$\tilde{}$" e "$\check{}$" são utilizadas de forma análoga);

- seguindo o mesmo padrão adotado na literatura, o preditor h passos à frente de Y_{t+h}^2 é representado por $\hat{Y}_{t+h}^2$ ao invés de $\widehat{Y_{t+h}^2}$. Para evitar confusão, denotamos o quadrado de $\hat{Y}_{t+h}$ como $(\hat{Y}_{t+h})^2$ (o mesmo segue para "˜" e "ˇ").

No artigo Lopes e Prass (2013a) consideram-se processos FIEGARCH(p,d,q) e provam-se os resultados descritos a seguir. Utilizamos as palavras Proposição, Lema, Teorema e Corolário em *itálico* quando nos referimos aos correspondentes no artigo. Como mencionado em Lopes e Prass (2013b), tais resultados são também válidos para os processos SFIEGARCH$(p,d,q)_s$ e as provas são análogas às apresentadas em Lopes e Prass (2013a).

Seja $\{X_t\}_{t\in\mathbb{Z}}$ um processo FIEGARCH(p,d,q) estacionário. Então, para todo $n\in\mathbb{Z}$ e $h>0$,

- os preditores h passos à frente de X_{n+h} e X_{n+1}^2, dado $\mathcal{F}_n$, são, respectivamente (veja o *Lema* 2),
$$\hat{X}_{n+h}=0 \quad \text{e} \quad \hat{X}_{n+1}^2=h_{n+1}^2;$$
- os preditores de X_{n+h}^2 e h_{n+h}^2 coincidem, isto é, $\hat{X}_{n+h}^2=\hat{h}_{n+h}^2$ (veja o parágrafo após a prova do *Lema* 2);
- o preditor h passos à frente de $\ln(h_{n+h}^2)$, dado $\mathcal{F}_n$, é escrito como (veja a *Proposição* 4)
$$\hat{\ln}(h_{n+h}^2)=\omega+\sum_{k=h-1}^{\infty}\lambda_{d,k}g(Z_{n+h-1-k})=\omega+\sum_{k=0}^{\infty}\lambda_{d,k+h-1}g(Z_{n-k}),$$
e tem erro quadrático médio igual a 0, se $h=1$ e
$$\mathbb{E}\left(\left[\ln(h_{n+h}^2)-\hat{\ln}(h_{n+h}^2)\right]^2\right)=\sigma_g^2\sum_{k=0}^{h-2}\lambda_{d,k}^2, \quad \text{se } h\geq 2;$$
com σ_g^2 dada em (4.10);
- definindo-se $\check{h}_t^2:=\exp\{\hat{\ln}(h_t^2)\}$, $\check{X}_t^2:=\check{h}_t^2$ e tomando o preditor h passos à frente de $\ln(X_{n+h}^2)$, dado $\mathcal{F}_n$, como $\check{\ln}(X_{n+h}^2):=\ln(\check{X}_{n+h}^2)=\ln(\check{h}_{n+h}^2)=\hat{\ln}(h_t^2)$, obtém-se (veja a *Proposição* 5)
$$\mathbb{E}\left(\left[\ln(X_{n+h}^2)-\check{\ln}(X_{n+h}^2)\right]^2\right)=\sigma_g^2\sum_{k=0}^{h-2}\lambda_{d,k}^2+\mathbb{E}\left(\ln(Z_{n+h}^2)\right),$$
com σ_g^2 dada em (4.10).

Lopes e Prass (2013a) propõe ainda um preditor h passos à frente para h_{n+h}^2 derivado a partir de uma expansão de Taylor de ordem 2 para a função $\ln(\cdot)$. Tal preditor é dado por

$$\tilde{h}_{n+h}^2:=\exp\left\{\mathbb{E}(\ln(h_{n+h}^2)|\mathcal{F}_n)\right\}+\frac{1}{2}\mathbb{E}(\left[\ln(h_{n+h}^2)-\mathbb{E}(\ln(h_{n+h}^2)|\mathcal{F}_n)\right]^2)\exp\left\{\mathbb{E}(\ln(h_{n+h}^2)|\mathcal{F}_n)\right\} \tag{6.1}$$

e está associado à $\check{h}_{n+h}^2$ através da relação

$$\tilde{h}_{n+h}^2=\begin{cases}\exp\{\hat{\ln}(h_{n+h}^2)\}=\check{h}_{n+h}^2, & \text{se} \quad h=1;\\ \exp\{\hat{\ln}(h_{n+h}^2)\}\left(1+\frac{1}{2}\sigma_g^2\sum_{k=0}^{h-2}\lambda_{d,k}^2\right)=\check{h}_{n+h}^2\left(1+\frac{1}{2}\sigma_g^2\sum_{k=0}^{h-2}\lambda_{d,k}^2\right), & \text{se} \quad h>1.\end{cases} \tag{6.2}$$

É fácil ver que $\check{X}_{n+h}^2$ (de forma análoga, $\check{h}_{n+h}^2$), para todo $h>0$ e $n\in\mathbb{Z}$, não depende da existência do momento de ordem 2 de $\{X_t\}_{t\in\mathbb{Z}}$, podendo ser utilizado sempre que $\mathbb{E}(Z_0^2)<\infty$. Obviamente, o erro quadrático médio associado a esse preditor só está bem definido se $\mathbb{E}(X_t^4)\leq\infty$.

No artigo Lopes e Prass (2013b) derivamos a expressão exata para $\hat{h}_t^2$, sob a hipótese de que $\mathbb{E}(X_t^2) < \infty$ (veja o *Teorema* 4.1) e obtemos o erro quadrático médio do preditor, quando $\mathbb{E}(X_t^4) < \infty$. Observamos que, por exemplo, quando Z_0 possui distribuição t-Student padronizada a hipótese de que $\mathbb{E}(X_t^2) < \infty$ não é satisfeita se os parâmetros θ e γ não satisfazem $\lambda_{d,k}\gamma + |\lambda_{d,k}\theta| < 0$, para todo $k \in \mathbb{N}$ (veja o teorema A1.2 em Nelson, 1991). Em Lopes e Prass (2013b) mostra-se que os preditores $\check{h}_{n+h}^2$, $\tilde{h}_{n+h}^2$ e $\hat{h}_{n+h}^2$ estão associados através da relação $\hat{h}_{n+1}^2 = \check{h}_{n+1}^2 = \tilde{h}_{n+1}^2$ e

$$\hat{h}_{n+h}^2 = \check{h}_{n+h}^2 \prod_{\ell=0}^{h-2} \mathbb{E}\big(\exp\{\lambda_{d,\ell} g(Z_0)\}\big) = \tilde{h}_{n+h}^2 \left[1 + \frac{1}{2}\sigma_g^2 \sum_{k=0}^{h-2} \lambda_{d,k}^2\right]^{-1} \prod_{\ell=0}^{h-2} \mathbb{E}\big(\exp\{\lambda_{d,\ell} g(Z_0)\}\big),$$

para todo $n \in \mathbb{Z}$ e $h > 1$.

Ressaltamos que na aplicação empírica apresentada em Lopes e Prass (2013b) utilizam-se processos ARMA(p_1, q_1)-SFIEGARCH$(p_2, d, q_2)_s$. Sendo assim, no referido artigo, considera-se os processos $\{r_t\}_{t\in\mathbb{Z}}$ definidos através de (veja Lopes e Prass, 2013b, página 20)

$$r_t = \mu + \sum_{k=0}^{\infty} \psi_k X_{t-k} := \mu + \psi(\mathcal{B}) X_t, \quad \text{para todo } t \in \mathbb{Z}, \tag{6.3}$$

onde $\mu \in \mathbb{R}$, $\{\psi_k\}_{k\in\mathbb{Z}}$ é uma sequência real satisfazendo $\sum_{k=0}^{\infty} \psi_k^2 < \infty$ e $\{X_t\}_{t\in\mathbb{Z}}$ é um processo SFIEGARCH$(p, d, q)_s$ com $\sup_{t\in\mathbb{Z}}\{\mathrm{Var}(X_t^2)\} < \infty$. Em seguida, mostra-se que os processos ARMA(p_1, q_1)-SFIEGARCH$(p_2, d, q_2)_s$ satisfazem (6.3) e prova-se que (veja o *Teorema* 4.3)

$$\hat{r}_{n+h}^2 = \mu^2 + \sum_{k=0}^{h-1} \psi_k^2 \hat{h}_{n+h-k}^2 + \sum_{j=h}^{\infty}\sum_{\ell=h}^{\infty} \psi_j \psi_\ell X_{n+h-j} X_{n+h-\ell} + 2\mu \sum_{i=h}^{\infty} \psi_i X_{n+h-i}, \quad \text{para todo } h > 0,$$

onde $\hat{h}_{n+h}^2$ é a previsão h passos à frente, dado $\mathcal{F}_n$, para X_{n+h}^2 (equivalentemente, para h_{n+h}^2), para todo $h \geq 1$. Em particular, $\hat{h}_{n+1}^2 = h_{n+1}^2$, para todo $n \in \mathbb{N}$.

6.2 *Processos SFIEGARCH com Variância Infinita*

Nesta seção tratamos da previsão h passos à frente para os processos α-SFIEGARCH$(p, d, q)_s$, definidos na Seção 4.3. Lembramos que, no caso de processos com variância finita, os preditores h passos à frente para as variáveis de interesse são, em geral, definidos de forma a minimizar o erro quadrático médio. Em particular, para um processo Gaussiano $\{Y_t\}_{t\in\mathbb{Z}}$, o preditor $\hat{Y}_{t+h}$ da variável Y_{t+h} que minimiza o erro quadrático médio $\mathbb{E}([(Y_{t+h} - \hat{Y}_{t+h}]^2)$ também minimiza a probabilidade de grandes desvios, dada por $\mathbb{P}(|Y_{t+h} - \hat{Y}_{t+h}| > K)$, para todo $K > 0$ (veja Cline e Brockwell, 1985). Por outro lado, quando tratamos de processos com variância infinita, a medida erro quadrático médio deixa de fazer sentido. Sendo assim, é necessário definir um novo critério para a seleção do melhor preditor.

No caso dos processos lineares, é possível encontrar na literatura técnicas alternativas para a seleção do melhor preditor h passos à frente. Dentre elas, podemos citar a minimização do erro médio absoluto ou a técnica pseudo-espectral (veja Cline e Brockwell, 1985, e referências ali contidas). Entretanto, Cline e Brockwell (1985) observam que a maioria das técnicas propostas são muito complexas, de baixa aplicabilidade ou requerem o conhecimento da distribuição das inovações do processo. Sendo assim, Cline e Brockwell (1985) propõe o uso do critério denominado medida de dispersão, utilizado, por exemplo, por Blattberg e Sargent (1971) e Stuck (1978) em problemas envolvendo variáveis aleatórias estáveis. Ressaltamos que em Cline e Brockwell (1985) a medida de dispersão $\mathrm{disp}(\cdot)$ é utilizada em um contexto mais geral, isto é, assumindo-se apenas que o processo de inovação é uma sequência de variáveis aleatórias independentes e identicamente distribuídas (não

necessariamente estáveis) com um determinado decaimento para as caudas da distribuição (veja mais detalhes a seguir). Embora o foco de Cline e Brockwell (1985) sejam os processos ARMA(p,q), a medida $\mathrm{disp}(\cdot)$ pode também ser empregada para processo com longa dependência e variância infinita, como mostra Kokoszka (1996).

Para os processos não lineares com heteroscedasticidade condicional podemos citar Chen e You (2012) que trata dos processos denominados *stable power*-GARCH(p,q). Entretanto, enquanto Cline e Brockwell (1985) e Kokoszka (1996) discutem com rigor matemático as propriedades do preditores proposto, considerando o caso mais geral possível, Chen e You (2012) tratam apenas da previsão para o caso $p=q=1$. Além disso, ao contrário de Cline e Brockwell (1985) e Kokoszka (1996), em Chen e You (2012) não é apresentada nenhuma medida que justifique a escolha do preditor proposto. Os autores dizem apenas que aplicam a teoria geral de previsão para a volatilidade, isto é, definem o preditor h passos à frente através da esperança condicional do processo, de forma análoga ao caso de variância finita (nas palavras dos autores: *"In this paper we present a Power GARCH model with stable errors and apply the general theory of volatility forecasting to it."*).

No que segue, consideramos os processos α-SFIEGARCH$(p.d.q)_s$ e derivamos as condições necessárias para a existência do momento de ordem r das variáveis aleatórias X_t e h_t, para todo $t\in\mathbb{Z}$. Até o presente momento não foi possível derivar condições suficientes para que $\mathbb{E}(|X_t|^r)$ e $\mathbb{E}(h_t^r)$ sejam finitas para algum $0<r<\alpha$, exceto no caso $\alpha=2$ (veja Lopes e Prass, 2013b). A dificuldade para provar tal resultado segue do fato que (veja a prova do Teorema 6.2)

$$\mathbb{E}(h_t^r)=\exp\left\{\frac{r\omega}{2}\right\}\prod_{k=0}^{\infty}\mathbb{E}\left(\exp\left\{\frac{r}{2}\lambda_{d,k}g(Z_{-1-k})\right\}\right),\quad \text{para } 0<r<\alpha.$$

Para provar que $\mathbb{E}(h_t^r)$ existe, é necessário mostrar que o produto infinito nesta expressão converge para uma constante não nula. Para mostrar tal resultado é imprescindível que saibamos a expressão exata da integral dada em (6.6).

Por outro lado (veja o Teorema 6.3), o cálculo da esperança condicional (dado $\mathcal{F}_n$) de X_{n+h}^r e h_{n+h}^r, para $h>0$, $n\in\mathbb{Z}$ e $0<r<\alpha$ envolve apenas um número finito de termos do tipo $\mathbb{E}\left(\exp\left\{\frac{r}{2}\lambda_{d,\ell}g(Z_{t+h-1-\ell})\right\}\right)$. Sendo assim, satisfeitas as condições necessárias dadas no Teorema 6.3, os preditores obtidos através da esperança condicional para as variáveis aleatórias X_{n+h}^r e h_{n+h}^r, para $h>0$, $n\in\mathbb{Z}$ e $0<r<\alpha$, de fato estão bem definidos, no sentido que são variáveis aleatórias finitas com probabilidade 1. Alternativamente, definimos também a medida de dispersão $\mathrm{disp}(\cdot)$ e mostramos como utilizá-la para obter preditores h-passos à frente para os processos associados à um α-SFIEGARCH. Observamos que os preditores obtidos dessa maneira não dependem da existência do momento de order r das variáveis aleatórias X_t e h_t, para todo $t\in\mathbb{Z}$.

Na Proposição 6.1 e no Teorema 6.1, a seguir, tratamos, respectivamente, da existência de $\mathbb{E}(e^{\mathfrak{b}Z})$ e $\mathbb{E}(e^{\mathfrak{b}g(Z)})$, onde Z uma variável aleatória $S\alpha S$, com $\alpha\in(1,2]$, $g(\cdot)$ é definida em (4.5) e $\mathfrak{b}\in\mathbb{R},\mathfrak{b}\neq 0$. O resultado do Teorema 6.1 é fundamental para a prova do Teorema 6.2, que fornece condições necessárias para a existência do momento de ordem r das variáveis aleatórias X_t e h_t, para todo $t\in\mathbb{Z}$.

Proposição 6.1. *Seja Z uma variável aleatória $S\alpha S$, com $\alpha\in(1,2]$. Então, $\mathbb{E}(e^{\mathfrak{b}Z})=\infty$, para todo $\mathfrak{b}\in\mathbb{R},\mathfrak{b}\neq 0$.*

Prova: Pela expressão (3.20), quando $\alpha\in(1,2]$ e $\beta=0$ temos

$$p_Z(z;\alpha,0)=\frac{1}{\pi}\int_0^{\infty}\cos\left(zt\right)e^{-t^{\alpha}}dt,\quad \text{para todo } z\in\mathbb{R}.$$

Portanto,

$$\mathbb{E}(e^{bZ}) = \int_{-\infty}^{\infty} e^{\mathfrak{b}z}\left[\frac{1}{\pi}\int_0^{\infty}\cos(zt)e^{-t^{\alpha}}dt\right]dz = \frac{1}{\pi}\int_0^{\infty} e^{-t^{\alpha}}\left[\int_{-\infty}^{\infty}\cos(zt)e^{\mathfrak{b}z}dz\right]dt, \tag{6.4}$$

sendo que a última igualdade na expressão (6.4) segue pelo teorema de Fubini.

De Gradshteyn e Ryzhik (2000) (veja a expressão 2.663, ítem 3) temos

$$\int \cos(zt)e^{\delta z}dz = \frac{e^{\delta z}\left[\delta\cos(zt) + t\sin(zt)\right]}{\delta^2 + t^2}, \quad \text{para todo } \delta, t \in \mathbb{R}. \tag{6.5}$$

Observe que, no intervalo de integração $(-\infty, 0]$, a expressão (6.5) é finita se, e somente se, $\delta \geq 0$ e, em $[0, \infty]$, (6.5) é finita se, e somente se, $\delta \leq 0$. Portanto, conclui-se que $\mathbb{E}(e^{\mathfrak{b}Z}) = \infty$, para todo $\mathfrak{b} \in \mathbb{R}, \mathfrak{b} \neq 0$. ∎

Teorema 6.1. *Seja Z uma variável aleatória $S\alpha S$, com $\alpha \in (1, 2]$, $g(\cdot)$ a função definida em (4.5), com $|\theta| + |\gamma| > 0$, e $\mathfrak{b} \in \mathbb{R}, \mathfrak{b} \neq 0$. Então $\mathbb{E}(\exp\{\mathfrak{b}g(Z)\}) < \infty$ se, e somente se, $\mathfrak{b}\gamma + |\mathfrak{b}\theta| < 0$.*

Prova: Observe que $g(\cdot)$ pode ser reescrita como

$$g(Z_t) = \begin{cases} (\theta+\gamma)Z_t - \gamma\mathbb{E}(|Z_t|), & Z_t \geq 0, \\ (\theta-\gamma)Z_t - \gamma\mathbb{E}(|Z_t|), & Z_t < 0, \end{cases} \quad \text{para todo } t \in \mathbb{Z}.$$

Portanto, análogo à (6.4), temos

$$\begin{aligned}\mathbb{E}(e^{\mathfrak{b}g(Z_0)}) &= \frac{1}{\pi}\int_0^{\infty} e^{-t^{\alpha}}\left[\int_{-\infty}^{\infty}\cos(zt)e^{\mathfrak{b}g(z)}dz\right]dt \\ &= \frac{1}{\pi}e^{\gamma\mathbb{E}(|Z_0|)}\left(\int_0^{\infty} e^{-t^{\alpha}}\left[\int_0^{\infty} e^{\mathfrak{b}(\gamma-\theta)z}dt\right]dz + \int_0^{\infty} e^{-t^{\alpha}}\left[\int_0^{\infty} e^{\mathfrak{b}(\gamma+\theta)z}dt\right]dz\right).\end{aligned} \tag{6.6}$$

De (6.5) e (6.6) conclui-se que, para $\mathbb{E}(e^{\mathfrak{b}g(Z_0)})$ ser finita é necessário que $\mathfrak{b}(\gamma-\theta) < 0$ e $\mathfrak{b}(\gamma+\theta) < 0$, ou seja, $\mathfrak{b}\gamma + |\mathfrak{b}\theta| < 0$.

Assuma agora que $\mathfrak{b}\gamma + |\mathfrak{b}\theta| \leq 0$. Aplicando-se o resultado de (6.5) em (6.6), obtemos

$$\begin{aligned}\mathbb{E}(e^{\mathfrak{b}g(Z)}) &= \frac{1}{\pi}e^{\gamma\mathbb{E}(|Z_0|)}\left(\int_0^{\infty} e^{-t^{\alpha}}\left[\int_0^{\infty} e^{\mathfrak{b}(\gamma-\theta)z}dt\right]dz + \int_0^{\infty} e^{-t^{\alpha}}\left[\int_0^{\infty} e^{\mathfrak{b}(\gamma+\theta)z}dt\right]dz\right) \\ &= \frac{1}{\pi}e^{\gamma\mathbb{E}(|Z_0|)}\left(\int_0^{\infty}\frac{-\mathfrak{b}(\gamma-\theta)}{\mathfrak{b}^2(\gamma-\theta)^2+t^2}e^{-t^{\alpha}}dt + \int_0^{\infty}\frac{-\mathfrak{b}(\gamma+\theta)}{\mathfrak{b}^2(\gamma+\theta)^2+t^2}e^{-t^{\alpha}}dt\right) \\ &\leq \frac{1}{\pi}e^{\gamma\mathbb{E}(|Z_0|)}\left(\int_0^{1}\left[\frac{-\mathfrak{b}(\gamma-\theta)}{\mathfrak{b}^2(\gamma-\theta)^2+t^2} + \frac{-\mathfrak{b}(\gamma+\theta)}{\mathfrak{b}^2(\gamma+\theta)^2+t^2}\right]e^{-t^{\alpha}}dt + 2\,\mathfrak{M}\int_1^{\infty} e^{-t^{\alpha}}dt\right),\end{aligned}$$

onde $\mathfrak{M} = \max\{-\mathfrak{b}(\gamma-\theta), -\mathfrak{b}(\gamma+\theta)\}$.

Dado que (expressão 3.462, ítem 9*, em Gradshteyn e Ryzhik, 2000)

$$\int_0^{\infty} t^{\nu-1}e^{-\delta t^{\alpha}}dt = \frac{1}{\alpha}\delta^{-\frac{\nu}{\alpha}}\Gamma\left(\frac{\nu}{\alpha}\right), \quad \text{para todo } \nu > 0,\ \delta > 0 \text{ e } \alpha > 0,$$

conclui-se que $\int_1^{\infty} e^{-t^{\alpha}}dt < \infty$ (basta tomar $\nu = \delta = 1$ na equação acima) . Além disso, observando-se ainda que $-\mathfrak{b}(\gamma-\theta)[\mathfrak{b}^2(\gamma-\theta)^2+t^2]^{-1} - \mathfrak{b}(\gamma+\theta)[\mathfrak{b}^2(\gamma+\theta)^2+t^2]^{-1}$ é uma função limitada no interval fechado $[0, 1]$, conclui-se que $\mathbb{E}(e^{\mathfrak{b}g(Z)}) < \infty$. ∎

Teorema 6.2. *Seja $\{X_t\}_{t\in\mathbb{Z}}$ um processo α-SFIEGARCH$(p, d, q)_s$, dado na Definição 4.2, com $|\theta| + |\gamma| > 0$ e $d < 1 - 1/\alpha$. Então, se $\mathbb{E}(|X_t|^r) < \infty$ e $\mathbb{E}(h_t^r) < \infty$, para algum $0 < r < \alpha$, necessariamente, $\lambda_{d,k}\gamma + |\lambda_{d,k}\theta| < 0$, para todo $k \in \mathbb{N}$, onde $\{\lambda_{d,k}\}_{k\in\mathbb{N}}$ são os coeficientes do polinômio $\lambda(\cdot)$, definido em (4.6).*

Prova: Pela Proposição 3.2, $0 < r < \alpha$ implica $\mathbb{E}(|Z_t|^r) < \infty$. Da independência de h_t e Z_t conclui-se que, se $\mathbb{E}(|X_t|^r) = \mathbb{E}(h_t^r)\mathbb{E}(|Z_t|^r)$, para todo $t \in \mathbb{Z}$ e, pela estacionariedade estrita de $\{X_t\}_{t\in\mathbb{Z}}$, $\{h_t\}_{t\in\mathbb{Z}}$ e $\{Z_t\}_{t\in\mathbb{Z}}$, os valores das esperanças não dependem de $t \in \mathbb{Z}$. Além disso, é imediato que, $\mathbb{E}(|X_t|^r) = \mathbb{E}(|X_0|^r) < \infty$ se, e somente se, $\mathbb{E}(h_t^r) = \mathbb{E}(|h_0|^r) < \infty$. Sendo assim, é suficiente provar o resultado para $|h_0|^r$, $0 < r < \alpha$.

Das expressões (4.17), (4.6) e da independência de $\{Z_t\}_{t\in\mathbb{Z}}$, segue que, se o momento de ordem r da variável aleatória h_0 existe, ele pode ser escrito como

$$\mathbb{E}(|h_0|^r) = \exp\left\{\frac{r\omega}{2}\right\}\prod_{k=0}^{\infty}\mathbb{E}\left(\exp\left\{\frac{r}{2}\lambda_{d,k}g(Z_{-1-k})\right\}\right), \quad \text{para todo } 0 < r < \alpha. \tag{6.7}$$

Sendo assim, para que $\mathbb{E}(|h_0|^r)$ seja finita, é necessário que cada termo do produto acima seja finito. Portanto, dado que $r > 0$, o Teorema 6.1 implica que, necessariamente, $\lambda_{d,k}\gamma + |\lambda_{d,k}\theta| < 0$, para todo $k \in \mathbb{N}$. ∎

Para obter o preditor h passos à frente para $|X_{t+h}|^r$ observamos que, pelo Teorema 3.8,

$$\mathbb{E}(|Z_0|^r) = \frac{2^{r-1}}{r\int_0^\infty u^{-r-1}\sin^2(u)du}\Gamma\Big(1-\frac{r}{\alpha}\Big), \quad \text{para todo } 0 < r < \alpha.$$

Logo,

$$\mathbb{E}(|X_{n+h}|^r|\mathcal{F}_n) = \mathbb{E}(h_{n+h}^r|\mathcal{F}_n)\frac{2^{r-1}}{r\int_0^\infty u^{-r-1}\sin^2(u)du}\Gamma\Big(1-\frac{r}{\alpha}\Big), \quad \text{para todo} \quad h > 0.$$

Em particular, $\mathbb{E}(h_{n+1}^r|\mathcal{F}_n) = h_{n+1}^r$ e, portanto, o preditor 1 passo à frente para $|X_{n+1}|^r$, dado $\mathcal{F}_n$, é

$$\mathbb{E}(|X_{n+1}|^r|\mathcal{F}_n) = h_{n+1}^r\frac{2^{r-1}}{r\int_0^\infty u^{-r-1}\sin^2(u)du}\Gamma\Big(1-\frac{r}{\alpha}\Big).$$

O Teorema 6.3 abaixo fornece a expressão para $\mathbb{E}(h_{n+h}^r|\mathcal{F}_n)$, quando $h > 1$.

Observação 6.1. Análogo à Seção 6.1, quando consideramos a esperança condicional, o preditor h passos à frente de h_{n+h}^r é representado por $\hat{h}_{n+h}^r$ ao invés de $\widehat{h_{n+h}^r}$. Caso necessário, o valor de $\hat{h}_{n+h} := \mathbb{E}(h_{n+h}|\mathcal{F}_n)$ elevado na potência r, será denotado por $(\hat{h}_{n+h})^r$.

Teorema 6.3. *Seja $\{X_t\}_{t\in\mathbb{Z}}$ um processo α-SFIEGARCH$(p,d,q)_s$, com $\lambda_{d,k}\gamma + |\lambda_{d,k}\theta| < 0$, para todo $k \in \mathbb{N}$. Então, para todo $n \in \mathbb{Z}$, $\hat{h}_{n+h}^r := \mathbb{E}(h_{n+h}^r|\mathcal{F}_n)$ é dada por*

$$\hat{h}_{n+h}^r = exp\left\{\frac{\omega r}{2}\right\}\prod_{k=h-1}^{\infty}\exp\left\{\frac{r}{2}\lambda_{d,k}g(Z_{n+h-1-k})\right\}\prod_{\ell=0}^{h-2}\mathbb{E}\left(\exp\left\{\frac{r}{2}\lambda_{d,\ell}g(Z_0)\right\}\right), \quad \textit{para todo } h > 1. \tag{6.8}$$

Prova: Observe que a hipótese $\lambda_{d,k}\gamma + |\lambda_{d,k}\theta| < 0$, para todo $k \in \mathbb{N}$, assegura que, para todo $\ell \in \{0,\cdots,h-2\}$, $\mathbb{E}\left(\exp\left\{\frac{r}{2}\lambda_{d,\ell}g(Z_0)\right\}\right) < \infty$. O restante da prova segue utilizando-se os mesmos argumentos da prova do *Teorema* 4.1 em Lopes e Prass (2013b). ∎

Nos parágrafos anteriores apresentamos preditores h passos à frente, para $|X_{n+h}|^r$ e h_{n+h}^r, para todo $h > 0$ e $0 < r < \alpha$. Tais preditores foram obtidos considerando-se o método da esperança condicional. É fácil ver que, basta substituir $\mathbb{E}(|Z_t|^r)$ por $\mathbb{E}(Z_t^r)$ nas equações derivadas considerando-se $|X_{t+h}|^r$ e obtém-se um preditor para X_{t+h}^r, para todo $h > 0$ e $0 < r < \alpha$. Embora relativamente simples, tais expressões tornam-se pouco atrativas em aplicações práticas pois envolvem o cálculo de $h-1$ integrais para cada passo h desejado. Uma alternativa para esse problema seria substituir o valor $\mathbb{E}\left(\exp\left\{\frac{r}{2}\lambda_{d,\ell}g(Z_0)\right\}\right)$, para cada $\ell \in \{0,\cdots,h-2\}$, pelo respectivo estimador amostral (para mais detalhes, veja a *Seção* 5.2 em Lopes e Prass, 2013b).

No que segue definimos a medida de dispersão $\text{disp}(\cdot)$ e os preditores h passos à frente, para $|X_{n+h}|^r$ e h^r_{n+h}, para todo $h > 0$ e $0 < r < \alpha$, obtidos a partir dela. As expressões para tais preditores são mais simples e, portanto, mais atrativas para aplicações práticas.

Definição 6.1 (Medida de Dispersão). Seja $\{\xi_t\}_{t\in\mathbb{Z}}$ uma sequência de variáveis aleatórias i.i.d. cuja função de distribuição satisfaz

$$\lim_{t\to\infty} \frac{\mathbb{P}(|\xi_0| > xt)}{\mathbb{P}(|\xi_0| > t)} = x^{-\alpha}, \quad \text{para todo } x > 0 \text{ e algum } \alpha > 0.$$

Considere uma sequência real $\{c_k\}_{k\in\mathbb{Z}}$, com $\sum_{k=-\infty}^{\infty} |c_k|^\alpha < \infty$, para a qual a variável aleatória $Y = \sum_{k=-\infty}^{\infty} c_k\xi_k$ esteja bem definida. Então, a *dispersão* de Y, denotada $\text{disp}(Y)$, é dada por

$$\text{disp}(Y) = \sum_{k=-\infty}^{\infty} |c_k|^\alpha.$$

Observação 6.2. Ressaltamos que, em Kokoszka (1996), a medida de dispersão é definida sob a hipótese de que $\xi_t \sim S_\alpha(\beta, \sigma, \mu)$. Na Definição 6.1 assume-se apenas que $\{\xi_t\}_{t\in\mathbb{Z}}$ possui uma função de distribuição de variação lenta no infinito, com índice de cauda $\alpha > 0$. Essa definição é equivalente à adotada em Cline e Brockwell (1985).

Cline e Brockwell (1985) observam que, se $\{\xi_t\}_{t\in\mathbb{Z}}$ é uma sequência de variáveis aleatórias $S\alpha S$, com parâmetro de escala σ, então Y é uma variável aleatória estável simétrica, mais especificamente,

$$Y \overset{d}{=} \left[\sum_{k=-\infty}^{\infty} |c_k|^\alpha\right]^{\frac{1}{\alpha}} \xi_0, \quad \text{ou seja,} \quad Y \sim S_\alpha(0, \sigma_Y, 0), \text{ com } \sigma_Y := \sigma \left[\sum_{k=-\infty}^{\infty} |c_k|^\alpha\right]^{\frac{1}{\alpha}},$$

onde $\overset{d}{=}$ denota igualdade em distribuição. Nesse caso, diz-se que $\text{disp}(Y)$ é a dispersão de Y em relação à ξ_0.

No caso dos processos α-SFIEGARCH$(p, d, q)_s$, sabemos que (veja o *Teorema* 1 em Prass, Lopes e Crato, 2013)

$$\mathbb{P}(|g(Z_0)| > x) \sim C_{\alpha,\sigma} \left(|\gamma + \theta|^\alpha + |\theta - \gamma|^\alpha\right) x^{-\alpha}, \quad \text{quando } x \to \infty,$$

onde

$$C_{\alpha,\sigma} = \sigma^\alpha \left(\int_0^\infty x^{-\alpha} \sin(x) dx\right)^{-1} = \frac{\sigma^\alpha}{\Gamma(1-\alpha)\cos\left(\frac{\pi\alpha}{2}\right)}, \quad \text{para } \alpha \in (1, 2].$$

Além disso, para $(1 - d)\alpha > 1$, temos $\sum_{k=0}^{\infty} |\lambda_{d,k}|^\alpha < \infty$ (segue do *Teorema* 1.2 em Lopes e Prass, 2013b) e a série $\sum_{k=0}^{\infty} \lambda_{d,k} g(Z_{t-1-k})$ está bem definida e converge com probabilidade 1, para todo $t \in \mathbb{Z}$ (veja o *Teorema* 2 em Prass, Lopes e Crato, 2013). Segue que a medida de dispersão $\text{disp}(\cdot)$, dada na Definição 6.1, pode ser aplicada considerando-se o processo $\{\ln(h^r_t) - \frac{r}{2}\omega\}_{t\in\mathbb{Z}}$, para qualquer $r > 0$.

No Teorema 6.4 que segue apresentamos o melhor preditor linear h passos à frente para $\ln(h^r_{t+h})$, dado $\mathcal{F}_n$, em termos da medida de dispersão.

Teorema 6.4. *Seja $\{\ln(h^2_t)\}_{t\in\mathbb{Z}}$ o processo estocástico definido pela expressão (4.2), com $Z_0 \sim S\alpha S$, $\alpha \in (1, 2]$. Para qualquer $r > 0$ fixo, defina $Y_t := \ln(h^r_t) - \frac{r\omega}{2}$, para todo $t \in \mathbb{Z}$, Então, para todo $h > 0$ e $n \in \mathbb{N}$, existe uma única sequência $\{\mathfrak{c}_k(h)\}_{k\in\mathbb{Z}}$, tal que*

$$\text{disp}\left(Y_{n+h} - \sum_{k=0}^{\infty} \mathfrak{c}_k(h) Y_{n-k}\right) = \min_{\{u_k\}_{k\in\mathbb{Z}}} \left\{\text{disp}\left(Y_{n+h} - \sum_{k=0}^{\infty} u_k Y_{n-k}\right)\right\},$$

onde o mínimo é tomado sobre todas as sequências $\{u_k\}_{k\in\mathbb{Z}}$ *que satisfazem* $\sum_{k=0}^{\infty}|u_k|<\infty$, *se* $d>0$, *e* $\sum_{k=0}^{\infty}|u_k|^{\nu}<\infty$, *para algum* $1\leq\nu<\alpha$, *se* $d<0$. *A sequência* $\{\mathfrak{c}_k(h)\}_{k\in\mathbb{Z}}$ *é dada por*

$$\mathfrak{c}_k(h)=-\sum_{j=0}^{k-1}\lambda_{d,k}\tilde{\lambda}_{d,k+h-j},\quad \textit{para todo } j\in\mathbb{N},$$

onde $\{\lambda_{d,k}\}_{k\in\mathbb{Z}}$ *e* $\{\tilde{\lambda}_{d,k}\}_{k\in\mathbb{Z}}$ *são, respectivamente, os coeficientes dos polinômios* $\lambda(\cdot)$ *e* $\lambda^{-1}(\cdot)$, *para* $\lambda(\cdot)$ *definido em* (4.6). *Além disso,*

$$\operatorname{disp}\left(Y_{n+h}-\sum_{k=0}^{\infty}\mathfrak{c}_k(h)Y_{n-k}\right)=\left(\frac{r}{2}\right)^{\alpha}\sum_{j=0}^{k-1}|\lambda_{d,k}|^{\alpha}.$$

Prova: Com as devidas correspondências, os resultados seguem imitando-se a prova do teorema 3.1 em Kokoszka (1996). ∎

Pela prova do Teorema 6.3 (veja mais detalhes em Kokoszka, 1996), conclui-se que o melhor preditor linear (em relação à medida de dispersão) para $Y_{n+h}:=\ln(h_{n+h}^r)-\frac{r\omega}{2}$, denotado por $\mathrm{Y}_{r,n}^{\mathrm{DISP}}(h)$, pode ser escrito como

$$\mathrm{Y}_{r,n}^{\mathrm{DISP}}(h)=\lambda^{-1}(\mathcal{B})\lambda^{(h)}(\mathcal{B})Y_n=\frac{r}{2}\lambda^{(h)}(\mathcal{B})g(Z_{n-1}),\quad \text{para cada } h>0, \tag{6.9}$$

onde $\lambda^{(h)}(\cdot)$ é o polinômio definido por

$$\lambda^{(h)}(z)=\sum_{k=0}^{\infty}\lambda_{d,k+h}z^k,\quad \text{para cada } h>0,$$

onde $\{\lambda_{d,k}\}_{k\in\mathbb{Z}}$ são os coeficientes do polinômio $\lambda(\cdot)$, definido em (4.6).

Observação 6.3. A expressão (6.9) mostra que, para o caso de variância finita, o preditor de mínima dispersão coincide com o melhor preditor linear em termos de erro quadrático médio. Além disso, a variável aleatória $\mathrm{Y}_{r,n}^{\mathrm{DISP}}(h)$ é o melhor preditor para Y_{n+h} em termos da norma $\mathcal{L}^r$, para todo $1<r<\alpha$ (veja Kokoszka, 1996, e referências ali contidas).

Define-se então um preditor h passos à frente para h_{n+h}^r, denotado por $\mathrm{H}_{r,n}^{\mathrm{DISP}}(h)$, como sendo

$$\mathrm{H}_{r,n}^{\mathrm{DISP}}(h):=\exp\left\{\mathrm{Y}_{r,n}^{\mathrm{DISP}}(h)+\frac{r\omega}{2}\right\},\quad \text{para todo } h>0,$$

onde $\mathrm{Y}_{r,n}^{\mathrm{DISP}}(h)$ é o preditor de mínima dispersão para $Y_{n+h}:=\ln(h_{n+h}^r)-\frac{r\omega}{2}$. Além disso, análogo ao caso da esperança condicional, definimos o preditor h passos à frente para $|X_{n+h}|^r$, denotado por $|\mathrm{X}|_{r,n}^{\mathrm{DISP}}(h)$, como sendo

$$|\mathrm{X}|_{r,n}^{\mathrm{DISP}}(h):=\mathrm{H}_{r,n}^{\mathrm{DISP}}(h)\mathbb{E}(|Z_0|^r)=\exp\left\{\mathrm{Y}_{r,n}^{\mathrm{DISP}}(h)+\frac{r\omega}{2}\right\}\frac{2^{r-1}}{r\int_0^{\infty}u^{-r-1}\sin^2(u)du}\Gamma\left(1-\frac{r}{\alpha}\right),$$

para todo $h>0$. Note que, para $h=1$, esse preditor coincide com $\mathbb{E}(|X_{n+1}|^r|\mathcal{F}_n)$. Observamos que as propriedades dos preditores para h_{n+h}^r e $|X_{n+h}|^r$, definidos por meio da medida de dispersão $\operatorname{disp}(\cdot)$, ainda são um assunto em aberto.

Capítulo 7

Conclusões e Futuros Trabalhos

Neste trabalho apresentamos uma nova classe de modelos pertencente à família ARCH (Engle, 1982). Os modelos introduzidos, denominados FIEGARCH *com sazonalidade* e denotados por SFIEGARCH generalizam os processos FIEGARCH introduzidos por Bollerslev e Mikkelsen (1996). Definimos e estudamos as propriedades teóricas dos modelos SFIEGARCH em dois contextos diferentes. Primeiramente, consideramos os processos para os quais as inovações possuem variância finita. Em um segundo momento, estendemos a definição do processo SFIEGARCH para inovações com distribuição α-estável, quando $1 < \alpha < 2$, portanto, com variância infinita. Nesse caso, denotamos os processos por α-SFIEGARCH. Os processos SFIEGARCH foram abordados no artigo Lopes e Prass (2013b), enquanto que os α-SFIEGARCH foram discutidos em Prass, Lopes e Crato (2013).

Derivamos as condições necessárias e suficientes para que os processos SFIEGARCH estejam bem definidos. Tratamos da invertibilidade, estacionariedade (fraca e estrita), ergodicidade e representação espectral destes processos. Além disso, derivamos a expressão exata para o preditor h passos à frente para a variável aleatória h^2_{n+h} (e, consequentemente, para X^2_{n+h}), dado $\mathcal{F}_n$, onde h_{n+h} é a volatilidade estocástica associada ao processo SFIEGARCH, para todo $n \in \mathbb{Z}$ e $h > 0$. Uma aplicação de tais modelos à série dos log-retornos intradiários do índice S&P500 é apresentada em Lopes e Prass (2013b). No caso α-SFIEGARCH, derivamos as condições necessárias e suficientes para a existência, invertibilidade, estacionariedade estrita e ergodicidade dos processos. Além disso, derivamos a expressão para preditores h passos à frente para h^r_{n+h} e X^r_{n+h}, dado $\mathcal{F}_n$, para $r > 0$, levando-se em conta a esperança condicional ou a medida $\mathrm{disp}(\cdot)$, definida no Capítulo 6.

Descrevemos também os principais métodos para estimação, tanto do parâmetro de longa dependência d, quanto dos demais parâmetros dos modelos SFIEGARCH. Os métodos heurísticos descritos englobam a estatística R/S, KPSS e V/S. Outros métodos considerados incluem os estimadores da máxima verossimilhança (MLE), quase-máxima verossimilhança (QMLE) e pseudo-máxima verossimilhança (PMLE). As propriedades assintóticas desses estimadores, para os processos SFIEGARCH permanecem um assunto em aberto, a ser abordado em um futuro trabalho.

Apresentamos também o estimador de Whittle e sua versão modificada, que pode ser aplicada para processos com distribuição α-estável. Ressaltamos que Zaffaroni (2009) mostra que, para os processos FIEGARCH cujas inovações apresentam variância finita, o estimador de Whittle é consistente e assintóticamente normal. Embora, intuitivamente, acredita-se que o mesmo resultado valha para os processos SFIEGARCH, a prova não é análoga à apresentada em Zaffaroni (2009). Tal assunto será abordado posteriormente.

Ressaltamos ainda que, no caso dos processos α-SFIEGARCH, resultados assintóticos para o estimador de Whittle modificado são um assunto em aberto. A prova apresentada em Kokoszka e Taqqu (1999) não pode ser adaptada para processos SFIEGARCH pois, ao contrário do que ocorre nos processsos ARFIMA, a densidade espectral do processo $\{\ln(X^*_t)\}_{t\in\mathbb{Z}}$ não pode ser decomposta como produto de duas funções $f_1(\cdot)$ e $f_2(\cdot)$, onde $f_1(\cdot)$ depende apenas dos parâmetros da função

de distribuição e $f_2(\cdot)$ depende apenas dos demais coeficientes do modelo. Além disso, simulações indicam que, no caso de variância infinita, assumindo-se $\sigma = 1$, o estimador correspondente não apresenta um desempenho aceitável. Observamos que fixando-se $\sigma = 1$, no caso Gaussiano ($\alpha = 2$), o modelo não está corretamente especificado pois $Z_0 \sim S_\alpha(0, \sigma, 0)$ implica $Z_0 \sim \mathcal{N}(0, 2\sigma^2)$, para qualquer $\sigma > 0$. Por esse motivo os resultados da simulação não nos surpreendem. Espera-se que, assumindo σ desconhecido e estimando-se esse valor juntamente com os demais parâmetros, a estimação dos parâmetros do modelo FIEGARCH melhore. Abordaremos esse problema em outra ocasião.

Tratamos ainda da metodologia Bayesiana para a estimação dos parâmetros do modelo. O procedimento adotado é descrito em Prass, Lopes e Achcar (2013) e Prass, Lopes e Crato (2013). Como proposta para um futuro trabalho, pretendemos modificar o conjunto de condições iniciais I_0, atualmente assumido como sendo $I_0 := \{g(Z_k) < 0; k \leq 0\}$, a fim de verificar sua influência no algoritmo utilizado.

Outro tema abordado foi a previsão para os processos SFIEGARCH e α-SFIEGARCH. Os resultados para os processos SFIEGARCH foram apresentados nos artigos Lopes e Prass (2013a,b). Na Seção 6.2 consideramos os processos α-SFIEGARCH$(p, d, q)_s$. Dentre os resultados obtidos, destacamos a obtenção de condições necessárias para a existência do momento de ordem r das variáveis aleatórias X_t e h_t, para todo $t \in \mathbb{Z}$ (dadas na definição dos processos α-SFIEGARCH); a obtenção dos preditores h passos à frente, através do método da esperança condicional; a definição da medida de dispersão $\mathrm{disp}(\cdot)$ e a utilização de tal medida para obter preditores h-passos à frente para os processos associados à um α-SFIEGARCH. Até o presente momento não foi possível derivar condições suficientes para que $\mathbb{E}(|X_t|^r)$ e $\mathbb{E}(h_t^r)$ sejam finitas para algum $0 < r < \alpha$, exceto no caso $\alpha = 2$ (veja Lopes e Prass, 2013b). A dificuldade na prova de tal resultado segue do fato que (veja a prova do Teorema 6.2)

$$\mathbb{E}(h_t^r) = \exp\left\{\frac{r\omega}{2}\right\} \prod_{k=0}^{\infty} \mathbb{E}\left(\exp\left\{\frac{r}{2}\lambda_{d,k} g(Z_{t-1-k})\right\}\right), \quad \text{para } 0 < r < \alpha.$$

Assim, para provar que $\mathbb{E}(h_t^r)$ existe, é necessário mostrar que o produto infinito na expressão acima converge para uma constante não nula. Para mostrar tal resultado é imprescindível que saibamos a expressão exata da integral dada em (6.6). Outro assunto ainda em aberto são as propriedades dos preditores para h_{n+h}^r e $|X_{n+h}|^r$, definidos por meio da medida de dispersão $\mathrm{disp}(\cdot)$.

Além dos problemas em aberto citados nos parágrafos anteriores, pretende-se também analisar o comportamento dos preditores h passos à frente, quando os parâmetros do modelo são estimados utilizando-se diferentes métodos. Por exemplo, desejamos saber se a previsão obtida utilizando-se o MLE é mais acurada do que a obtida pelo PMLE, quando a verdadeira distribuição não foi corretamente especificada.

Referências Bibliográficas

Abramowitz, M. e I.A. Stegun (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Mineola: Dover Publications.

Annis, A.A. e E.H. Lloyd (1976). "The Expected Value of the Adjusted Rescaled Hurst Range of Independent Normal Summands." *Biometrika*, vol. 63, 111-116.

Baillie, R.; T. Bollerslev e H.O. Mikkelsen (1996). "Fractionally Integrated Generalised Autoregressive Conditional Heteroskedasticity". *Journal of Econometrics*, vol. 74, 3-30.

Belov, I.A. (2005). "On the Computation of the Probability Density Function of α-Stable Distributions". Em: *Mathematical Modeling and Analysis. Proceedings of the 10th International Conference MMA 2005*. Technika, 333-341.

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

Bergström, H. (1952). "On Some Expansions of Stable Distributions". *Arkiv for Mathematik*, vol. 2, 375-378.

Besag, J. (1975). "Statistical Analysis of Non-lattice Data". *The Statistician*, vol. 24, 179-195.

Bisognin C. (2007), "Estimação e Previsão em Processos SARFIMA$(p,d,q)\times(P,D,Q)_s$ na Presença de Outliers". Tese de Doutorado no Programa de Pós Graduação em Matemática. Universidade Federal do Rio Grande do Sul, Porto Alegre.

Bisognin, C. e S.R.C. Lopes (2007). " Estimating and Forecasting the Long-Memory Parameter in the Presence of Periodicity". *Journal of Forecasting*, vol. 26, 405-427.

Bisognin, C. e S.R.C. Lopes (2009). "Properties of Seasonal Long Memory Processes". *Mathematical and Computer Modelling*, vol. 49, 1837-1851.

Blattberg, R. e T. Sargent (1971). "Regression with non-Gaussian Stable Disturbances: Some Sampling Results". *Econometrica*, vol. 39, 501-510.

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

Bollerslev, T. (2010). "Glossary to ARCH (GARCH)". Em Bollerslev T.; J.R. Russell e M.W. Watson (editores), *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, Oxford: Oxford University Press, 137-163.

Bollerslev, T. e H.O. Mikkelsen (1996). "Modeling and Pricing Long Memory in Stock Market Volatility". *Journal of Econometrics*, vol. 73, 151-184.

Bordignon, S.; M. Caporin e F. Lisi (2007). "Generalized Long-Memory GARCH for Intradaily Volatility Modelling". *Computational Statistics & Data Analysis*, vol. 51, 5900-5912.

Bordignon, S.; M. Caporin e F. Lisi (2009). "Periodic Long-Memory GARCH models". *Econometric Reviews*, vol. 28(1), 60-82.

Breidt, F.J, N. Crato e P. de Lima (1998). "The Detection and Estimation of Long Memory in Stochastic Volatility". *Journal of Econometrics*, vol. 83(1-2), 325-348.

Brockwell, P.J. e R.A. Davis (1991). *Time Series: Theory and Methods*. New York: Springer-Verlag.

Chambers, J.M.; C.L. Mallows e B.W. Stuck (1976). "A Method for Simulating Stable Random Variables", *Journal of the American Statistical Association*, vol. 71, 340-344.

Chen, H. e C. You (2012). "Forecasting of Time Sereis with Stable Errors". Em: *Proceedings of 2012 International Conference on Measurement, Information and Control*, 298-302.

Cline, D.B.H. e P.J. Brockwell (1985). "Linear Prediction of ARMA Processes with Infinite Variance". *Stochastic Processes and their Applications*, vol. 19, 281-296.

Crato, N. (1996). "Some Results on the Spectral Analysis of Nonstationary Time Series". *Portugaliae Mathematica*, vol. 53(2), 179-186.

Ding, Z.; C.W.J. Granger e R.F. Engle (1993). "A Long Memory Property of Stock Market Returns and a New Model". *Journal of Empirical Finance*, vol. 1, 83-106.

Engle, R.F. (1982). "Autoregressive Conditional Heteroskedasticity with Estimates of Variance of U.K. Inflation". *Econometrica*, vol. 50, 87-1008.

Engle, R.F e T. Bollerslev (1986). "Modeling the Persistence of Conditional Variances", *Econometric Reviews*, vol. 5, 1-50.

Feller, W. (1951). "The Asymptotic Distributions of the Range of Sums of Independent Random Variables". *Annals of Mathematical Statistics*, vol. 22, 427-432.

Feller, W. (1971). *An Introdution to Probability Theory and its Applications*, vol. 2. New York: Wiley. 2ª ed.

Giraitis, L.; P. Kokoszka; R. Leipus e G. Teyssière (2003). "Rescaled Variance and Related Tests for Long Memory in Volatility and Levels". *Journal of Econometrics*, vol. 112, 265-294.

Giraitis, L. e R. Leipus (1995). "A Generalized Fractionally Differencing Approach in Long-Memory Modeling". *Lithuanian Mathematical Journal*, vol. 35(1), 53 - 65.

Gnedenko, B.V. e A.N. Kolmogorov (1954). *Limit Distribution for Sums of Independent Random Variables*. Reading: Addison-Wesley.

Gradshteyn, I.S. e I.M. Ryzhik (2000). *Table of Integrals, Series, and Products*. New York: Academic Press.

Granger, C.W.J. e R. Joyeux (1980). "An introduction to Long Memory Time Series Models and Fractional Differencing". *Journal of Time Series Analysis*, vol. 1(1), 15-9.

Hosking, J.R.M. (1981). "Fractional Differencing". *Biometrika*, vol. 68, 165-167.

Hosking, J.R.M. (1984), "Modelling Persistence in Hydrological Time Series Using Fractional Differencing". *Water Resources Research*, vol. 20(12), 1898 - 1908.

Hurst, H.E. (1951). "Long-term Storage in Reservoirs". *Transactions of the American Society of Civil Engineers*, vol. 116, 770-799.

Hurst, H.E. (1957). "A Suggested Statistical Model of Some Time Series which Occur in Nature". *Nature*, vol. 180, 494-494.

Kennedy, D. (1976). "The Distribution of the Maximum Brownian Excursion". *Journal of Applied Probabability*, vol. 13, 371-376.

Karlin, S. e H.M. Taylor (1975). *A First Course in Stochastic Processes.* New York: Academic Press. 2ª ed.

Kokoszka, P.S. (1996). "Prediction on Infinite Variance Fractional ARIMA". *Probability and Mathematical Statistics*, vol. 16, 65-83.

Kokoszka, P.S. e M.S. Taqqu (1995). "Fractional ARIMA with stable Innovations". *Stochastic Processes and their Applications*, vol. 60, 19-47.

Kokoszka, P.S. e M.S. Taqqu (1999). "Discrete Time Parametric Models with Long Memory and Infinite Variance". *Mathematical and Computer Modelling*, vol. 29, 203-215.

Kozubowski, T.J.; A.K. Panorska e S.T. Rachev (2003). "Statistical Issues in Modeling Multivariate Stable Portfolios". In Rachev, S.T. (editor), *Handbook of Heavy Tails Distributions in Finance.* Amsterdam: Elsevier, 131-162.

Künsch, H. R. (1987). "Statistical Aspects of Self-similar Processes". Em: *Proceedings of the First World Congress of the Bernoulli Society*, Y. Prokhorov & V. V. Sazanov (Editores). Utrecht: VNU Science Press, 67-74.

Kwiatkowski, D.; P.C.B. Phillips; P. Schmidt e Y. Shin (1992). "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root?" *Journal of Econometrics*, vol. 54, 159-178.

Lee, D. e P. Schmidt (1996). "On the Power of the KPSS Test of Stationarity Against Fractionally-integrated Alternatives". *Journal of Econometrics*, vol. 73(1), 285-302.

Lo, A.W. (1991). "Long Term Memory in Stochastic Market Prices". *Econometrica*, vol. 59, 1279-1313.

Lopes, S.R.C (2008). "Long-Range Dependence in Mean and Volatility: Models, Estimation and Forecasting". Em *In and Out of Equilibrium 2*, V. Sidoravicius e M.E. Vares (Editores), vol. 60, 497-525. Basel: Birkhäuser.

Lopes, S.R.C. e T.S. Prass (2013a). "Fractionally Integrated Exponential Generalized Autoregressive Conditional Heteroskedastic Processes". Em Revisão.

Lopes, S.R.C. e T.S. Prass (2013b). "Seasonal FIEGARCH Processes". *Computational Statistics & Data Analysis*, vol. 68, 262-295.

Mandelbrot, B. (1965). "Une Classe de Processus Stochastiques Homothetiques a soi; Application a loi Climatologique de H.E. Hurst". *Comptes Rendus de l'Académie des Sciences, Paris*, vol. 240, 3274-3277.

Mandelbrot B.B and J. Van Ness (1968). "Fractional Brownian Motion, Fractional Noises and Applications". *SIAM Review*, vol. 10, 422-437.

Mandelbrot, B.B e J.R. Wallis (1968). "Noah, Joseph and Operational Hydrology". *Water Resources Research*, vol. 4, 909-918.

Mandelbrot, B.B. e J.R. Wallis (1969). "Computer Experiments with Fractional Gaussian Noises. Part One, Averages and Variances". *Water Resources Research*, vol. 5(1), 228-241.

McLeod, A.I. e K.W. Hipel (1978). "Preservation of the Rescaled Adjusted Range. A Reassessment of the Hurst Phenomenon". *Water Resources Research*, vol. 14, 491-508.

Meyer, R. e T. Yu (2000). "Bugs for a Bayesian Analysis of Stochastic Volatility Models". *Econometrics Journal*, vol. 3, 198-215.

Mittnik, S.; T. Doganoglu e D. Chenyao (1999). "Computing the Probability Density Function of the Stable Paretian Distribution". *Mathematical and Computer Modelling*, vol. 29, 235-240.

Montanari, A.; R. Rosso e M.S. Taqqu (1997). "Fractionally Differenced ARIMA Models Applied to Hydrologic Time Series: Identification, Estimation, and Simulation." *Water Resources Research*, vol. 33(5), 1035-1044.

Montanari, A.; R. Rosso e M.S. Taqqu (2000). "A Seasonal Fractional ARIMA Model Applied to the Nile River Monthly Flows at Aswan. *Water Resources Research*, vol. 36(5), 1249–1259.

Nelson, D.B. (1991). "Conditional Heteroskedasticity in Asset Returns: A New Approach". *Econometrica*, vol. 59, 347-370.

Newey, W.K. e K.D. West, (1987). "A Simple Positive Semidefinite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". *Econometrica*, vol. 55, 703-708.

Nolan, J.P. (1997). "Numerical Calculation of Stable Densities and Distribution Functions". *Communication in Statistics - Stochastic Models*, vol. 13, 759-774.

Nolan, J.P. (1999). "An Algorithm for Evaluating Stable Densities in Zolotarev's (M) Parametrization". *Mathematical and Computer Modelling*, vol. 29, 229-233.

Panorska, A.K.; S. Mittnik e A. Rachev (1995). "Stable GARCH Models for Financial Time Series". *Applied Mathematics Letters*, vol. 8(5), 33-37.

Porter-Hudak, S. (1990). "An Application of the Seasonal Fractionally Differenced Model to the Monetary Aggregates". *Journal of the American Statistical Association*, vol. 85(410), 338-344.

Prass, T.S. (2008). *Análise e Estimação de Medidas de Risco em Processos* FIEGARCH. Dissertação de Mestrado no Programa de Pós-Graduação em Matemática. Universidade Federal do Rio Grande do Sul, Porto Alegre.

Prass, T.S.; J-M. Bravo; R.T. Clarke; W. Collischonn e S.R.C. Lopes (2012). "Comparison of Forecasts of Mean Monthly Water-Level in the Paraguay River, Brazil, from two Fractionally-Differenced Models". *Water Resources Research*, vol. 48, W05502.

Prass, T.S.; R.T. Clarke; W. Collischonn e S.R.C. Lopes (2011). "Joint Deterministic and Stochastic Modelling of Discharge in Large Brazilian Rivers". Em: *ISI 2011. Atas do International Statistical Institute.*

Prass, T.S. e S.R.C. Lopes (2012). "Quasi-likelihood Estimation on SFIEGARCH Process: a Simulation Study". Em *Proceedings of the Joint Statistical Meetings. Alexandria, VA: American Statistical Association*, vol. 1, 1313-1323.

Prass, T.S. e S.R.C. Lopes (2013). "VaR, Teste de Estresse e MaxLoss na Presença de Heteroscedasticidade e Longa Dependência na Volatilidade". *Revista Brasileira de Estatística*. No Prelo.

Prass, T.S; S.R.C. Lopes e J.A. Achcar (2013). "MCMC Bayesian Estimation in FIEGARCH Models". Em Revisão.

Prass, T.S.; S.R.C. Lopes e N. Crato (2013). "SFIEGARCH Processes with α-Stable Innovations: a Bayesian Approach for Parameter Estimation". Em Fase de Preparação.

Prass, T.S.; S.R.C. Lopes; J.G. Dorea; R.C. Marques e K.G. Brandão (2012). "Amazon Forest Fires Between 2001 and 2006 and Birth Weight in Porto Velho". *Bulletin of Environmental Contamination and Toxicology*, vol. 89(1), 1-7.

Press, W.H.; B.P. Flannery; S.A. Teukolsky e W.T. Vetterling (1988). *Numerical Recipes in C*. Cambridge: Cambridge University Press. 2ª ed.

Robinson, P.M. (1995). "Gaussian Semiparametric Estimation of Long Range Dependence". *The Annals of Statistics*, vol. 23, 1630-1661.

Roussas, G.G. (2005). *Measure-Theoretic Probability*. San Diego: Elsevier.

Samorodnitsky, G. e M.S. Taqqu (1994). *Stable Non-Gaussian Random Processes. Stochastic Models with Infinite Variance*. New York: Chapman & Hall.

Stoer, J. e R. Bulirsch (2002). *Introduction to Numerical Analysis*. New York: Springer. 3ª ed.

Stout, W.F. (1974). *Almost Sure Convergence*. New York: Academic Press.

Stuck, B.W. (1978). "Minimum Error Dispersion Linear Filtering of Scalar Symmetric Stable Processes". *IEEE Transactions on Automatic Control*, vol. 23, 507-509.

Taqqu, M.S.; V. Teverovsky e W. Willinger (1995). "Estimators for Long-Range Dependence: an Empirical Study". *Fractals*, vol. 3(4), 785-788.

Teverovsky, V.; M.S. Taqqu e W. Willinger (1999). "A Critical Look at Lo's Modiffied R/S Statistic". *Journal of Statistical Planning and Inference*, vol. 80, 211-227.

Tucker, H.G. (1967). *A Graduate Course in Probability*. New York: Academic Press.

Wei, W. (1990). *Time Series Analysis. Univariate and Multivariate Methods.* New York: Addison-Wesley.

Whittle, P. (1953). "Estimation and Information in Stationary Time Series". *Arkiv för Mathematik*, vol. 2(5), 423-434.

Woodward, W.A.; Q.C. Cheng e H.L. Gray (1998). "A k-Factor GARMA Long-Memory Model". *Journal of Time Series Analysis*, vol. 19(4), 485-504.

Zaffaroni, P. (2009). "Whittle Estimation of EGARCH and Other Exponential Volatility Models". *Journal of Econometrics*, vol. 151(2), 190-200.

Zivot, E. e J. Wang (2005). *Modeling Financial Time Series with S-PLUS*. New York: Springer-Verlag. 2ª ed.

Zolotarev, V.M. (1986). *One Dimensional Stable Distribution*, vol. 65 de "Translations of Mathematical Monographs". *American Mathematical Society*. Traduzido a partir da versão original de 1983 escrita em russo.

Apêndice A

Quadratura Numérica

Nesse apêndice apresentamos o método da quadratura numérica para a aproximação de integrais definidas. Tratamos em especial das regras de quadratura Gaussiana e de Laguerre, utilizadas no Capítulo 3 (veja Seção 3.3.3).

Os métodos de integração numérica são muito utilizados na prática. Tais métodos permitem calcular o valor aproximado de uma integral definida, mesmo sem conhecer uma expressão analítica para a sua primitiva. A idéia básica é utilizar a aproximação para a integral da função $f(\cdot)$, dada por

$$\int_a^b f(x)dx \approx \sum_{i=0}^{n} c_i f(x_i), \tag{A.1}$$

onde c_i, para $i = 0, \cdots, n$, são coeficientes reais e x_i, para $i = 0, \cdots, n$, são pontos do intervalo $[a, b]$. A aproximação dada na expressão (A.1) é denominada *quadratura numérica*.

No que segue apresentamos os métodos (ou regras) de quadratura Gaussiana e de Laguerre.

A.1 Quadratura Gaussiana

Uma regra de quadratura é uma aproximação da integral de uma função estabelecida como um somatório com pesos dos valores assumidos pela função em pontos específicos dentro do domínio de integração. A regra de quadratura Gaussiana é um caso particular dentre as regras de quadratura e recebeu esse nome em homenagem a Carl Friedrich Gauss (1777 - 1855).

Uma quadratura Gaussiana de n pontos, para $n \in \mathbb{N}$, é uma regra de quadratura construída para produzir um resultado exato para polinômios de grau menor ou igual a $2n - 1$, dada uma escolha adequada dos pontos x_i e pesos w_i para $i = 1, \cdots, n$. Por convenção, o domínio de integração de tal regra é o intervalo $[-1, 1]$, isto é,

$$\int_{-1}^{1} f(x)\, dx \approx \sum_{i=1}^{n} w_i f(x_i).$$

Press et al. (1988) ou, mais recentemente, Stoer e Bulirsch (2002), mostram que os pontos usados para avaliar a função são exatamente as raízes de um polinômio pertencente a uma classe de polinômios ortogonais. Para a quadratura Gaussiana de n pontos o polinômio associado é o n-ésimo polinômio de Legendre. Tais polinômios foram assim nomeados em homenagem a Adrien-Marie Legendre (1752 - 1833) e são definidos, de forma recorrente, por

$$P_0(x) = 1, \quad P_1(x) = x \quad \text{e} \quad (n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x), \quad \text{para todo} \quad n \geq 1.$$

Além disso, os pontos $\{x_i\}_{i=1}^n$ são as raízes do polinômio $P_n(\cdot)$ e os pesos $\{w_i\}_{i=1}^n$ são dados por (veja Abramowitz e Stegun, 1972)

$$w_i = \frac{2}{(1-x_i^2)\,(P_n'(x_i))^2},$$

onde $P_n'(\cdot)$ denota a derivada da função $P_n(\cdot)$.

A.1.1 Mudança de Intervalo para a Quadratura Gaussiana

Se o intervalo de integração for da forma $[a, b]$, uma mudança de variáveis deve ser feita antes de aplicar a regra da quadratura Gaussiana. Essa mudança é dada por

$$\int_a^b f(x)\,dx = \frac{b-a}{2}\int_{-1}^1 f\left(\frac{b-a}{2}x + \frac{a+b}{2}\right)\,dx.$$

Dessa forma, pela regra da quadratura Gaussiana, obtém-se

$$\int_a^b f(x)\,dx \approx \frac{b-a}{2}\sum_{i=1}^n w_i f\left(\frac{b-a}{2}x_i + \frac{a+b}{2}\right).$$

A.2 Outras Formas de Quadratura Gaussiana

O problema de integração analisado anteriormente pode ser expresso de uma maneira mais geral. Isto é, acrescenta-se uma função peso, $\omega(\cdot)$, ao integrando e permite-se intervalos de integração diferentes de $[-1, 1]$. Dessa forma, o problema passa a ser a estimação da seguinte integral

$$\int_a^b \omega(x)\,f(x)\,dx,$$

para alguma função $\omega(\cdot)$ e um dado intervalo $[a, b]$. Note que, se $a = -1$, $b = 1$ e $\omega(x) = 1$, o problema é o mesmo tratado na Subseção A.1.

Diferentes escolhas para a função peso e para o intervalo de integração nos fornecem diferentes regras de quadratura. Na Tabela A.1 apresentamos as principais regras de quadratura utilizadas na literatura e os respectivos intervalos de integração, função peso e polinômios ortogonais associados. Para mais detalhes, veja Abramowitz e Stegun (1972).

A.2.1 Quadratura de Gauss-Laguerre

Os polinômios de Laguerre, assim nomeados em homenagem a Edmond Laguerre (1834 - 1886), são utilizados em uma das regras de quadratura apresentadas na Tabela A.1. Tais polinômios podem ser definidos de forma recursiva, por

$$L_0(x) = 1, \quad L_1(x) = 1 - x \quad \text{e} \quad L_{n+1}(x) = \frac{(2n+1-x)L_n(x) - nL_{n-1}(x)}{(n+1)}, \text{ para todo } n \geq 1. \quad \text{(A.2)}$$

A quadratura de Gauss-Laguerre, ou simplesmente quadratura de Laguerre, é uma extensão do método da quadratura Gaussiana, para aproximar o valor de uma integral da forma

$$\int_0^\infty e^{-x} f(x)\,dx. \quad \text{(A.3)}$$

Tabela A.1: Regras de Quadratura e seus Respectivos Intervalos de Integração, Funções Peso e Polinômios Ortogonais Associados.

Intervalo	$\omega(x)$	Polinômios Ortogonais	Quadratura
$[-1,1]$	1	de Legendre	Gaussiana
$(-1,1)$	$(1-x)^{\alpha}(1+x)^{\beta}$ $\alpha,\beta>-1$	de Jacobi	-
$(-1,1)$	$\frac{1}{\sqrt{1-x^2}}$	de Chebyshev (primeria classe)	Chebyshev-Gauss
$[-1,1]$	$\sqrt{1-x^2}$	de Chebyshev (segunda classe)	Chebyshev-Gauss
$[0,\infty)$	e^{-x}	de Laguerre	Gauss-Laguerre
$(-\infty,\infty)$	e^{-x^2}	de Hermite	Gauss-Hermite

Nesse caso, a aproximação para a integral da expressão (A.3), é dada por

$$\int_0^{+\infty} e^{-x} f(x)\, dx \approx \sum_{i=1}^{n} w_i f(x_i),$$

onde $\{x_i\}_{i=1}^n$ são as raízes do polinômio de Laguerre $L_n(\cdot)$, de ordem n, e os pesos $\{w_i\}_{i=1}^n$ são dados por (veja Abramowitz e Stegun, 1972)

$$w_i = \frac{x_i}{(n+1)^2[L_{n+1}(x_i)]^2},$$

onde $L_{n+1}(\cdot)$ é definido pela expressão (A.2).

Note que, mesmo que a função peso não apareça no integrando, podemos reescrever o integrando de forma a obter uma integral como a que aparece em (A.3). Por exemplo, no caso $\int_0^{\infty} f(x)dx$, temos

$$\int_0^{\infty} f\,(x)\, dx = \int_0^{\infty} e^{-x} e^{x} f\,(x)\, dx = \int_0^{\infty} e^{-x} g\,(x)\, dx,$$

onde $g(x) = e^{x} f\,(x)$. Aplica-se então a regra da quadratura de Laguerre à integral $\int_0^{\infty} e^{-x} g(x) dx$.

Apêndice B

Métodos de Integração: Estudo Comparativo

Nessa seção apresentamos um estudo comparativo sobre a eficiência dos diferentes métodos para o cálculo da função densidade de probabilidade de variáveis α-estáveis, descritos na Seção 3.3.

Quando $\alpha = 2$ (fixando-se $\beta = 0$), $S_\alpha(\beta, \sigma, \mu)$ é a função de distribuição Gaussiana com média μ e variância $2\sigma^2$, denotada por $\mathcal{N}(\mu, 2\sigma^2)$, cuja função densidade de probabilidade é dada por

$$p(x) = \frac{1}{\sqrt{4\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{4\sigma^2}\right\}, \quad -\infty < x < \infty. \tag{B.1}$$

Para todos os cálculos apresentados nessa seção fixamos $\sigma = 1$ e $\mu = 0$. Utilizaremos a notação que segue para a função densidade de probabilidade:

- $p(\cdot)$ para os valores obtidos utilizando-se a expressão (B.1).
- $p_{zol}(\cdot)$ para os valores obtidos utilizando-se a representação integral de Zolotarev (1986), dada no Teorema 3.7, e o método de integração utilizado por Nolan (1997).
- $p_\Delta(\cdot)$ para os valores obtidos utilizando-se a aproximação dada pela expressão (3.33).
- $p_{fft}(\cdot)$ para os valores obtidos utilizando-se o método da FFT, descrito na Subseção 3.3.2.
- $p_{quad}(\cdot)$ para os valores obtidos utilizando-se o método das duas quadraturas, descrito na Subseção 3.3.3.

Primeiramente, tomamos $\alpha = 2$ e comparamos os valores da função densidade de probabilidade estimados através dos diferentes métodos citados acima com os valores obtidos através da expressão (B.1). Para $1 < \alpha < 2$ não temos uma forma fechada da função densidade de probabilidade. Então, comparamos os valores estimados utilizando-se a FFT e o método das duas quadraturas com os valores estimados utilizando-se integração numérica direta. Esse procedimento é análogo ao adotado por Mittnik et al. (1999) e Belov (2005).

As estatísticas utilizadas para avaliar a precisão dos métodos são o desvio médio absoluto D_1 e o desvio máximo absoluto D_2 definidos, respectivamente, por

$$D_1 = \frac{1}{N}\sum_{i=1}^{N} |v_i - w_i| \quad \text{e} \quad D_2 = \max_{i=1,\cdots,N} |v_i - w_i|, \tag{B.2}$$

onde v e w são vetores de tamanho N que representam os valores da função estimados para dois diferentes métodos, para N diferentes valores de x.

No que segue, os valores da função densidade de probabilidade foram calculados para $N = 50$ valores de x_k no intervalo $[0, 5)$, definidos por $x_k = (0.1)i$, para $k = 0, \cdots, 49$. Além disso,

- utilizamos $\varepsilon = 10^{-14}$ para calcular (veja Seção 3.3.1)

$$p_{\Delta}(x_k) \approx \frac{1}{\pi}\int_0^{\Delta} \cos\left(x_k t + \beta \tan\left(\frac{\pi\alpha}{2}\right)(t - t^{\alpha})\right) e^{-t^{\alpha}} dt,$$

  onde $\Delta = \Delta(\alpha, \beta)$ é a raiz da equação $\Gamma\left(\frac{1}{\alpha}, \Delta^{\alpha}\right) = \alpha\pi\varepsilon$, com $\Gamma(\cdot,\cdot)$ definida em (2.2);

- fixamos $N = 2^{14}$ e $h = 0.01$ para calcular (veja Seção 3.3.2)

$$p_{\mathit{fft}}(x_k) \approx s(-1)^{k-1-\frac{N}{2}} \sum_{n=1}^{N} (-1)^{n-1} \varphi\big(2\pi s\big[n - 1 - N/2\big]\big) e^{\frac{-\mathrm{i}2\pi(n-1)(k-1)}{N}}, \quad \text{para } k = 1, \cdots, N.$$

  Quando necessário, utilizamos interpolação para obter $p_{\mathit{fft}}(x_k)$;

- utilizamos $T = 8$ (Tabela B.1) e $T \in \{8, 16, 24\}$ (Tabela B.2), para calcular (veja Seção 3.3.3)

$$p_{quad}(x_k) = \int_0^T f(x_k, t; \alpha, \beta) dt + \int_0^{\infty} f(x_k, u + T; \alpha, \beta) du,$$

  com $f(x, t; \alpha, \beta) = \frac{1}{\pi}\cos\big(h(x, t; \alpha, \beta)\big) e^{-t^{\alpha}}$ e $h(x, t; \alpha, \beta)$, dada na Observação (3.9).

Na Tabela B.1 apresentamos os valores das estatísticas D_1 e D_2, dadas em (B.2), para $\alpha = 2$ e $\beta = 0$. Utilizamos como base de comparação a função $p(\cdot)$ dada pela expressão (B.1). Observamos que o método de integração numérica que utiliza as equações de Zolotarev (1986) (valores em negrito) é o que apresenta a menor diferença em relação aos valores obtidos pela expressão (B.1). Observa-se pouca diferença entre os desempenhos dos outros três métodos, quando comparados entre si.

Tabela B.1: Valores das medidas de diferença D_1 e D_2 para $v = \big(p(x_0), \cdots, p(x_{49})\big)$, $\alpha = 2$, $\beta = 0$ e $w_L := \big(p_L(x_0), \cdots, p_L(x_{49})\big)$, para cada $L \in \{zol, \Delta, \mathit{fft}, quad\}$.

w	w_{zol}	w_{Δ}	w_{fft}	w_{quad}
D_1	**1.002020**e-17	1.203828e-14	2.991660e-14	4.252059e-14
D_2	**1.110223**e-16	6.766809e-14	1.282585e-13	4.357625e-14

Dado que, dentre os métodos de integração numérica direta, aquele que utiliza as equações de **?** foi o que apresentou melhores resultados, utilizamos esse método como base para comparação quando $1 < \alpha < 2$.

Na Tabela B.2 apresentamos os valores das estatísticas D_1 e D_2, dadas em (B.2), para $\alpha \in \{1.25, 1.5, 1.75\}$ e $\beta \in \{-1, -0.5, 0, 0.5, 1\}$. Utilizamos como base de comparação a função $p_{zol}(\cdot)$.

Pelos resultados apresentados na Tabela B.2 observa-se que, para todos os métodos considerados, os valores das estatísticas D_1 e D_2 diminuem à medida que o valor de α aumenta. Além disso, esses valores, em geral, diminuem à medida que os valores de β aumentam. O método da FFT foi o que apresentou os maiores valores para essas estatísticas. Tal resultado é consistente com o reportado por Mittnik et al. (1999).

Embora de fácil implementação, o método das duas quadraturas mostrou-se mais instável do que o esperado. Na Tabela B.2 observa-se que o valor mais apropriado para o ponto de truncamento T varia com α. Entretanto, não podemos afirmar com certeza como se dá essa relação. Os resultados encontrados na Tabela B.2 indicam que,

- para $\alpha = 1.25$ (quando α está mais próximo de 1), o valor $T = 16$ conduz a resultados muito melhores do que $T = 8$. Porém, a diferença entre os valores obtidos com $T = 16$ e $T = 24$ é muito pequena.

- para $\alpha = 1.5$, observa-se pouca variação dos valores das estatísticas D_1 e D_2 em relação ao valor de T;
- para $\alpha = 1.75$ (quando α está mais próximo de 2), valores menores de T fornecem melhores estimativas da função densidade de probabilidade.

Tabela B.2: Valores das medidas de diferença D_1 e D_2, para $\alpha \in \{1.25, 1.5, 1.75\}$ e $\beta \in \{-1, -0.5, 0, 0.5, 1\}$, utilizando $v = \big(p_{zol}(x_0), \cdots, p_{zol}(x_{49})\big)$ e $w_L := \big(p_L(x_0), \cdots, p_L(x_{49})\big)$, para cada $L \in \{\Delta, \mathit{fft}, \mathit{quad}\}$.

α	1.25		1.50		1.75	
β	D_1	D_2	D_1	D_2	D_1	D_2
	$w = w_\Delta$					
-1.0	5.41e-08	1.83e-06	9.19e-09	4.23e-08	3.19e-09	1.54e-08
-0.5	1.78e-08	1.79e-08	8.62e-09	4.43e-08	3.05e-09	1.60e-08
0.0	1.79e-08	1.79e-08	7.99e-09	8.03e-09	2.97e-09	1.64e-08
0.5	2.21e-08	8.69e-08	1.31e-08	2.58e-07	2.31e-09	1.67e-08
1.1	2.09e-08	8.77e-08	1.13e-08	4.74e-08	1.88e-09	2.59e-09
β	$w = w_{\mathit{fft}}$, $N = 2^{14}$, $h = 0.01$					
-1.0	1.03e-05	1.06e-05	2.41e-06	2.49e-06	4.17e-07	4.32e-07
-0.5	1.03e-05	1.04e-05	2.38e-06	2.42e-06	4.10e-07	4.18e-07
0.0	1.02e-05	1.02e-05	2.34e-06	2.35e-06	4.03e-07	4.04e-07
0.5	1.00e-05	1.01e-05	2.30e-06	2.34e-06	3.95e-07	4.02e-07
1.1	9.80e-06	1.01e-05	2.27e-06	2.34e-06	3.88e-07	4.02e-07
β	$w = w_{quad}$, $T = 8$					
-1.0	2.69e-04	6.06e-04	1.87e-08	3.34e-08	7.59e-11	7.61e-11
-0.5	3.44e-04	6.36e-04	1.98e-08	3.34e-08	7.58e-11	7.59e-11
0.0	2.81e-04	6.39e-04	1.80e-08	3.34e-08	7.58e-11	7.58e-11
0.5	1.66e-04	3.90e-04	1.48e-08	2.97e-08	7.57e-11	7.58e-11
1.1	1.16e-04	2.42e-04	1.14e-08	2.33e-08	7.56e-11	7.58e-11
β	$w = w_{quad}$, $T = 16$					
-1.0	1.09e-08	2.18e-08	3.23e-09	3.26e-09	5.12e-10	5.14e-10
-0.5	1.12e-08	2.49e-08	3.22e-09	3.23e-09	5.11e-10	5.12e-10
0.0	1.09e-08	2.49e-08	3.21e-09	3.21e-09	5.10e-10	5.10e-10
0.5	1.04e-08	1.87e-08	3.20e-09	3.21e-09	5.09e-10	5.10e-10
1.1	1.04e-08	1.59e-08	3.18e-09	3.21e-09	5.08e-10	5.10e-10
	$w = w_{quad}$, $T = 24$					
-1.0	2.71e-08	2.77e-08	8.95e-09	9.05e-09	1.56e-09	1.57e-09
-0.5	2.68e-08	2.72e-08	8.89e-09	8.95e-09	1.56e-09	1.56e-09
0.0	2.65e-08	2.65e-08	8.84e-09	8.84e-09	1.55e-09	1.56e-09
0.5	2.62e-08	2.65e-08	8.79e-09	8.84e-09	1.55e-09	1.56e-09
1.1	2.58e-08	2.64e-08	8.73e-09	8.84e-09	1.55e-09	1.56e-09

Apêndice C

Artigo Lopes e Prass (2013a)

Autores: Sílvia R. C. Lopes e Taiane S. Prass

Título: Theoretical Results on Fractionally Integrated Exponential Generalized Autoregressive Conditional Heteroskedastic Processes

Revista: Physica A. Em Revisão

Ano: 2013

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$(q,d,0)$, 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 also 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 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 it 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.

ARCH-type models, proposed by [1], constitute one of the main classes of econometric models used for representing the dynamic evolution of volatilities. Another popular one is the class of Stochastic Volatility (SV) models (see, [2] and references therein). In both cases, ARCH-type and SV models, the stochastic process $\{X_t\}_{t\in\mathbb{Z}}$ can be written as

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

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 by the

*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

past informations until time $t-1$. An important difference between these two classes is that, 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 $\{Z_t\}_{t\in\mathbb{Z}}$. Therefore, the volatility of a SV process is specified as a latent variable which is not directly observable and this can make the estimation challenging, which is a known drawback of this class of models.

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

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

(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 (GARCH) processes, proposed by [3], and the Exponential GARCH (EGARCH) processes, proposed by [4]. These models are given, respectively, by (1) and (2) below by setting $d = 0$. The usual definition of σ_t^2 for a GARCH(p^*, q) model, namely,

$$\sigma_t^2 = \omega + \sum_{i=1}^{p^*} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \quad \text{for all} \quad t \in \mathbb{Z},$$

is obtained from (1) by letting $p^* := \max\{p, q\}$ and $\alpha(z) = \sum_{i=1}^{p^*} \alpha_i z^i := \beta(z) - \phi(z)$, where $\beta(z) := 1 - \sum_{j=1}^{q} \beta_j z^j$ and $\phi(z) := 1 - \sum_{k=1}^{p} \phi_k z^k$.

ARCH, GARCH and EGARCH are all short memory models. Among the generalizations that capture the effects of long-memory characteristic in the conditional variance are the Fractionally Integrated GARCH (FIGARCH), proposed by [5], and the Fractionally Integrated EGARCH (FIEGARCH), introduced by [6]. For a FIGARCH(p, d, q), σ_t^2 is given by

$$\left[1 - \sum_{j=1}^{q} \beta_j \mathcal{B}^k\right] \sigma_t^2 = \omega + \left(1 - \sum_{j=1}^{q} \beta_j \mathcal{B}^k - \left[1 - \sum_{k=1}^{p} \phi_k \mathcal{B}^k\right](1-\mathcal{B})^d\right) X_t^2, \quad \text{for all } t \in \mathbb{Z}, \tag{1}$$

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

$$\begin{aligned} \ln(\sigma_t^2) &= \omega + \frac{1 - \sum_{i=1}^{p} \alpha_i \mathcal{B}^i}{1 - \sum_{j=1}^{q} \beta_j \mathcal{B}^j} (1-\mathcal{B})^{-d} \big(\theta Z_{t-1} + \gamma[|Z_{t-1}| - \mathbb{E}(|Z_{t-1}|)]\big) \\ &:= \omega + \frac{\alpha(\mathcal{B})}{\beta(\mathcal{B})} (1-\mathcal{B})^{-d} g(Z_{t-1}), \quad \text{for all } t \in \mathbb{Z}, \end{aligned} \tag{2}$$

where $\mathcal{B}$ is the backward shift operator defined by $\mathcal{B}^k(X_t) = X_{t-k}$, for all $k \in \mathbb{N}$, and $(1-\mathcal{B})^d$ is the operator defined by its Maclaurin series expansion as,

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

with $\Gamma(\cdot)$ the gamma function.

FIEGARCH models have not only the capability of modeling clusters of volatility (as in the ARCH and GARCH models) and capturing its asymmetry[3] (as in the EGARCH models) but they also take into account the characteristic of long memory in the volatility (as in the FIGARCH models, with the advantage

[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.

of been weakly stationary if $d < 0.5$). Besides non-stationarity (in the weak sense), another drawback of the FIGARCH(p, d, q) models is that we must have $d \geq 0$ and the polynomial coefficients in its definition must satisfy some restrictions so 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.

Some authors argue that the long memory behavior observed in the sample autocorrelation and periodogram functions of financial time series could actually be caused by the non-stationarity property. According to [7], long range behavior could be just an artifact due to structural changes. On the other hand, [7] also argue that, when modeling return series with large sample size, considering a single GARCH model is unfeasible and that the best alternative would be to update the parameter values along the time. As an alternative to the traditional heteroskedastic models, [8] presents a regime switching model that, combined with heavy tailed distributions, presents the long memory characteristic.

It is our belief that FIEGARCH models are a competitive alternative for modeling large sample sized data, especially because they avoid parameter updating. Also, as we prove in this work, FIEGARCH processes are weakly stationary if and only if $d < 0.5$ and hence, non-stationarity can be easily identified. Moreover, [9] analyze the daily returns of the Tunisian stock market and rule out the random walk hypothesis. According to the authors, the rejection of this hypothesis seems to be due to substantial non-linear dependence and not to non-stationarity in the return series and, after comparing several ARCH-type models they concluded that a stationary FIEGARCH model provides the best fit for the data. Furthermore, [10] presents a sub period investigation of long memory and structural changes in volatility. The authors consider FIEGARCH models to examine the long run persistence of stock return volatility for 23 developing markets for the period of January 2000 to October 2007. No clear evidence that long memory characteristic could be attributed to structural changes in volatility was found.

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, [10] and [11], 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 ([12],[13],[14]). In this work we present a theoretical study on the main properties of FIEGARCH(p, d, q) processes, for any $p, q \geq 0$.

One of the contributions of the paper is to extend, for any p and q, the results already known in the literature for $p, q \in \{0, 1\}$ or $d = 0$. In particular, we provide the expressions for the asymmetry and kurtosis measures of FIEGARCH(p, d, q) process, for all $p, q \geq 0$. These results extends the one in [11] where only the case $p = 0$ and $q = 1$ was considered and only the kurtosis measure was derived.

Another contribution of this work is the ARFIMA representation of $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, when $\{X_t\}_{t\in\mathbb{Z}}$ is a FIEGARCH process, which is derived in the paper. This results is very useful in model identification and parameter estimation since the literature of ARFIMA models is well developed (see [15] and references therein) and, to the best of our knowledge, this result is absent in the literature.

To derive the properties of $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, we first investigate the conditions for the existence of power series representation for $\lambda(z) = \alpha(z)[\beta(z)]^{-1}(1 - z)^{-d}$ and the behavior of the coefficients in this representation. This study is fundamental not only for simulation purposes but also to draw conclusions on the autocorrelation and spectral density functions decay of the non-observable process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ and the observable one $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$. 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 forecast for the processes $\{X_t\}_{t\in\mathbb{Z}}$ and $\{X_t^2\}_{t\in\mathbb{Z}}$. We also present the h-step ahead forecast for both $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ and $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ processes, with their respective mean square error forecast. To the best of our knowledge, formal proofs for these expressions are not given in the literature of FIEGARCH(p, d, q) processes.

We also present a simulation study including generation, estimation and forecasting features of FIEGARCH models. 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 [16])[4].

[4]The asymptotic properties for the quasi-likelihood method are well established for ARCH/GARCH models (see, for instance,

Therefore, we consider here a simulation study to investigate the finite sample performance of the estimator. Since it is expected that, the better the fit, the better the forecasting, we also investigate the fitted models' forecasting performance.

The paper is organized as follows: Section 2 presents the formal definition of FIEGARCH process and its theoretical properties. We give a recurrence formula to obtain the coefficients in the power series expansion of the polynomial that describes the volatility and we show their asymptotic properties. 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 also presented and analyzed. The asymmetry and kurtosis measures of any stationary FIEGARCH process are also presented. Section 3 presents the theoretical results regarding the forecasting. Section 4 presents a Monte Carlo simulation study including the generation of FIEGARCH time series, estimation of the model parameters and the forecasting 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 present the *Fractionally Integrated Exponential Generalized Autoregressive Conditional Heteroskedastic* process (FIEGARCH). This class of processes, introduced by [6], describes not only the volatility varying on time and the volatility clusters (known as ARCH/GARCH effects) but also the volatility long-range dependence and its asymmetry.

Here, we present some results related to the existence, stationarity and ergodicity for these processes. We analyze the autocorrelation and the spectral density functions decay for both $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ and $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ processes. Conditions for the existence of a series expansion for the polynomial that describes the volatility are given and a recurrence formula to calculate the coefficients of this expansion is presented. We also discuss the coefficients asymptotic behavior. We observe that if $\{X_t\}_{t\in\mathbb{Z}}$ is a FIEGARCH(p,d,q) process then $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is an ARFIMA(q,d,p) process and we prove that, under mild conditions, $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is an ARFIMA$(q,d,0)$ process with correlated innovations. We present the expression for the kurtosis and the asymmetry measures for any stationary FIEGARCH(p,d,q) process.

Throughout the paper, 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.

From now on, let $(1-\mathcal{B})^d$ be the operator defined by its Maclaurin series expansion as,

$$(1-\mathcal{B})^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} := \sum_{k=0}^{\infty} \delta_{d,k}\,\mathcal{B}^k := \delta_d(\mathcal{B}), \tag{3}$$

where $\Gamma(\cdot)$ is the gamma function, $\mathcal{B}$ is the backward shift operator defined by $\mathcal{B}^k(X_t) = X_{t-k}$, for all $k \in \mathbb{N}$, and the coefficients $\delta_{d,k}$ are such that $\delta_{d,0} = 1$ and $\delta_{d,k-1} = \delta_{d,k-1}\big(\frac{k-1-d}{k}\big)$, for all $k \geq 1$.

Remark 1. Note that expression (3) 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 (3), that is

$$(1-\mathcal{B})^{-d} = \sum_{j=0}^{\infty} \delta_{-d,j}\mathcal{B}^j := \sum_{k=0}^{\infty} \pi_{d,k}\mathcal{B}^k, \tag{4}$$

where $\pi_{d,j} = \delta_{-d,j}$, for all $j \in \mathbb{N}$. Moreover, $\pi_{d,k} \sim \frac{1}{\Gamma(d)\,k^{1-d}}$, as $k \to \infty$ (see [14]). Therefore, $\pi_{d,k} = O(k^{d-1})$, as k goes to infinity.

[17], [18], [19], [20] and [21]) and also for EGARCH models (see, for instance, [22]).

Suppose that $\{Z_t\}_{t\in\mathbb{Z}}$ is a sequence of independent and identically distributed (i.i.d.) random variables, with zero mean and variance equal to one. Let $\alpha(\cdot)$ and $\beta(\cdot)$ be the polynomials of order p and q defined, respectively, by

$$\alpha(z) = \sum_{i=0}^{p}(-\alpha_i)z^i = 1 - \sum_{i=1}^{p}\alpha_i z^i \quad \text{and} \quad \beta(z) = \sum_{j=0}^{q}(-\beta_j)z^j = 1 - \sum_{j=1}^{q}\beta_j z^j, \tag{5}$$

with $\alpha_0 = \beta_0 = -1$. We assume that $\beta(z) \neq 0$, if $|z| \leq 1$, and that $\alpha(\cdot)$ and $\beta(\cdot)$ have no common roots. These conditions assure that the operator $\frac{\alpha(\mathcal{B})}{\beta(\mathcal{B})}$ is well defined.

Definition 1. Let $\{X_t\}_{t\in\mathbb{Z}}$ be the stochastic process defined as

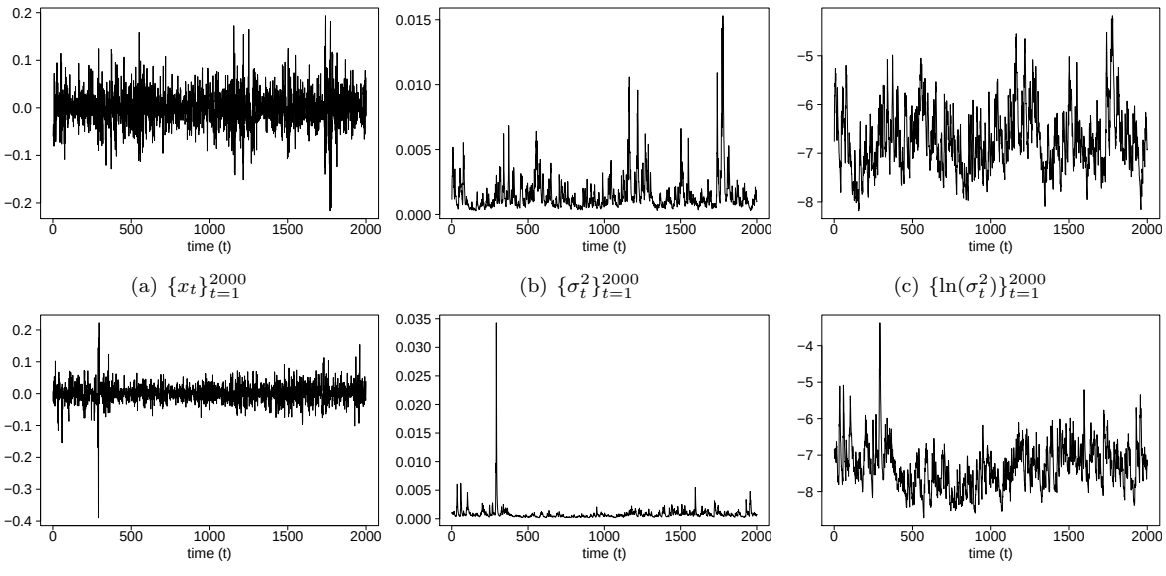
$$X_t = \sigma_t Z_t, \tag{6}$$

$$\ln(\sigma_t^2) = \omega + \frac{\alpha(\mathcal{B})}{\beta(\mathcal{B})}(1-\mathcal{B})^{-d} g(Z_{t-1}), \quad \text{for all } t \in \mathbb{Z}, \tag{7}$$

where $\omega \in \mathbb{R}$ and $g(\cdot)$ is defined by

$$g(Z_t) = \theta Z_t + \gamma\left[|Z_t| - \mathbb{E}(|Z_t|)\right], \quad \text{for all } t \in \mathbb{Z}, \quad \text{with } \theta, \gamma \in \mathbb{R}. \tag{8}$$

Then $\{X_t\}_{t\in\mathbb{Z}}$ is a *Fractionally Integrated* EGARCH *process*, denoted by FIEGARCH(p, d, q).



(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 $n = 2,000$ observations, considering $Z_0 \sim \mathcal{N}(0, 1)$ (first row) and $Z_0 \sim$ 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.

Example 1. Figure 1 presents samples from FIEGARCH$(0, d, 1)$ processes, with $n = 2,000$ observations, considering two different underlying distributions. To obtain these samples we consider Definition 1 and two different distributions for Z_0. For this simulation 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 the distribution for Z_0 is the Generalized Error Distribution (GED), with tail-thickness parameter $\nu = 1.5$.

Remark 2. Note that, in Definition 1, no conditions on the parameter d are imposed. Necessary and sufficient conditions on the parameter d, to guarantee the existence of the stochastic process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$, satisfying (7), are discussed in the sequel. Also notice that when $d = 0$, we obtain the well known EGARCH process.

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 are equivalent to (2). For instance, [23] defines the conditional variance of a FIEGARCH process through the equation

$$\left[1 - \sum_{j=1}^{q} \beta_j \mathcal{B}^j\right](1-\mathcal{B})^d \ln(\sigma_t^2) = a + \sum_{i=0}^{p} [\psi_i |Z_{t-1-i}| + \gamma_i Z_{t-1-i}].$$

This is the definition considered, for instance, in the software S-Plus (see [23]) and it is equivalent to (2) whenever $d > 0$, $a = -\gamma\mathbb{E}(|Z_t|)[1 - \sum_{i=1}^p \alpha_i]$, $\psi_0 = \gamma$, $\gamma_0 = \theta$, $\psi_i = -\gamma\alpha_i$ and $\gamma_i = -\theta\alpha_i$ for all $1 \leq i \leq p$. This equivalence is mentioned in [16] and a detailed proof is provided in [14]. In [11] only the case $p = 0$ and $q = 1$ is considered and $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is defined as

$$(1-\phi\mathcal{B})(1-\mathcal{B})^d \ln(\sigma_t^2) = \omega^* + \alpha\big[|Z_{t-1}| - \sqrt{2/\pi}\big] + \gamma^* Z_{t-1}, \quad \text{for all } t \in \mathbb{Z}, \tag{9}$$

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 [24]. Notice that, by setting $\phi = \beta_1$, $\alpha = \gamma$ and $\gamma^* = \theta$, (9) is equivalent to (7) if and only if the equality $\omega^* = (1-\beta)(1-\mathcal{B})^d\omega$ holds.

Remark 3. We observe that the theory presented here can be easily adapted to a more general framework than (7) (which uses the same notation as in [6]) by considering

$$\ln(\sigma_t^2) = \omega_t + \sum_{k=0}^{\infty} \lambda_k g(Z_{t-1-k}) := \omega_t + \lambda(\mathcal{B})g(Z_{t-1}), \quad \text{for all } t \in \mathbb{Z}, \tag{10}$$

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 white noise process with variance not necessarily equal to one and $g(\cdot)$ is any measurable function. In particular, Theorems 1 and 2 below, which are stated and proved in [4], assume that $\ln(\sigma_t^2)$ is given by (10) (the notation was adapted to reflect the one used in this work), with $\{Z_t\}_{t\in\mathbb{Z}}$ and $g(\cdot)$ as in Definition 1. Although (10) is more general than (7), in practice the applicability of the model is somewhat limited given that the parameter estimation is far more complicated when compared to the model (7).

Notice that the function $g(\cdot)$ can be rewritten as

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

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. 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. *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 (8) 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*

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

Proof. See [14]. □

Theorem 1 below provides a criterion for stationarity and 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 (10) 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 (5), leading to an EGARCH(p, q) process. For this model, the criterion in Theorem 1 holds whenever the roots of $\beta(\cdot)$ are outside 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 (7), leading to a FIEGARCH(p, d, q) process.

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

$$X_t = \sigma_t Z_t; \qquad Z_t \sim i.i.d., \quad \mathbb{E}(Z_t) = 0 \quad \textit{and} \quad \mathrm{Var}(Z_t) = 1; \tag{12}$$

$$\ln(\sigma_t^2) = \omega_t + \sum_{k=1}^{\infty} \lambda_k g(Z_{t-k}), \quad \lambda_1 = 1; \qquad g(Z_t) = \theta Z_t + \gamma\left[|Z_t| - \mathbb{E}(|Z_t|)\right]; \tag{13}$$

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 $\mathrm{Cov}\big(\ln(\sigma_t^2), \ln(\sigma_{t-k}^2)\big) = \mathrm{Var}\big(g(Z_t)\big)\sum_{j=1}^{\infty}\lambda_j\lambda_{j+k}$.

Proof. See theorem 2.1 in [4]. □

Theorem 2 shows the existence of the rth moment for the random variables X_t and σ_t^2, defined by (12)-(13), when $\sum_{k=1}^{\infty}\lambda_k^2 < \infty$ and the distribution of Z_0 is the Generalized Error Distribution (GED).

Theorem 2. *Define $\{\sigma_t^2, X_t\}_{t\in\mathbb{Z}}$ by (12)-(13), 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 rth moments of $e^{-\omega_t}\sigma_t^2$ and $e^{-\omega_t/2}X_t$, as $t \to \infty$:*

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

where plim *denotes the limit in probability.*

Proof. See theorem 2.2 in [4]. □

From now on, let $\lambda(\cdot)$ be the polynomial defined by

$$\lambda(z) = \frac{\alpha(z)}{\beta(z)}(1-z)^{-d} := \sum_{k=0}^{\infty}\lambda_{d,k}z^k, \quad \text{for all } |z| < 1, \tag{14}$$

where $\alpha(\cdot)$ and $\beta(\cdot)$ are defined in (5). 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 disc $\{z : |z| < 1\}$ (if $d \leq 0$, in the closed disk $\{z : |z| \leq 1\}$). Therefore, it has a unique power series representation and (7) can be rewritten, equivalently, as

$$\ln(\sigma_t^2) = \omega + \sum_{k=0}^{\infty} \lambda_{d,k} g(Z_{t-1-k}) = \omega + \lambda(\mathcal{B}) g(Z_{t-1}), \quad \text{for all } t \in \mathbb{Z}. \tag{15}$$

Notice that, with this definition we obtain a particular case of parametrization (10).

Theorem 3 below gives the convergence order of the coefficients $\lambda_{d,k}$, 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 (14). Then, for all $k \in \mathbb{N}$, the coefficients $\lambda_{d,k}$ satisfy*

$$\lambda_{d,k} \sim \frac{1}{\Gamma(d)\, k^{1-d}} \frac{\alpha(1)}{\beta(1)}, \qquad \text{as } k \to \infty. \tag{16}$$

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

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

$$\beta(z)^{-1} := f(z) = \sum_{k=0}^{\infty} f_k z^k, \quad \text{where } f_k = \frac{f^{(k)}(0)}{k!}, \text{ for all } k \in \mathbb{N}. \tag{17}$$

From expressions (4), (14) and (17) it follows that

$$\lambda(z) = \sum_{k=0}^{\infty} \left[\sum_{i=0}^{\min\{p,k\}} (-\alpha_i) \sum_{j=0}^{k-i} \pi_{d,k-i-j} f_j \right] z^k. \tag{18}$$

From (18), 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 $0 \leq j \leq m$ and $0 \leq i \leq p$. 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 $m_0 < m + 1 \leq j \leq k - i$ and $0 \leq i \leq p$. 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. □

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 (14). The coefficients $\lambda_{d,k}$, for all $k \in \mathbb{N}$, are given by*

$$\lambda_{d,0} = 1 \quad \text{and} \quad \lambda_{d,k} = -\alpha_k^* + \sum_{i=0}^{k-1} \lambda_i \sum_{j=0}^{k-i} \beta_j^* \delta_{d,k-i-j}, \quad \text{for all} \quad k \geq 1, \tag{19}$$

where the coefficients $\delta_{d,k}$, for all $k \in \mathbb{N}$, are given in (3) and

$$\alpha_m^* := \begin{cases} \alpha_m, & \text{if} \quad 0 \leq m \leq p; \\ 0, & \text{if} \quad m > p \end{cases} \quad \text{and} \quad \beta_m^* := \begin{cases} \beta_m, & \text{if} \quad 0 \leq m \leq q; \\ 0, & \text{if} \quad m > q. \end{cases} \tag{20}$$

Proof. Let $\lambda(\cdot)$ be defined by (14). Consequently,

$$\alpha(z) = \beta(z)(1-z)^d \sum_{k=0}^{\infty} \lambda_{d,k} z^k. \tag{21}$$

By defining β_k^* as in expression (20), for all $k \in \mathbb{N}$, and upon considering expression (3), observing that $\delta_{d,0} = -1 = \beta_0$, the right hand side of expression (21) can be rewritten as

$$\begin{aligned} \beta(z)(1-z)^d \sum_{k=0}^{\infty} \lambda_{d,k} z^k &= \left[\sum_{k=0}^{\infty} \left(\sum_{j=0}^{k} -\beta_j^* \delta_{d,k-j} \right) z^k \right] \sum_{k=0}^{\infty} \lambda_{d,k} z^k = \sum_{k=0}^{\infty} \left[\sum_{i=0}^{k} \lambda_{d,i} \left(-\sum_{j=0}^{k-i} \beta_j^* \delta_{d,k-i-j} \right) \right] z^k \\ &= \sum_{k=0}^{\infty} \left[\lambda_{d,k} - \sum_{i=0}^{k-1} \lambda_{d,i} \sum_{j=0}^{k-i} \beta_j^* \delta_{d,k-i-j} \right] z^k. \end{aligned} \tag{22}$$

Now, by setting α_k^* as in expression (20), for all $k \in \mathbb{N}$, from expression (22) one concludes that the equality (21) holds if and only if,

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

Therefore, expression (19) holds. It is easy to see that by replacing the coefficients $\lambda_{d,k}$, given by (19), in the expression (22), for all $k \in \mathbb{N}$, we get $\sum_{k=0}^{\infty} (-\alpha_k^*) z^k = \alpha(z)$, which completes the proof. □

The applicability of Theorem 1 to long memory 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 (7), which leads to a long memory model whenever $d > 0$.

Corollary 1. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a* FIEGARCH(p,d,q) *process, given in Definition 1. 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.*

Proof. Let $\{\lambda_{d,k}\}_{k\in\mathbb{N}}$ be defined by (14) and rewrite (7) as (15). 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 $k \geq 1$. □

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 (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, [25], [15]). 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,

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

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 4. The proof of **P1** can be found in [25], theorem 13.2.2. Regarding **P2**, in the literature one usually find that an ARFIMA(p, d, q) process is invertible for $|d| < 0.5$ (see, for instance, [25], theorem 13.2.2). However, [26] already proved that this range can be extended to $d \in (-1, 0.5)$, for an ARFIMA$(0, d, 0)$ and, more recently, [27] 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 stationary and ergodic processes. However, as mentioned in [4], this does not imply weakly 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, given in Definition 1. 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(v) *random variables, with $v > 1$, 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 $r > 0$.*

Proof. Let $\{\lambda_{d,k}\}_{k\in\mathbb{N}}$ be defined by (14) and rewrite (7) as (15). 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. □

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(v) 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

$$A_X := \frac{\mathbb{E}(X_t^3)}{\big[\mathbb{E}(X_t^2)\big]^{3/2}} = \frac{\mathbb{E}(\sigma_t^3)\mathbb{E}(Z_t^3)}{\big[\mathbb{E}(\sigma_t^2)\big]^{3/2}} \quad \text{and} \quad K_X := \frac{\mathbb{E}(X_t^4)}{\big[\mathbb{E}(X_t^2)\big]^{2}} = \frac{\mathbb{E}(\sigma_t^4)\mathbb{E}(Z_t^4)}{\big[\mathbb{E}(\sigma_t^2)\big]^{2}}, \quad \text{for all } t \in \mathbb{Z}. \tag{23}$$

An expression for K_X (as a function of the FIEGARCH model parameters) was already given in [11] by assuming that $\{Z_t\}_{t\in\mathbb{Z}}$ is a Gaussian white noise with variance equal to one, $d > 0$, $p = 0$, $q = 1$ and by defining $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ through expression (9). According, 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 bellow we consider stationary FIEGARCH(p, d, q) processes (therefore, $d < 0.5$) with $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ defined by (7) 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 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, given in Definition 1. If $\mathbb{E}(X_0^3) < \infty$, the asymmetry measure of $\{X_t\}_{t\in\mathbb{Z}}$ is given by*

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

and, if $\mathbb{E}(X_0^4) < \infty$, the kurtosis measure of $\{X_t\}_{t\in\mathbb{Z}}$ is given by

$$K_X = \mathbb{E}(Z_0^4)\,\frac{\prod_{k=0}^{\infty}\mathbb{E}\big(\exp\{2\lambda_{d,k}g(Z_0)\}\big)}{\big[\prod_{k=0}^{\infty}\mathbb{E}\big(\exp\{\lambda_{d,k}g(Z_0)\}\big)\big]^{2}},$$

where $\lambda_{d,k}$ are given in (14) *and $g(\cdot)$ is defined by* (8)*.*

Proof. Let $\{X_t\}_{t\in\mathbb{Z}}$ be any stationary FIEGARCH(p, d, q) process and $\lambda(\cdot)$ be the polynomial defined by (14). Notice that, since $\{g(Z_t)\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables, from (7) it follows that

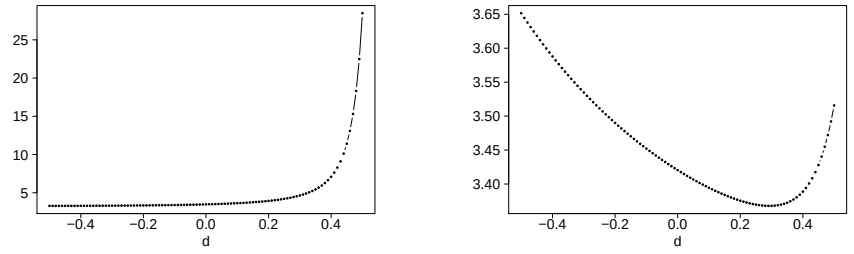
$$\mathbb{E}(\sigma_0^r) = e^{\frac{r\omega}{2}}\prod_{k=0}^{\infty}\mathbb{E}\Big(\exp\Big\{\frac{r}{2}\lambda_{d,k}g(Z_0)\Big\}\Big), \quad \text{for all } r > 0. \tag{24}$$

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 (24) converges, for any $r \leq 3$ (analogously, $r \leq 4$). Upon replacing (24) in (23) we conclude the proof. □

Example 2. Figure 2 shows the theoretical value of the kurtosis measure, as a function of the parameter d, for any 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) 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 $-0.5 \leq d < 0.3$ and increase for $0.3 \leq d \leq 0.5$.



(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 $\beta_1 = -0.6860$.

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 (6), 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, [28] and [29] consider models such that X_t can be written as in (6), but σ_t can have a more general definition than (7). While [28] consider maximum likelihood and Whittle's method of estimation in the class of exponential volatility models, especially the EGARCH ones, [29] consider 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 in the case where X_t can be written as in (6), and σ_t is defined through the expression (7) 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, given in Definition 1. 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*

$$\gamma_{\ln(X^2)}(h) = \sigma_g^2 \sum_{k=0}^{\infty} \lambda_{d,k}\lambda_{d,k+|h|} + \mathrm{Var}\big(\ln(Z_t^2)\big)\mathbb{I}_{\{0\}}(h) + \lambda_{d,|h|-1}\mathcal{K}\mathbb{I}_{\mathbb{Z}^*}(h), \quad \textit{for all } h \in \mathbb{Z}. \tag{25}$$

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

Proof. 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 (14) and rewrite (7) as (15).

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 [30]). 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 (25). From the definition of $\ln(X_t^2)$, it follows that

$$\begin{aligned}\mathrm{Cov}\big(\ln(X_{t+h}^2), \ln(X_t^2)\big) = \mathrm{Cov}\big(\ln(\sigma_{t+h}^2), \ln(\sigma_t^2)\big) + \mathrm{Cov}\big(\ln(Z_{t+h}^2), \ln(Z_t^2)\big) \\ + \mathrm{Cov}\big(\ln(\sigma_{t+h}^2), \ln(Z_t^2)\big) + \mathrm{Cov}\big(\ln(Z_{t+h}^2), \ln(\sigma_t^2)\big). \qquad (26)\end{aligned}$$

Theorem 1 shows that

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

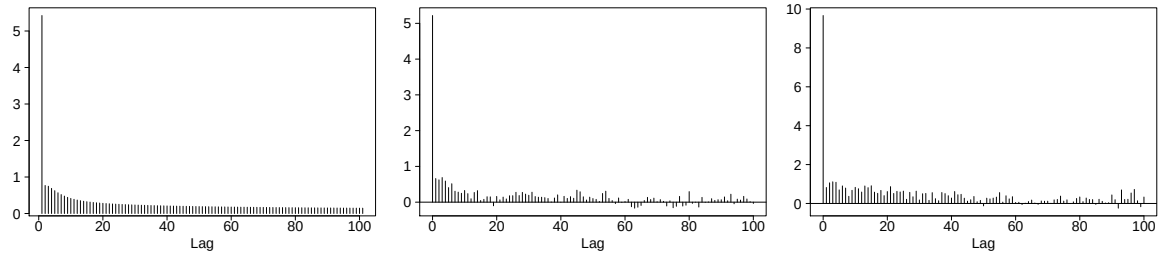
From the independence of the random variables $\ln(Z_t^2)$, for all $t \in \mathbb{Z}$, and from expression (15), we have

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

where $\mathcal{K} = \mathrm{Cov}\big(g(Z_0), \ln(Z_0^2)\big)$. Since $\mathrm{Cov}\big(\ln(Z_{t+h}^2), \ln(\sigma_t^2)\big) = \mathrm{Cov}\big(\ln(\sigma_{u-h}^2), \ln(Z_u^2)\big)$, with $u = t + h$, 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 (26) we conclude that the autocovariance function of $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is given by (25). □



(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 $\{\ln(r_t^2)\}_{t=1}^{n}$, where $\{r_t\}_{t=1}^{n}$ 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. Since the sample autocovariance function is an estimator of the theoretical autocovariance function, it is expected that their graphics will have the same behavior. The similarity between the decay in the graphs in Figure 3 (b) and (c) indicates that a FIEGARCH model seems appropriate for fitting the Bovespa index log-returns time series.

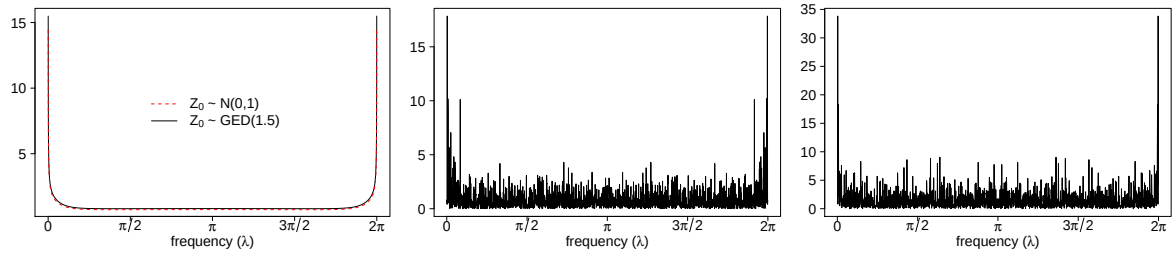
Example 4. Theorem 4 provides the expression for the autocorrelation function of $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$. The spectral density function of the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is given by (see [29])

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

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 any 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 $\mathrm{Var}(\ln(Z_0^2))$, used in the computation of $f_{\ln(X_t^2)}(\cdot)$, are given in Table 6. 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 e $\mathcal{K}$ 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) 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) indicates that the probability distribution of Z_0 may not be evident from the periodogram function given the similarity between the graphs 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 $Z_0 \sim$ GED(1.5) as shown in Table 6. Moreover, one observes that the graphs in Figure 4 (b) and (c) present similar behavior, indicating that a FIEGARCH model may be adequate to fit the data. On the other hand, Figure 4 (a) shows evidence that the underlying probability distribution of $\{r_t\}_{t=1}^{1717}$ may not be the same as $\{x_t\}_{t=1}^{2,000}$. 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 $< 2.2\times10^{-16}$). 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 we consider the same time series $\{x_t\}_{t=1}^{2000}$ as in Figure 4 (b) and perform the Kolmogorov-Smirnov hypothesis test as described in [25], pages 339 - 342, considering both cases $Z_0 \sim \mathcal{N}(0,1)$ and $Z_0 \sim \text{GED}(1.5)$. Recall that

- 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 on ploting the Kolmogorov-Smirnov boundaries

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

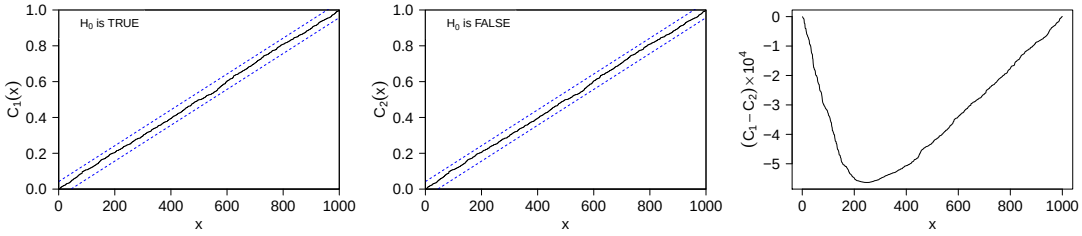
  and the function $C(x)$ defined as

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

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

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

  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 $1 \leq x \leq m$.



(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 $\{x_t\}_{t=1}^{2,000}$ 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) $Z_0 \sim \text{GED}(1.5)$ (therefore, the null hypothesis is false). In all cases, $d = 0.3578$, $\theta = -0.1661$, $\gamma = 0.2792$, $\omega = -7.2247$ and $\beta_1 = 0.6860$. On (c), the difference between the values of $C(x)$ (multiplied by 10^4) assuming, respectively, $Z_0 \sim \mathcal{N}(0,1)$ and $Z_0 \sim \text{GED}(1.5)$.

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 $0 \leq x \leq 1000$. 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 results is very useful in model identification and parameter estimation since the literature of ARFIMA models is well developed (see [15] and references therein).

Theorem 5. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a* FIEGARCH(p,d,q) *process, given in Definition 1. 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 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

$$\gamma_\varepsilon(h)=\begin{cases}\sigma_g^2\sum_{i=|h|}^{p}\alpha_i\alpha_{i-|h|}+\mathcal{K}\sum_{i=0}^{p}\alpha_i\phi_{i+|h|+1}+\mathcal{K}\sum_{i=|h|-1}^{p}\alpha_i\phi_{i-|h|+1}+\sigma_\ell^2\sum_{i=|h|}^{\infty}\phi_i\phi_{i-|h|}, & \textit{if}\quad 0\le|h|\le p;\\ \mathcal{K}\alpha_p\phi_1+\sigma_\ell^2\sum_{i=p+1}^{\infty}\phi_i\phi_{i-(p+1)}, & \textit{if}\quad |h|=p+1;\\ \sigma_\ell^2\sum_{i=|h|}^{\infty}\phi_i\phi_{i-|h|}, & \textit{if}\quad |h|>p+1,\end{cases}\tag{27}$$

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

$$\phi(z):=\beta(z)(1-z)^d=\sum_{k=0}^{\infty}\phi_k z^k,\quad \textit{for } |z|<1,\tag{28}$$

σ_g^2 given in (11)*, $\mathcal{K}=\mathrm{Cov}(g(Z_0),\ln(Z_0^2))$, $\{\alpha_i\}_{i=0}^p$ given in* (5) *and $\sigma_\ell^2:=\mathrm{Var}(\ln(Z_0^2))$.*

Proof. Let $\{X_t\}_{t\in\mathbb{Z}}$ be a FIEGARCH process. From expressions (6) and (7) 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 $0\le|\mathbb{E}(\ln(Z_0^2))|\le\mathbb{E}(|\ln(Z_0^2)|)\le[\mathbb{E}([\ln(Z_0^2)]^2)]^{1/2}$, 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 (28). Assume, for the moment, that $\mathrm{Var}(\varepsilon_t^2)<\infty$, for all $t\in\mathbb{Z}$. It follows that

$$\begin{aligned}\mathrm{Cov}(\varepsilon_{t+h},\varepsilon_t)=&\,\mathrm{Cov}\big(\alpha(\mathcal{B})g(Z_{t+h-1}),\alpha(\mathcal{B})g(Z_{t-1})\big)+\mathrm{Cov}\big(\phi(\mathcal{B})\ln(Z_{t+h}^2),\phi(\mathcal{B})\ln(Z_t^2)\big)\\&+\mathrm{Cov}\big(\alpha(\mathcal{B})g(Z_{t+h-1}),\phi(\mathcal{B})\ln(Z_t^2)\big)+\mathrm{Cov}\big(\phi(\mathcal{B})\ln(Z_{t+h}^2),\alpha(\mathcal{B})g(Z_{t-1})\big).\end{aligned}\tag{29}$$

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

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

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

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

and

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

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

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

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

Therefore, all four terms in expression (29) do not depend on $t \in \mathbb{Z}$ and expression (27) holds. Now, to validate expression (27) we only need to show that $\mathrm{Var}(\varepsilon_t) < \infty$, for all $t \in \mathbb{Z}$. Notice that, since $\mathbb{E}(\varepsilon_t) = 0$, it follows that $\mathbb{E}(\varepsilon_t^2) = \mathrm{Var}(\varepsilon_t) = \gamma_{\varepsilon}(0)$. Upon replacing $h = 0$ in (29) one obtains

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

By hypothesis, $\mathbb{E}([\ln(Z_0^2)]^2) < \infty$ and $d \in (-0.5, 0.5)$. It follows that $\mathrm{Var}(\ln(Z_0^2)) < \infty$ and $\sum_{i=0}^{\infty} \phi_i^2 < \infty$. We also know that $\mathrm{Var}(g(Z_0)) < \infty$. In order to show that $\mathcal{K} < \infty$, notice that $\mathcal{K} := \mathrm{Cov}(g(Z_0), \ln(Z_0^2)) = \mathbb{E}(g(Z_0)\ln(Z_0^2))$ and, since $\mathbb{E}(Z_0^2) = 1$ and $\mathrm{Var}(\ln(Z_0^2)) < \infty$, from Hölder's inequality, we have $\mathbb{E}(|Z_0|) < \infty$ and $\mathbb{E}(\ln(Z_0^2)) < \infty$. Then from (8) 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

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

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

3. Forecasting

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a FIEGARCH(p, d, q) process, given in Definition 1, 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 $\mathcal{F}_t := \sigma(\{Z_s\}_{s\leq t})$, 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 (7), 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 [15] and [31]). 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 5. 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 "$\hat{}$" to denote the h-ahead step 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 "$\tilde{}$" and "$\check{}$" 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 "$\tilde{}$" and "$\check{}$");
- we follow the approach usually considered in the literature and denote the h-ahead step forecast Y_{t+h}^2 as $\hat{Y}_{t+h}^2$ instead of $\widehat{Y_{t+h}^2}$. If necessary, to avoid confusion, we will denote the square of $\hat{Y}_{t+h}$ as $(\hat{Y}_{t+h})^2$ (analogously for "$\tilde{}$" and "$\check{}$").

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_{n+1}^2, given $\mathcal{F}_n$.

Lemma 1. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a* FIEGARCH(p, d, q) *process, given in Definition 1 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 $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$.*

Proof. 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 $\mathbb{E}(X_t|\mathcal{F}_{t-1}) = \mathbb{E}(\sigma_t Z_t|\mathcal{F}_{t-1}) = \sigma_t\mathbb{E}(Z_t|\mathcal{F}_{t-1}) = 0$. Therefore, the process $\{X_t\}_{t\in\mathbb{Z}}$ is a martingale difference with respect to $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$. □

Lemma 2. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a stationary* FIEGARCH(p, d, q) *process, given by Definition 1. 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_{n+1}^2, given $\mathcal{F}_n$, are, respectively, $\hat{X}_{n+h} = 0$ and $\hat{X}_{n+1}^2 = \sigma_{n+1}^2$.*

Proof. 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. □

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

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

While for ARCH/GARCH models, $\mathbb{E}(\sigma_{n+h}^2|\mathcal{F}_t)$ can be easily calculated, for FIEGARCH processes, what is easy to derive is the expression for the h-step ahead forecast for the process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$, for any $h > 1$. The expressions for $\hat{\ln}(\sigma_{n+h}^2) := \mathbb{E}(\ln(\sigma_{n+h}^2)|\mathcal{F}_t)$ 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 $\check{\sigma}_{n+h}^2 := \exp\{\hat{\ln}(\sigma_{n+h}^2)\}$, for all $h > 0$.

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

$$\hat{\ln}(\sigma_{n+h}^2) = \omega + \sum_{k=0}^{\infty} \lambda_{d,k+h-1}\, g(Z_{n-k}), \quad \textit{for all } h > 0. \tag{30}$$

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

$$\mathbb{E}\big(\big[\ln(\sigma^2_{n+h}) - \hat{\ln}(\sigma^2_{n+h})\big]^2\big) = \sigma_g^2 \sum_{k=0}^{h-2} \lambda^2_{d,k}, \quad \textit{if } h \geq 2, \tag{31}$$

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

Proof. Let $\hat{\ln}(\sigma^2_{n+h}) := \mathbb{E}(\ln(\sigma^2_{n+h})|\mathcal{F}_n)$. Note that $\mathbb{E}(g(Z_t)|\mathcal{F}_n) = \mathbb{E}(g(Z_t)) = 0$, for all $t > n$, and $\mathbb{E}(g(Z_t)|\mathcal{F}_n) = g(Z_t)$, for all $t \leq n$. By (15), one has

$$\hat{\ln}(\sigma^2_{n+h}) = \omega + \sum_{k=0}^{\infty} \lambda_{d,k}\, \mathbb{E}(g(Z_{n+h-1-k})|\mathcal{F}_n) = \omega \; + \sum_{k=h-1}^{\infty} \lambda_{d,k}\, g(Z_{n+h-1-k}),$$

and expression (30) follows.

Since $\ln(\sigma^2_{n+h})$ is a function of $\{g(Z_s)\}_{s\leq n+h-1}$ and $\{g(Z_t)\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables with zero mean and variance $\sigma_g^2 := \mathbb{E}([g(Z_0)]^2)$, we conclude that

$$\mathbb{E}\big(\big[\ln(\sigma^2_{n+h}) - \hat{\ln}(\sigma^2_{n+h})\big]^2\big) = \mathbb{E}\bigg(\bigg[\sum_{k=0}^{h-2} \lambda_{d,k}\, g(Z_{n+h-1-k})\bigg]^2\bigg) = \sigma_g^2 \sum_{k=0}^{h-2} \lambda^2_{d,k}, \quad \text{if } h \geq 2,$$

and zero if $h = 1$. □

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 (30), 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

$$\check{\ln}(X^2_{n+h}) := \ln(\check{X}^2_{n+h}) = \ln(\check{\sigma}^2_{n+h}) = \hat{\ln}(\sigma^2_{n+h}), \quad \text{for all} \quad h > 0. \tag{32}$$

From expressions (6) and (32), 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 forecast for the h-step ahead forecast of $\ln(X^2_{n+h})$, defined through expression (32).

Proposition 5. *Let $\check{\ln}(X^2_{n+h})$, for all $h > 0$, be the h-step ahead forecast of $\ln(X^2_{n+h})$, given the filtration $\mathcal{F}_n = \sigma(\{Z_s\}_{s\leq n})$, defined by expression* (32). *Then the mean square error forecast is given by*

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

Proof. By expression (32), $\check{\ln}(X^2_{n+h}) := \hat{\ln}(\sigma^2_{n+h})$, for all $h > 0$. Thus, from expression (6) and from Proposition 4, we have

$$\begin{aligned}\mathbb{E}\big(\big[\ln(X^2_{n+h}) - \check{\ln}(X^2_{n+h})\big]^2\big) &= \mathbb{E}\big(\big[\ln(X^2_{n+h}) - \ln(\hat{\sigma}^2_{n+h})\big]^2\big) = \mathbb{E}\big(\big[\ln(\sigma^2_{n+h}) + \ln(Z^2_{n+h}) - \ln(\hat{\sigma}^2_{n+h})\big]^2\big) \\ &= \mathbb{E}\bigg(\bigg[\sum_{k=0}^{h-2} \lambda_{d,k}\, g(Z_{n+h-1-k}) + \ln(Z^2_{n+h})\bigg]^2\bigg). \end{aligned} \tag{33}$$

By expanding the right hand side of expression (33) and using the fact that $\{Z_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables, the proposition follows immediately. □

Remark 6. If the values of X_t and σ_t are known only for $t \in \{1, \cdots, n\}$ then the h-step ahead forecast $\hat{\ln}(\hat{\sigma}^2_{n+h})$ of $\ln(\sigma^2_{n+h})$, is approximated by

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

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

$$\mathbb{E}\big(\big[\ln(\sigma^2_{n+h}) - \hat{\ln}(\sigma^2_{n+h})\big]^2\big) = \sigma_g^2\bigg(\sum_{k=0}^{h-2}\lambda^2_{d,k} + \sum_{k=n+h-1}^{\infty}\lambda^2_{d,k}\bigg), \quad \text{and}$$

$$\mathbb{E}\big(\big[\ln(X^2_{n+h}) - \check{\ln}(X^2_{n+h})\big]^2\big) = \sigma_g^2\bigg(\sum_{k=0}^{h-2}\lambda^2_{d,k} + \sum_{k=n+h-1}^{\infty}\lambda^2_{d,k}\bigg) + \mathbb{E}(\big[\ln(Z^2_{n+h})\big]^2, \quad \text{for all } h > 0.$$

From Jensen's inequality, one concludes that

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

so that $\mathbb{E}(\check{\sigma}^2_{n+h} - \sigma^2_{n+h}) = \mathbb{E}(\mathbb{E}(\check{\sigma}^2_{n+h} - \sigma^2_{t+h}|\mathcal{F}_n)) = \mathbb{E}(\check{\sigma}^2_{n+h} - \hat{\sigma}^2_{n+h}) \le 0$, for all $h > 0$. In fact, from (16) and (30), we have

$$\check{\sigma}^2_{n+h} := \exp\{\hat{\ln}(\sigma^2_{n+h})\} = \exp\left\{\omega + \sum_{k=0}^{\infty}\lambda_{d,k+h-1}\, g(Z_{n-k})\right\} \xrightarrow{h\to\infty} e^{\omega} = \exp\{\mathbb{E}(\ln(\sigma^2_0))\}. \tag{34}$$

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

$$\begin{aligned}\sigma^2_{n+h} &= \exp\big\{\mathbb{E}(\ln(\sigma^2_{n+h})|\mathcal{F}_n)\big\} + \big[\ln(\sigma^2_{n+h}) - \mathbb{E}(\ln(\sigma^2_{n+h})|\mathcal{F}_n)\big]\exp\big\{\mathbb{E}(\ln(\sigma^2_{n+h})|\mathcal{F}_n)\big\} \\ &\quad + \frac{1}{2}\big[\ln(\sigma^2_{n+h}) - \mathbb{E}(\ln(\sigma^2_{n+h})|\mathcal{F}_n)\big]^2\exp\big\{\mathbb{E}(\ln(\sigma^2_{n+h})|\mathcal{F}_n)\big\} + R_{n+h}, \quad \text{for all } h > 0, \end{aligned} \tag{35}$$

From expression (35), a natural choice is to define a h-step ahead predictor for σ^2_{n+h} as

$$\tilde{\sigma}^2_{n+h} := \exp\big\{\mathbb{E}(\ln(\sigma^2_{n+h})|\mathcal{F}_n)\big\} + \frac{1}{2}\mathbb{E}(\big[\ln(\sigma^2_{n+h}) - \mathbb{E}(\ln(\sigma^2_{n+h})|\mathcal{F}_n)\big]^2)\exp\big\{\mathbb{E}(\ln(\sigma^2_{n+h})|\mathcal{F}_n)\big\}, \tag{36}$$

for all $h > 0$.

From expressions (30), (31) and (36) one concludes that $\check{\sigma}^2_{n+h}$ and $\tilde{\sigma}^2_{n+h}$ are related through the equation

$$\tilde{\sigma}^2_{n+h} = \begin{cases} \exp\{\hat{\ln}(\sigma^2_{n+h})\} = \check{\sigma}^2_{n+h}, & \text{if} \quad h = 1; \\ \exp\{\hat{\ln}(\sigma^2_{n+h})\}\bigg(1 + \frac{1}{2}\sigma_g^2\sum_{k=0}^{h-2}\lambda^2_{d,k}\bigg) = \check{\sigma}^2_{n+h}\bigg(1 + \frac{1}{2}\sigma_g^2\sum_{k=0}^{h-2}\lambda^2_{d,k}\bigg), & \text{if} \quad h > 1. \end{cases} \tag{37}$$

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 (35), we easily conclude that, for all $h > 1$,

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

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

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

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 quasi-likelihood estimator and also the forecasting on 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). The set of parameters considered in this study is the same as in [12] and [13][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

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 value consider in this simulation study are given in Table 2. For each model we consider $re = 1,000$ replications, with sample size $N = 5,050$.

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	-	-	-	-	-

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 (14). For this simulation study the infinite sum (14) 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,

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

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 speed of the convergence 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 $k \in \{0; 10; 100; 1{,}000; 5{,}000; 10{,}000; 20{,}000; 50{,}000; 100{,}000\}$, for each simulated model Mi, $i \in \{1, \cdots, 6\}$. Note that, for $k \geq 5{,}000$, the coefficient values decrease slowly. We also report in Table 3 $Q_1(k)$ and $Q_2(k)$ values for the

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	0	10	100	1,000	5,000	10,000	25,000	50,000	100,000
M1 := FIEGARCH$(2,d,1)$									
$\lambda_{d,k}$	1	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$(0,d,4)$									
$\lambda_{d,k}$	1	-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$(0,d,1)$									
$\lambda_{d,k}$	1	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$(0,d,1)$									
$\lambda_{d,k}$	1	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$(1,d,1)$									
$\lambda_{d,k}$	1	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$(1,d,0)$									
$\lambda_{d,k}$	1	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

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, while $Q_2(k)$ converges to 1 faster for model M1 than for the other models.

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 Definition 1 and the equality in (14), 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 7. For parameter estimation and forecasting we shall consider sub-samples from these time series, with size $n \in \{2,000; 5,000\}$. The sub-samples of size $n = 2,000$ correspond to the last 2,000 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 $n = 2,000$ is the approximated size of the observed time series considered in [13]. The value $n = 5,000$ 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

[5] [12] present a Monte Carlo simulation study on risk measures estimation in time series derived from FIEGARCH process. [13] analyze a portfolio composed by 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.

$t \in \mathbb{Z}$, 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

$$\ln(\ell(\boldsymbol{\eta}; x_1, \cdots, x_n)) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\left[\ln(\sigma_t^2) + \frac{x_t^2}{\sigma_t^2}\right],. \tag{38}$$

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 variable $\ln(\sigma_t^2)$, for $t \in \{1, \cdots, n\}$. Then we use these estimated values to solve (38). 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 $\hat{\eta}_k$ denotes the estimate of η in the k-th replication, where $k \in \{1, \cdots, re\}$, $re = 1,000$ and η is any vector parameter given in Table 2. To access the performance of 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

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

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

4.4. Estimation Results

Table 4 summarizes the results on the parameter estimation procedure. Figures 6 - 11 present the kernel distribution of the parameter estimators for each considered model when $n \in \{2,000; 5,000\}$. These graphs help to illustrate the results presented in Table 4.

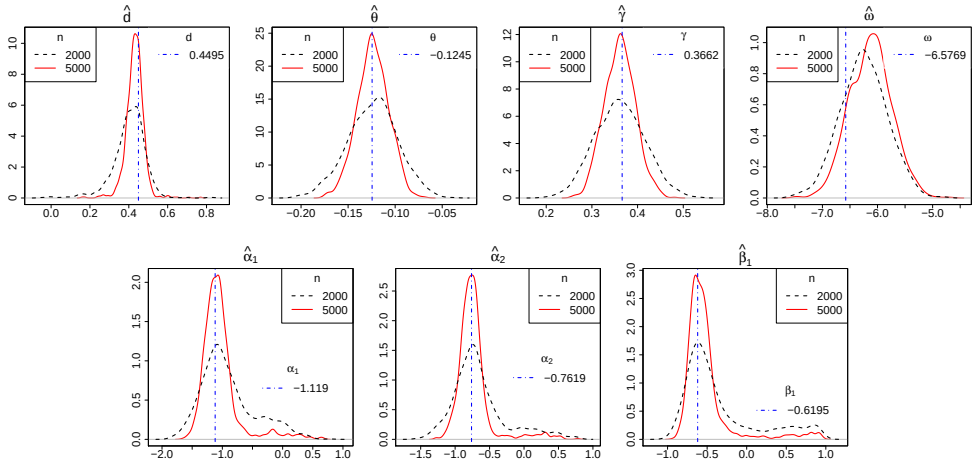


Figure 6: Kernel density function of the estimates for model M1, for $n \in \{2,000; 5,000\}$.

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

the parameters d (in particular, models M2, M3 and M4), α_i (model M2) 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 $n = 2,000$ and $n = 5,000$ are similar, one observes that, as expected, the observations tend to concentrate closer to the mean when $n = 5,000$.

Table 4: Estimation results for the simulated FIEGARCH models.

Sample Size (n)		2,000					5,000			
Parameter (η)	$\bar{\eta}$	sd	$bias$	mae	mse	$\bar{\eta}$	sd	$bias$	mae	mse
M1 := FIEGARCH$(2, d, 1)$; $re = 1,000$										
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$(0, d, 4)$; $re = 1,000$										
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$(0, d, 1)$; $re = 1,000$										
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$(0, d, 1)$; $re = 1,000$										
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$(1, d, 1)$; $re = 1,000$										
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$(1, d, 0)$; $re = 1,000$										
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

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

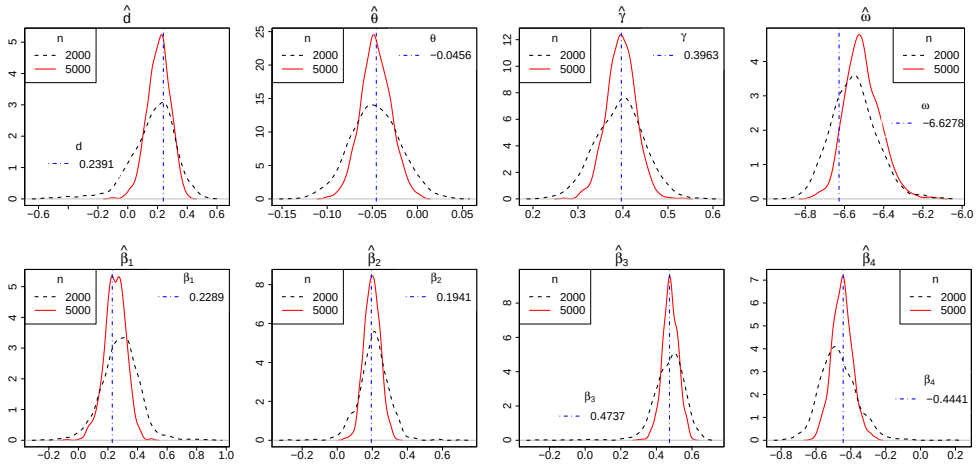


Figure 7: Kernel density function of the estimates for model M2, for $n \in \{2,000;\ 5,000\}$.

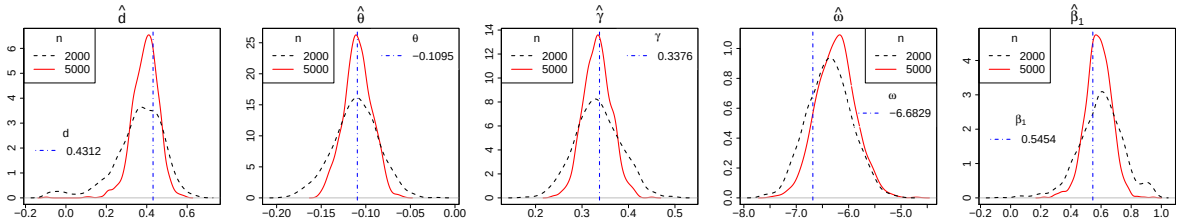


Figure 8: Kernel density function of the estimates for model M3, for $n \in \{2,000;\ 5,000\}$.

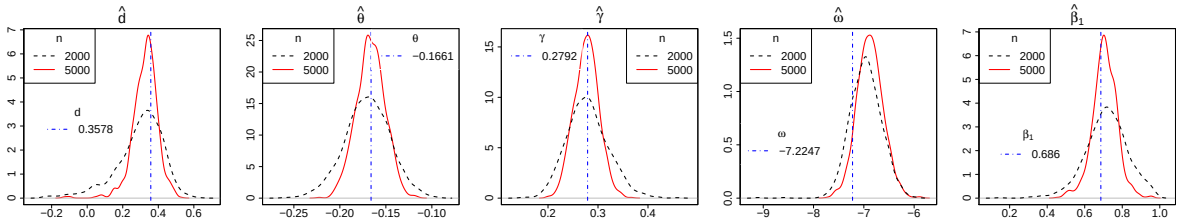


Figure 9: Kernel density function of the estimates for model M4, for $n \in \{2,000;\ 5,000\}$.

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 $q = 1$).

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 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 region for model M5 and, for model M1, not only $p = 2$ but also $d = 0.4495$, which implies a more complex

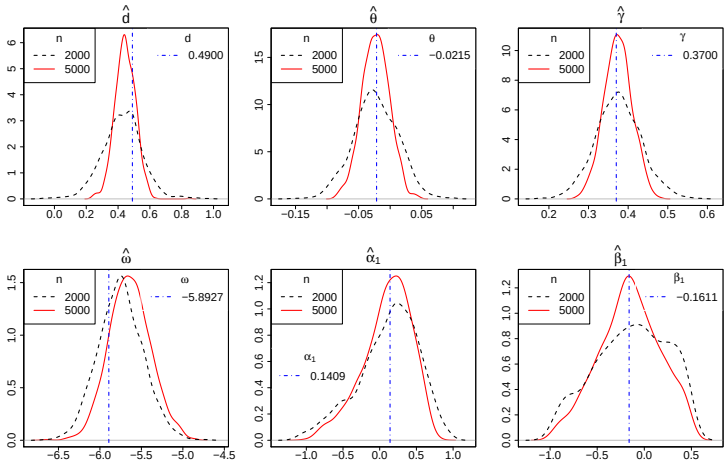


Figure 10: Kernel density function of the estimates for model M5, for $n \in \{2,000;\ 5,000\}$.

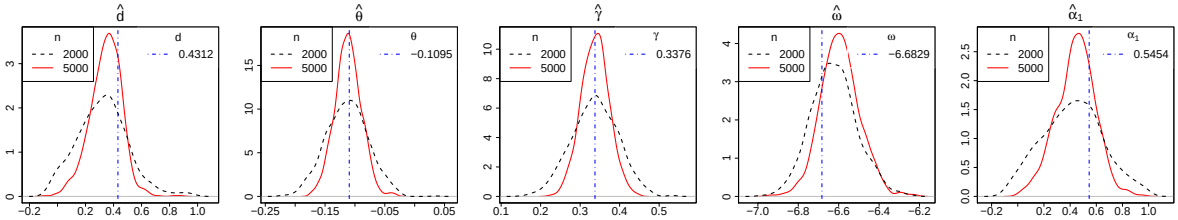


Figure 11: Kernel density function of the estimates for model M6, for $n \in \{2,000;\ 5,000\}$.

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 $n = 2,000$ 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.

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 $n \in \{2,000; 5,000\}$, we repeat steps **F1** - **F5** below.

F1: Replace the true parameters 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:

$$\sigma_1 = e^{\hat{\omega}0.5}; \quad z_1 = \frac{x_1}{\sigma_1}; \quad \sigma_t = \exp\left\{\frac{\hat{\omega}}{2} + \frac{1}{2}\sum_{k=0}^{n-1}\hat{\lambda}_{d,k}\left[\hat{\theta}z_{t-1-k} + \hat{\gamma}(|z_{t-1-k}| - \sqrt{2/\pi})\right]\right\} \quad \text{and} \quad z_t = \frac{x_t}{\sigma_t},$$

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 (30) and (36), 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 $1,000$ 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 $N = 5,000$ and sub-samples $n \in \{2,000; 5,000\}$. This table also reports the mean square error (mse) of forecast, defined as

$$mse(Y_{N+h}) := \frac{1}{re}\sum_{k=1}^{re} \left(Y_{N+h}^{(k)} - \check{Y}_{N+h}^{(k)}(n)\right)^2, \quad \text{for any } \ h \in \{1, \cdots, 5\} \ \text{ and } \ n \in \{2,000; 5,000\},$$

where $re = 1,000$ is the number of replications, Y_{N+h} denotes the true value of σ_{N+h}^2 (or X_{N+h}^2) and $\check{Y}_{N+h}^{(k)}(n)$ 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) conclude that,

- when we consider σ_{N+h}^2, 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

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

  where $\sigma_g^2 := \mathbb{E}([g(Z_0)]^2)$ is given in (11);
- when we consider X_{N+h}^2, the mse is usually high, if compared to the mean simulated and mean predicted values. Therefore, we conclude that $\tilde{X}_{n+h}^2 := \tilde{\sigma}_{n+h}^2$ is a poor estimator for X_{n+h}^2. This result is not a surprise since the main purpose of FIEGARCH models is to estimate the logarithm of the conditinal variance of the process and not the process itself;

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 $N = 5,000$ and $n \in \{2,000; 5,000\}$ 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.

n			2,000			5,000		
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$(2, d, 1)$; $re = 1,000$								
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$(0, d, 4)$; $re = 1,000$								
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$(0, d, 1)$; $re = 1,000$								
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$(0, d, 1)$; $re = 1,000$								
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$(1, d, 1)$; $re = 1,000$								
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$(0, d, 1)$; $re = 1,000$								
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

- as expected, in all cases, the models' forecasting performance improves as n increases. Notice, however, that the difference in the mse values, from $n = 2,000$ to $n = 5,000$, 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 1,000 replications for:

- the simulated values σ^2_{N+h} and X^2_{N+h} obtained from model Mi, for each $i \in \{1, \cdots, 6\}$, $N = 5,000$ 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 $N^* = N + h$, $N = 5,000$ 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 $n = 5,000$). The remaining predicted values $\{\check{\sigma}^2_{N+h-1}(1)\}_{h=2}^{50}$ are calculated by updating the forecasting origin from $N = 5,000$ 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 $N = 5,000$ (without update). For all graphs the size of the sub-sample used for parameter estimation and forecasting is $n = 5,000$.

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.

From Figure 12 we observe that, for all models, the means for the one-step ahead forecast values $\check{\sigma}^2_{N^*+1}$, show the same behavior over the time as the means for the true values $\sigma^2_{N^*+1}$, where $N^* = N + h - 1$, $N = 5,000$ 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 values for $g(Z_t)$ as t increases ($t > N$).

Regarding the h-step ahead predictors $\check{\sigma}^2_{i,n+h}$ 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 (34), $\check{\sigma}^2_{N+h}$ 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 (37),

$$\begin{aligned}\tilde{\sigma}^2_{N+h} := \check{\sigma}^2_{N+h}\left(1 + \frac{1}{2}\sigma_g^2 \sum_{k=0}^{h-2} \lambda^2_{d,k}\right) &\xrightarrow{h\to\infty} e^{\omega(i)}\left(1 + \frac{1}{2}\sigma_g^2(i) \sum_{k=0}^{\infty} \lambda^2_{d,k}(i)\right) \\ &\approx e^{\omega(i)}\left(1 + \frac{1}{2}\sigma_g^2(i) \sum_{k=0}^{m} \lambda^2_{d,k}(i)\right) := L_2(i), \end{aligned} \tag{39}$$

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 (39), 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 $\lim_{h\to\infty} \check{\sigma}^2_{N+h}$ 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 $\lim_{h\to\infty} \check{\sigma}^2_{N+h}$ (respectively, $L_2(i)$ and $\lim_{h\to\infty} \tilde{\sigma}^2_{N+h}$) is expected since the former one is calculated using the true parameter values while $\check{\sigma}^2_{N+h}$ is obtained by considering the estimates for the parameter.

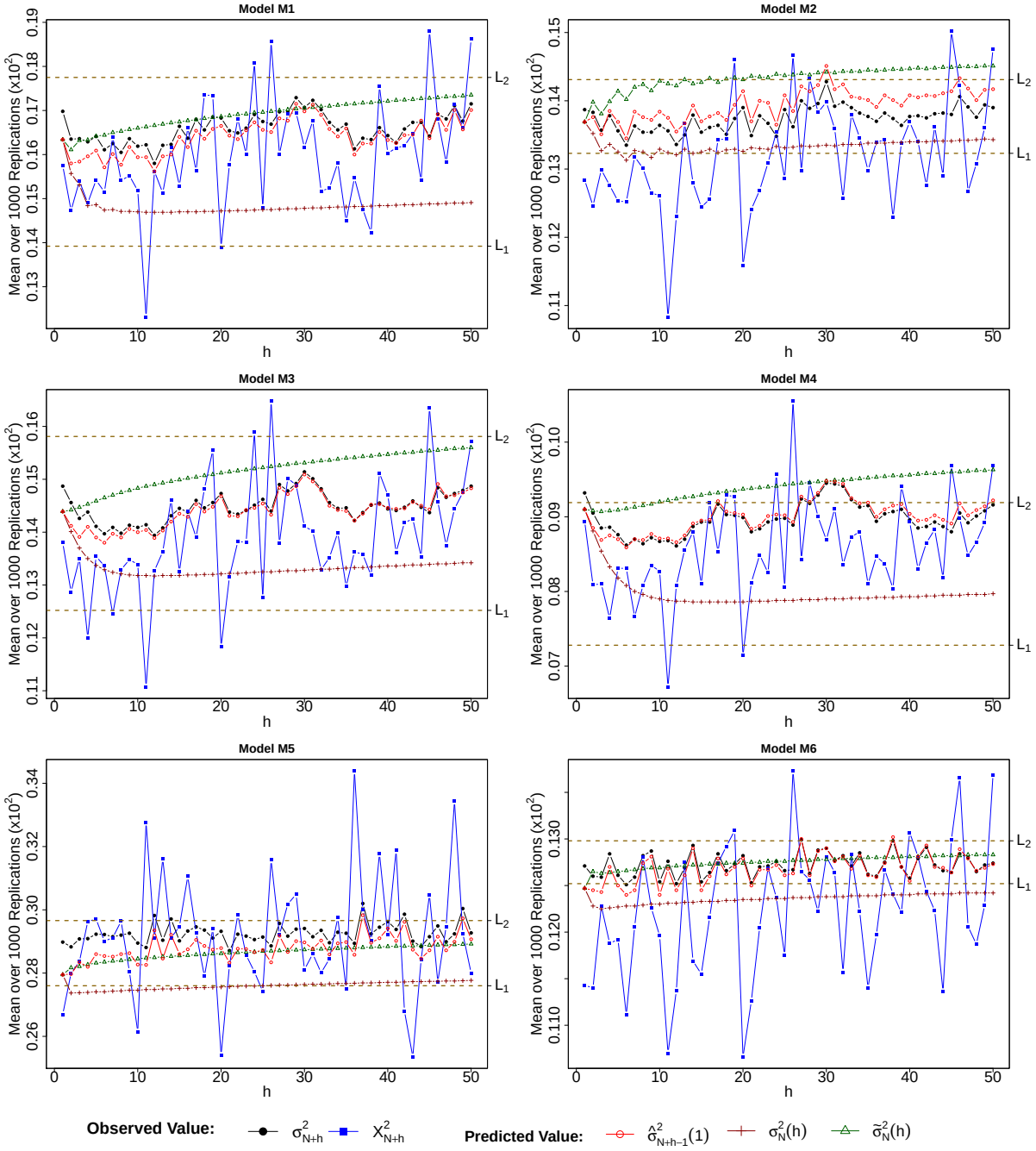
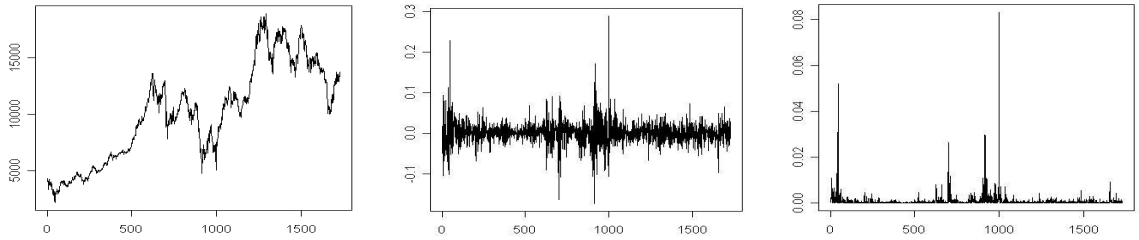


Figure 12: For each model Mi, $i \in \{1, \cdots, 6\}$: the simulated values for σ^2_{N+h}; the one-step ahead forecast $\check{\sigma}^2_{N^*+1} := \tilde{\sigma}^2_{N^*+1}$ (denoted in the graphs by $\hat{\sigma}^2_{N+h-1}(1)$), obtained by updating the forecasting origin to $N^* = N + h - 1$; the 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 $h \in \{1, \cdots, 50\}$, $N = 5,000$ and the size of the sub-sample used for parameter estimation and forecasting is $n = 5,000$. All values in the graphs correspond to the mean taken over 1,000 replications.

5. Analysis of an Observed Time Series

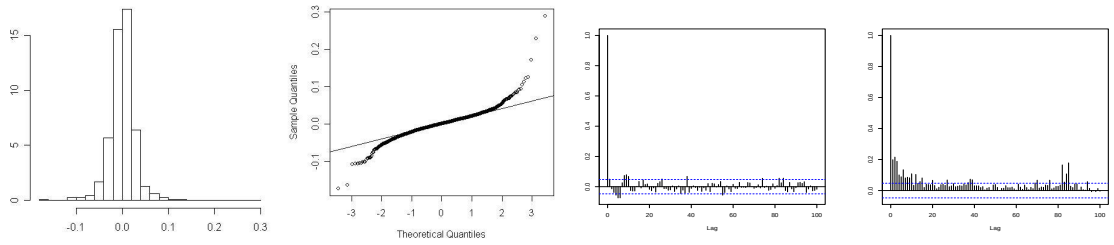
This section presents the analysis of the São Paulo Stock Exchange Index (Bovespa Index or IBovespa) log-return time series. We consider the FIEGARCH model, fully described in this paper, and we compare its forecasting performance with other ARCH-type models. The total number of observations for the 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 $t = 1,000$ (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 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 series presents almost no correlation, the sample autocorrelation of the square of the log-return series assumes high values for several lags, pointing to the existence of heteroskedasticity and possibly long memory. Notice that the periodogram of $\{\ln(r_t^2)\}_{t=1}^{1717}$, presented in Figure 4 (c), also indicates possibly long-memory in the conditional variance. Regarding the histogram and the QQ-Plot, we observe that the distribution of the log-return series seems approximately symmetric and leptokurtic.



(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 and (c) sample autocorrelation of the IBovespa log-return series and (d) sample autocorrelation of the square of the IBovespa log-return series.

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 [32]. Due to the magnitude of the data we multiply the time 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

[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 $r = 10$.

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.

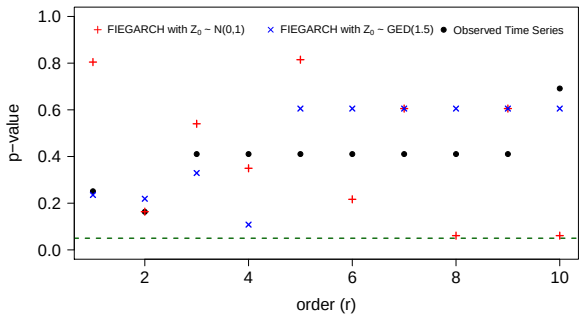


Figure 15: The p-values for the stationarity hypothesis test considering the moments of order $r \in \{1, \cdots, 10\}$. The dashed line corresponds to p-value = 0.05.

To analyze if the ergodicity property holds for the time series $\{r_t\}_{t=1}^{1717}$ we perform the test described in [33]. For comparison, we also apply this test to the simulated time series (only for sample size $n = 2,000$) 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 [33]. Moreover, for the simulated time series, the values in Table 7 correspond to the mean taken over 1,000 replications. Notice that the proportion of p-values smaller than 0.05 (equivalently, 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 [33] 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 take into account the long memory (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 eliminate 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 the 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. 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 memory 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)

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.

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 parameters values as model M4 considered in the simulated study in Section 4.

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

To fit a radial basis model to the data (no exogenous variables are considered) we assume that $\{r_t\}_{t\in\mathbb{Z}}$ can be written as (see [34], [35])

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

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 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

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

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 N neurons, for

some $N \in \mathbb{N}$, that is,

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

with $\boldsymbol{w}_1 = (a_1, \cdots, a_N, b_1, \cdots, b_N, \boldsymbol{c}_1, \cdots, \boldsymbol{c}_N)$, $\boldsymbol{w}_2 = (a_1^*, \cdots, a_N^*, b_1^*, \cdots, b_N^*, \boldsymbol{c}_1^*, \cdots, \boldsymbol{c}_N^*)$, $a_i, b_i, a_i^*, b_i^* \in \mathbb{R}$, $\boldsymbol{c}_i, \boldsymbol{c}_i^* \in \mathbb{R}^p$, $||\cdot||$ 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 $i \in \{1, \cdots, N\}$.

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

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

where $\hat{\boldsymbol{y}}_{n+h} = (\hat{r}_{n+h-1}, \cdots, \hat{r}_{n+h-1-p})$, 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

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

where, $e_{n+h} := \hat{r}_{n+h}^2 - r_{n+h}^2$, for $h \in \{1, \cdots, 20\}$ and $n = 1717$, is the h-step ahead forecast error. Note that, when considering the ARMA combined with ARCH-type models, from the ARMA(0,1) part of the models, $r_t = X_t - \theta_1 X_{t-1}$, where $X_t = \sigma_t Z_t$, for all $t \in \mathbb{Z}$. Since we define $\hat{r}_{t+h}^2 = \mathbb{E}(r_{t+h}^2 | \mathcal{F}_t)$ and σ_t^2 is $\mathcal{F}_{t-1}$-measurable, for all $t \in \mathbb{Z}$, by elementary calculations we conclude that, $\hat{r}_{n+1}^2 = \sigma_{n+1}^2 + \theta_1^2 X_n^2$ and $\hat{r}_{n+h}^2 = \hat{\sigma}_{n+h}^2 + \theta_1^2 \hat{\sigma}_{n+h-1}^2$, for all $h > 1$, with $\hat{\sigma}_{n+1}^2 = \sigma_{n+1}^2$. For EGARCH and FIEGARCH models, $\hat{\sigma}_{n+1}^2$ is replaced by $\tilde{\sigma}_{n+1}^2$, given in expression (36), and $\check{\sigma}_{n+h}^2 := \exp\{\hat{\ln}(\sigma_{n+h}^2)\}$, where $\hat{\ln}(\sigma_{n+h}^2)$ 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	$\hat{\sigma}_{t+h}^2$	$\hat{\sigma}_{t+h}^2$	$\hat{\sigma}_{t+h}^2$	$\tilde{\sigma}_{t+h}^2$	$\check{\sigma}_{t+h}^2$	$\tilde{\sigma}_{t+h}^2$	$\check{\sigma}_{t+h}^2$
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 square of the residuals from this model are still correlated and we use the model only for comparison. The ARMA(0,1)-ARCH(6) model performed similar 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 it also carries 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-memory 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 imply 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-memory 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 $N \in \{5, 10, \cdots, 45\}$ hidden neurons and $p \in \{1, 5, 10, 15\}$.

p	N	mae	mpe	max_{ae}	p	N	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 N for which the mpe value for the radial basis model is much smaller then 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 too close to zero.

6. Conclusions

Here we show 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 prove that if $\{X_t\}_{t\in\mathbb{Z}}$ is a FIEGARCH(p,d,q) process and $\mathbb{E}(\left[\ln(Z_0^2)\right]^2) < \infty$, 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 prove 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 $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ are given with their respective mean square error forecast. Since $\mathbb{E}(\sigma_{t+h}^2|\mathcal{F}_t)$ cannot be easily calculated for FIEGARCH models, we also discuss some alternative estimators for the h-step ahead forecast of σ_{t+h}^2, for all $h > 0$.

We present 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 are related to the real time series analyzed in [12]. Parameter estimation was performed by considering the well known quasi-likelihood method. We conclude that, given the complexity of FIEGARCH models, the quasi-likelihood method performs relatively well, which is 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 observe that the mean square error of forecast decreases as the sample size increases. However, while the conditional variance is well estimated, which is indicated by the small *mae* values, the estimator $\tilde{X}_{n+h}^2 := \tilde{\sigma}_{t+h}^2$, which is an approximation for $\hat{X}_{t+h}^2 := \mathbb{E}(X_{n+h}^2|\mathcal{F}_n) = \hat{\sigma}_{n+h}^2$, does not perform well in predicting X_{n+h}^2. This result is 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 present 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 conclude 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 too 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.

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] R.F. Engle, T. Bollerslev, Modeling the Persistence of Conditional Variances, Econometric Reviews, 5 (1986) 1-50.
[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] T. Mikosch, C. Stărică, Change of Structure in Financial Time Series, Long Range Dependence and the Garch Model, (1999), Preprint.

[8] L. Ming, Modeling Long Memory in Stock Market Volatility, Journal of Econometrics, 99(1) (2000) 139-171.
[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] 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.
[11] 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.
[12] T.S. Prass, S.R.C. Lopes, Risk Measures Estimation on FIEGARCH Processes, (2012) Working Paper.
[13] T.S. Prass, S.R.C. Lopes, VaR, Teste de Estresse e MaxLoss na Presença de Heteroscedasticidade e Longa Dependência na Volatilidade, (2012) Working Paper.
[14] 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.
[15] 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,.
[16] W. Palma, Long-Memory Time Series: Theory and Methods, John Wiley, New Jersey, 2007.
[17] 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.
[18] R. Lumsdaine, Asymptotic Properties of the Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) Models, Econometrica, 64 (1996) 575-596.
[19] I. Berkes, L. Horváth, P. Kokoszka, GARCH Processes: Structure and Estimation, Bernoulli, 9 (2003) 201-228.
[20] I. Berkes, L. Horváth, The Rate of Consistency of the Quasi-Maximum Likelihood Estimator, Statistics and Probability Letters, 61 (2003) 133-143.
[21] P. Hall, Q. Yao, Inference in ARCH and GARCH Models with Heavy-tailed Errors, Econometrica, 71 (2003) 285-317.
[22] 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.
[23] E. Zivot, J. Wang, Modeling Financial Time Series with S-PLUS, 2nd edition, Springer-Verlag, New York, 2005.
[24] S. Laurent, J.P Peters, G@RCH 4.0, Estimating and Forecasting ARCH models, Timberlake Consultants, 2005.
[25] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, Second Edition, Springer-Verlag, New York, 1991.
[26] P. Bloomfield, On Series Representations for Linear Predictors, The Annals of Probability, 13(1) (1985) 226-233.
[27] P. Bondon, W. Palma, A Class of Antipersistent Processes, Journal of Time Series Analysis, 28 (2007) 261-273.
[28] 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
[29] C.M. Hurvich, E. Moulines, P. Soulier, Estimating Long Memory in Volatility, Econometrica 73(4) (2005) 1283-1328.
[30] R. Durrett, Probability: Theory and Examples, Wadsworth & Brooks/Cole, Pacific Grove, 1991.
[31] 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.
[32] J. Grazzini, Analysis of the Emergent Properties: Stationarity and Ergodicity, Journal of Artificial Societies and Social Simulation 15(2) (2012) 7.
[33] I. Domowitz, M.A El-Gamal, A Consistent Nonparametric Test of Ergodicity for Time Series with Applications, Journal of Econometrics 102 (2001) 365-398.
[34] E. Giacomini, Neural Networks in Quantitative Finance. Master's Thesis, University of Berlin, (2003).
[35] 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.

Apêndice D

Artigo Lopes e Prass (2013b)

Autores: Sílvia R. C. Lopes e Taiane S. Prass

Título: Seasonal FIEGARCH Processes

Revista: Computational Statistics & Data Analysis, vol. 68, 262-295

Ano: 2013

Seasonal FIEGARCH Processes

Sílvia R.C. Lopes and Taiane S. Prass*

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

July 3, 2013

Abstract

Here we develop the theory of seasonal FIEGARCH processes, denoted by SFIEGARCH, establishing conditions for the existence, the invertibility, the stationarity and the ergodicity of these processes. We analyze their asymptotic dependence structure by means of the autocovariance and autocorrelation functions. We also present some properties regarding their spectral representation. All properties are illustrated through graphical examples and an application of SFIEGARCH models to describe the volatility of the S&P500 US stock index log-return time series in the period from December 13, 2004 to October 10, 2009 is provided.

Keywords. Long-Range Dependence, Volatility, Periodicity, FIEGARCH Process.

Mathematics Subject Classification (2010). 60G10, 62M10, 62M15, 91B84, 97M30.

Introduction

Introduced by Bollerslev and Mikkelsen (1996), FIEGARCH processes are one of the main models used to describe the volatility in financial time series. This class of models has not only the capability of capturing the asymmetry in the log-returns, as in the EGARCH models, but also it takes into account the characteristic of long memory in the volatility, as in the FIGARCH models, with the advantage of been weakly stationary. Lopes and Prass (2013) present a study on the theoretical properties of these processes, including results on the volatility forecast. The authors also analyze the finite sample performance of the quasi-likelihood estimator for four different FIEGARCH models and present the analysis of an observed time series. The simulated study presented by Lopes and Prass (2013) considers the same parameters values as the ones in the models adjusted to the observed time series considered in Prass and Lopes (2012, 2013).

More recently, economists have noticed that FIEGARCH models are not fully satisfactory, specially when modelling volatility of intra-daily financial returns. The main discovery is that volatility of high frequency financial time series shows long-range dependence merged with periodic behavior. According to Bordignon et al. (2007), these patterns, in the case of exchange rate returns, are generally attributed to different openings of European, Asian and North American markets superimposed each other. Similar patterns are found in stock markets, mainly due to the so-called time-of-day phenomena, such as market opening, closing operations, lunch-hour and overlapping effects. Once again, the focus is on the squared, log-squared and absolute returns. Periodic components are represented as marked peaks at some frequencies of the time series periodogram function and it can also

*Corresponding author. E-mail: taianeprass@gmail.com

be identified through a persistent cyclical behavior on the autocorrelation function with oscillations decaying very slowly. From the theoretical point of view, modelling and prediction of the volatility dynamics may be seriously affected if this empirical evidence is neglected.

Bordignon et al. (2007, 2009) introduced new GARCH-type models characterized by long memory behavior of periodic type. The generalized long memory GARCH (G-GARCH) introduces generalized periodic long-memory filters, based on Gegenbauer polynomials, into the equation describing the time-varying volatility of standard GARCH models. The periodic long-memory GARCH (PLM-GARCH) process represents a natural extension of the FIGARCH model proposed for modelling the volatility long-range persistence. Although periodic long memory versions of EGARCH (PLM-EGARCH) models were also considered in Bordignon et al. (2009) , we feel that there are several theoretical results related to these processes that were not yet explored. For instance, conditions for the existence, stationarity and ergodicity are yet to be established. Moreover, the autocovariance structure and the spectral representation of these processes are of extreme importance in both theoretical and practical point of view and hence, their study is an important matter.

Here we develop the theory of seasonal FIEGARCH processes, denoted by SFIEGARCH$(p,d,q)_s$, where p, d and q have the same meaning as in the so-called FIEGARCH(p,d,q) process and s is the length of the periodic component. This model is similar to the PLM-EGARCH process introduced by Bordignon et al. (2009) but, in the definition considered here, for any SFIEGARCH process $\{X_t\}_{t\in\mathbb{Z}}$, the process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is a SARFIMA one, where σ_t^2 is the conditional variance of X_t, for all $t \in \mathbb{Z}$. In particular, if $s = 1$, it is an ARFIMA(q,d,p) process (see Lopes, 2008). This result is useful for establishing whether the process $\{X_t\}_{t\in\mathbb{Z}}$ is well defined.

Results regarding the process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ are already known in the literature and can be found in Bisognin and Lopes (2009) and references therein. Moreover, for an SFIEGARCH process the sequence of random variables $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is not directly observable and we study its characteristics only to obtain the properties of the processes $\{X_t\}_{t\in\mathbb{Z}}$ and $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, which are the observable ones. In this work we extend the range of the parameter d for the invertibility and we present an alternative asymptotic expression for the autocovariance function $\gamma_{\ln(\sigma_t^2)}(\cdot)$. These results are useful to derive the exact and the asymptotic expressions for the autocovariance and spectral density functions of the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$.

The paper is organized as follows: in Section 1 we present the SFIEGARCH$(p,d,q)_s$ processes and we discuss the existence of a power series representation for the function $\lambda(z) = \frac{\alpha(z)}{\beta(z)}(1-z^s)^{-d}$ and the asymptotic behavior of the coefficients in this representation. A recurrence formula to calculate those coefficients is also provided. In Section 1 we also analyze the existence of the process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ and its invertibility property. This analysis is important to guarantee the existence of the process $\{X_t\}_{t\in\mathbb{Z}}$ itself. Section 2 is devoted to study the asymptotic dependence structure of both $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ and $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ processes, where $\{X_t\}_{t\in\mathbb{Z}}$ is an SFIEGARCH process. Section 3 presents the spectral representation of both processes $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ and $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$. Section 5 shows an application of SFIEGARCH models to describe the volatility of the S&P500 US stock index log-return time series in the period from December 13, 2004 to October 02, 2009. Section 6 presents the final conclusions. All proofs are presented in Appendix A.

1 SFIEGARCH Process

In this section we define the *Seasonal* FIEGARCH (SFIEGARCH) process which describes the volatility varying in time, volatility clusters (known as ARCH/GARCH effects), volatility periodic long-memory and asymmetry. Since the existence of a solution $\{X_t\}_{t\in\mathbb{Z}}$ for expression (1.1) depends on the existence of the stochastic process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ satisfying expression (1.2), we show that the random variable $\ln(\sigma_t^2)-\omega$ is finite with probability one, for all $t \in \mathbb{Z}$, if and only if $d < 0.5$. We show that $\{\ln(\sigma_t^2)-\omega\}_{t\in\mathbb{Z}}$ is an invertible process, with respect to $\{g(Z_t)\}_{t\in\mathbb{Z}}$, if and only if, $d \in (-1, 0.5)$, extending the range given in Bisognin and Lopes (2009). We also discuss the similarities between this model and the PLM-EGARCH model, introduced by Bordignon et al. (2009).

Hereafter, $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ denote, respectively, the floor and ceiling functions and $\mathbb{I}_A(\cdot)$ is the indicator function defined as $\mathbb{I}_A(z) = 1$, if $z \in A$, and 0, otherwise. Whenever $T = \mathbb{N}$ or $T = \mathbb{Z}$, we define $T^* := T\backslash\{0\}$. Throughout the paper, given two real/complex valued functions $f(\cdot)$ and $g(\cdot)$, $f(x) = O(g(x))$, means that $|f(x)| \leq c|g(x)|$, for some $c > 0$, as $x \to \infty$; $f(x) = o(g(x))$ means that $f(x)/g(x) \to 0$, as $x \to \infty$; $f(x) \sim g(x)$ means that $f(x)/g(x) \to 1$, as $x \to \infty$. 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 $x \geq x_0$. Similar definitions can be obtained upon replacing the functions $f(\cdot)$ and $g(\cdot)$ by sequences of real numbers $\{a_k\}_{k\in\mathbb{N}}$ and $\{b_k\}_{k\in\mathbb{N}}$ or if one considers any constant a or $-\infty$ instead of ∞.

Definition 1.1. Let $\{X_t\}_{t\in\mathbb{Z}}$ be the stochastic process defined by the expressions

$$X_t = \sigma_t Z_t, \tag{1.1}$$

$$\ln(\sigma_t^2) = \omega + \frac{\alpha(\mathcal{B})}{\beta(\mathcal{B})}(1-\mathcal{B}^s)^{-d} g(Z_{t-1}), \quad \text{for all } t \in \mathbb{Z}, \tag{1.2}$$

where $\{Z_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables, with zero mean and variance equal to one, $g(\cdot)$ is defined by

$$g(Z_t) = \theta Z_t + \gamma\big[|Z_t| - \mathbb{E}\big(|Z_t|\big)\big], \quad \text{with } \theta, \gamma \in \mathbb{R}, \quad \text{for all } t \in \mathbb{Z}, \tag{1.3}$$

$\omega \in \mathbb{R}$, $\mathcal{B}$ is the backward shift operator defined by $\mathcal{B}^{sk}(X_t) = X_{t-sk}$, for all $s, k \in \mathbb{N}$, $\alpha(\cdot)$ and $\beta(\cdot)$ are, respectively, polynomials of order p and q, with no common roots defined by

$$\alpha(z) = \sum_{i=0}^{p}(-\alpha_i)z^i \quad \text{and} \quad \beta(z) = \sum_{j=0}^{q}(-\beta_j)z^j, \tag{1.4}$$

with $\alpha_0 = -1 = \beta_0$, and $\beta(z) \neq 0$ in the closed disk $\{z : |z| \leq 1\}$, $d \in \mathbb{R}$ is the differencing parameter, $s \in \mathbb{N}^*$ is the length of the periodic component, $(1-\mathcal{B}^s)^{-d}$ is the seasonal difference operator, defined by its Maclaurin series expansion, namely,

$$(1-\mathcal{B}^s)^{-d} = \sum_{k=0}^{\infty}\frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)}(\mathcal{B}^s)^k := \sum_{k=0}^{\infty}\delta_{-d,k}\mathcal{B}^{sk} := \sum_{k=0}^{\infty}\pi_{d,k}\mathcal{B}^k, \tag{1.5}$$

where $\Gamma(\cdot)$ is the Gamma function, $\pi_{d,k} := 0$, if $k/s \notin \mathbb{N}$, and $\pi_{d,sj} = \delta_{-d,j} := \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)}$, for all $j \in \mathbb{N}$. Then, $\{X_t\}_{t\in\mathbb{Z}}$ is a seasonal FIEGARCH process, with seasonal period s and differencing parameter d, denoted by SFIEGARCH$(p, d, q)_s$.

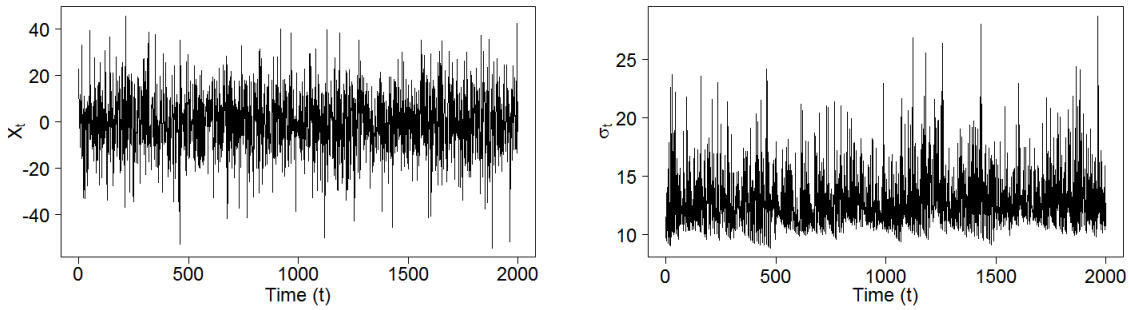
Remark 1.1. The assumptions that $\beta(z) \neq 0$ in the closed disk $\{z : |z| \leq 1\}$ and that $\alpha(\cdot)$ and $\beta(\cdot)$ have no common roots guarantee that the operator $\frac{\alpha(\mathcal{B})}{\beta(\mathcal{B})}$ is well defined.

Example 1.1. Figure 1.1 presents a simulated SFIEGARCH$(0, d, 0)_s$ time series $\{X_t\}_{t=1}^n$ and its conditional standard deviation $\{\sigma_t\}_{t=1}^n$, defined by expressions (1.1) and (1.2). For these graphs, $Z_0 \sim \mathcal{N}(0, 1)$, $\omega = 5.0$, $\theta = -0.25$, $\gamma = 0.24$, $d = 0.35$ and $s = 6$.

Remark 1.2. In this work we consider the case where the conditional variance σ_t^2 is defined through expression (1.2), with $\mathbb{E}(Z_0) = 0$, $\mathrm{Var}(Z_0) = \mathbb{E}(Z_0^2) = 1$ and the function $g(\cdot)$ defined by expression (1.3). However the results presented here can be easily extended if one considers $\mathrm{Var}(Z_0) = \sigma^2 \neq 1$ and replaces $g(\cdot)$ by any measurable function satisfying $\mathrm{Var}(g(Z_0)) < \infty$.

Observe that the series expansion of the operator $(1-\mathcal{B}^s)^d$ is obtained upon replacing $-d$ by d in expression (1.5). Moreover, when $d \in \mathbb{N}$, $(1-\mathcal{B}^s)^d$ is merely the seasonal difference operator $1 - B^s$ iterated d times. Thus, one can easily see that an equivalent definition for SFIEGARCH process is given if one replaces expression (1.2) by

$$\beta(\mathcal{B})(1-\mathcal{B}^s)^d(\ln(\sigma_t^2) - \omega) = \alpha(\mathcal{B})g(Z_{t-1}), \quad \text{for all } t \in \mathbb{Z}. \tag{1.6}$$



(a) (b)

Figure 1.1: Samples from an SFIEGARCH$(0, d, 0)_s$ processes, with $n = 2000$ observations, considering $Z_0 \sim \mathcal{N}(0, 1)$, $\omega = 5.0$, $\theta = -0.25$, $\gamma = 0.24$, $d = 0.35$ and $s = 6$. Panel (a) shows the time series $\{X_t\}_{t=1}^n$. Panel (b) presents the time series $\{\sigma_t\}_{t=1}^n$, where σ_t is the conditional standard deviation of X_t, for all $t \in \{1, \cdots, n\}$.

This expression is similar to the one in the definition of the PLM-EGARCH process, presented by Bordignon et al. (2009). For a PLM-EGARCH(p, m, d, q, s), the conditional variance σ_t^2 of X_t is defined through the equation

$$(1 - \mathcal{B}^s)^d \phi(\mathcal{B})\big(\ln(\sigma_t^2) - \omega\big) = a(\mathcal{B})Z_t + c(\mathcal{B})\big(|Z_t| - \mathbb{E}(|Z_t|)\big), \quad \text{for all } t \in \mathbb{Z}, \tag{1.7}$$

where $a(z) = \sum_{k=1}^p a_k z^k$ and $c(z) = \sum_{l=1}^m c_l z^l$ are polynomials of order p and m, respectively, $\phi(z) = \sum_{j=0}^{q-s} \phi_j z^j$ is a polynomial of order $q-s$, which satisfies $(1-\mathcal{B}^s)^d\phi(\mathcal{B}) = 1-b(\mathcal{B})$, where $b(z) = \sum_{i=1}^q b_i z^i$ is a polynomial of order q.

Notice that in the PLM-EGARCH, the polynomials $a(\cdot)$ and $c(\cdot)$ do not necessarily have the same order. Also, it is easy to see that, by setting

$$\begin{aligned} \phi_i &:= -\beta_i, \quad \text{for all } i = 0, \cdots, q, \\ a_{j+1} &:= -\theta\alpha_j \quad \text{and} \quad c_{j+1} := -\gamma\alpha_j, \quad \text{for all } j = 0, \cdots, p, \end{aligned} \tag{1.8}$$

one can rewrite the right hand side of expression (1.6) as the right hand side of (1.7). Under this point of view, the PLM-EGARCH model seems more general than the SFIEGARCH one. On the other hand, the left hand side of expression (1.6) is more general then the left hand side of (1.7). This is so because in the SFIEGARCH model, no restriction is made in the order of the product $\beta(z)(1 - z^s)^d$, allowing for the parameter d to be fractional.

Remark 1.3. It is immediate that, if $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is a stationary process with finite mean, then $\omega = \mathbb{E}(\ln(\sigma_t^2))$, for all $t \in \mathbb{Z}$. Also, if $d = 0$, we have the EGARCH(p, q) model proposed by Nelson (1991) and, if $s = 1$, we have the FIEGARCH(p, d, q) process defined by Bollerslev and Mikkelsen (1996). A study on the theoretical properties of FIEGARCH(p, d, q) process are presented in Lopes and Prass (2013).

From Definition 1.1, one easily concludes that the existence of the stochastic process $\{X_t\}_{t\in\mathbb{Z}}$ depends on the existence of the stochastic process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ which satisfies equation (1.2). The existence of a solution for equation (1.2) is discussed in the sequel.

From now on, let $\lambda(\cdot)$ be the polynomial defined as

$$\lambda(z) := \frac{\alpha(z)}{\beta(z)}(1 - z^s)^{-d} = \sum_{k=0}^{\infty} \lambda_{d,k} z^k, \quad |z| < 1, \tag{1.9}$$

where $\alpha(\cdot)$ and $\beta(\cdot)$ are defined in (1.4). Notice that, by definition, $\beta(z)$ has no roots in the closed disk $\{z : |z| \leq 1\}$, and $\alpha(\cdot)$ and $\beta(\cdot)$ have no common roots. Therefore, the function $\lambda(z)$ is analytic in

the open disc $\{z : |z| < 1\}$ and, if $d \leq 0$, in the closed disk $\{z : |z| \leq 1\}$. So, it has a unique power series representation and the operator given in (1.2) can be rewritten as $\frac{\alpha(\mathcal{B})}{\beta(\mathcal{B})}(1-\mathcal{B}^s)^{-d} = \sum_{k=0}^{\infty} \lambda_{d,k}\mathcal{B}^k = \lambda(\mathcal{B})$. This representation is more convenient and will be used from now on. In the following, we analyze the asymptotic behavior of the coefficients $\lambda_{d,k}$, for all $k \in \mathbb{Z}$, defined by expression (1.9). This result is fundamental for proving the results regarding the existence, invertibility, stationarity and ergodicity of SFIEGARCH processes.

It is immediate that, if $p \geq 0$ and $q = 0$, one can rewrite (1.9) as,

$$\lambda(z) = \begin{cases} (1-z^s)^{-d} = \sum_{k=0}^{\infty} \pi_{d,k} z^k, & \text{if } p = 0 = q; \\ \alpha(z)(1-z^s)^{-d} = \sum_{k=0}^{\infty} \Big[\sum_{i=0}^{\min\{p,k\}} (-\alpha_i \pi_{d,k-i}) \Big] z^k, & \text{if } p > 0 \text{ and } q = 0. \end{cases}$$

Thus, for all $r \in \{0, \cdots, s-1\}$ and all $k \in \mathbb{N}$,

$$\lambda_{d,sk+r} = \pi_{d,sk+r}, \quad \text{if} \quad p = 0 = q,$$

and, whenever $p > 0$ and $q = 0$,

$$\lambda_{d,sk+r} = -\sum_{j=0}^{\min\{p,sk+r\}} \alpha_j \pi_{d,sk+r-j} = \begin{cases} 0, & \text{if } p < r; \\ -\sum_{j=0}^{\min\{\lfloor \frac{p-r}{s} \rfloor, k\}} \alpha_{sj+r} \pi_{d,sk-sj}, & \text{otherwise.} \end{cases}$$

Consequently, given $\mathsf{r} > 0$,

$$\sum_{k=0}^{\infty} |\lambda_{d,k}|^{\mathsf{r}} < \infty \quad \text{if and only if} \quad \sum_{k=0}^{\infty} |\pi_{d,k}|^{\mathsf{r}} < \infty.$$

Theorem 1.1 bellow shows that this result also holds in the general case $p \geq 0$ and $q > 0$. The proofs of all results stated in this work are given in the Appendix.

Remark 1.4. By Stirling's formula and from lemma 3.1 in Kokoszka and Taqqu (1995), one easily concludes that

$$\pi_{d,sk} := \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)} = \frac{1}{\Gamma(d)k^{1-d}} + O(k^{d-2}) \sim \frac{1}{\Gamma(d)k^{1-d}}, \quad \text{as } k \to \infty. \tag{1.10}$$

Since (integral convergence test)

$$\sum_{k=1}^{m} k^{-\mathsf{r}(d-1)} \leq \int_1^m x^{-\mathsf{r}(d-1)} \mathrm{d}x = \frac{1}{1-r(d-1)} \Big|_1^m, \quad \text{for any } \mathsf{r} > 0,$$

converges to a finite constant as $m \to \infty$ if and only if $1-(1-d)\mathsf{r} < 0$, it follows that $\sum_{k=0}^{\infty} |\pi_{d,k}|^{\mathsf{r}} < \infty$ if and only if $(1-d)\mathsf{r} > 1$.

Theorem 1.1. *Let $\lambda_{d,k}$, for $k \in \mathbb{N}$, be the coefficients of the polynomial $\lambda(\cdot)$, given by expression (1.9). Let $f(z) := \frac{\alpha(z)}{\beta(z)} = \sum_{k=0}^{\infty} f_k z^k$. Then, for each $r \in \{0, \cdots, s-1\}$ and any $\nu > 0$, one has*

$$\lambda_{d,sk+r} = \pi_{d,sk}\mathscr{K}(sk+r) + o(k^{-\nu}), \quad \text{as } k \to \infty, \tag{1.11}$$

where $\mathscr{K}(\cdot)$ satisfies $\lim_{k\to\infty} \sum_{r=0}^{s-1} \mathscr{K}(sk+r) = \frac{\alpha(1)}{\beta(1)}$. Thus, $\sum_{r=0}^{s-1} \lambda_{d,sk+r} \sim \pi_{sk} \frac{\alpha(1)}{\beta(1)}$, as $k \to \infty$.

Theorem 1.2 presents an alternative asymptotic representation for the coefficients $\lambda_{d,k}$, as k goes to infinity. While expression (1.11) is more convenient for proving the asymptotic behavior of $\gamma_{\ln(X_t^2)}(\cdot)$ (see Theorem 2.3), expression (1.12) is useful for simulation purpose (see Remark 1.5).

Theorem 1.2. *Let $\lambda_{d,k}$, for $k \in \mathbb{N}$, be the coefficients of the polynomial $\lambda(\cdot)$, given by expression (1.9), with $d < 0.5$. Then, for each $r \in \{0, \cdots, s-1\}$, one can write*

$$\begin{aligned}\lambda_{d,sk+r} &= \frac{1}{\Gamma(d)k^{1-d}}\frac{\alpha(1)}{\beta(1)} - \frac{1}{\Gamma(d)k^{1-d}}\sum_{j=0}^{sk+r} f_j \mathbb{I}_{\mathbb{R}\setminus\mathbb{N}}\Big(\tfrac{|j-r|}{s}\Big) + O(k^{d-2}) \\ &= O(k^{d-1}) + O(k^{d-1})\mathbb{I}_{\mathbb{N}\setminus\{0,1\}}(s) + O(k^{d-2}), \qquad \text{as } k \to \infty,\end{aligned} \tag{1.12}$$

where $f(z) = \frac{\alpha(z)}{\beta(z)} = \sum_{k=0}^{\infty} f_k z^k$.

Remark 1.5. From Theorem 1.1 one observes that $\lambda_{d,sk+r}$ behaves asymptotically as the coefficient $\pi_{d,sk}$, as k goes to infinity. This property is very useful to prove the results stated in Section 2.1. On the other hand, from Theorem 1.2,

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

This approximation has a closed formula which also takes into account the magnitude of $\frac{\alpha(1)}{\beta(1)}$. Although this is a rough approximation, it can be used to estimate a truncation point m for $\lambda(\cdot)$ in Monte Carlo simulation studies. That is, given $\varepsilon > 0$, if one chooses $m \gg s\Big[\frac{1}{|\Gamma(d)|\varepsilon}\frac{\alpha(1)}{\beta(1)}\Big]^{\frac{1}{1-d}}$, one gets $|\lambda_{d,m}| < \varepsilon$.

In the following proposition we present a recurrence formula to calculate the coefficients $\lambda_{d,k}$, for all $k \in \mathbb{N}$. This recurrence formula is very useful in Monte Carlo simulation studies.

Proposition 1.1. *Let $\lambda(\cdot)$ be the polynomial defined by (1.9). Suppose $\alpha(\cdot)$ and $\beta(\cdot)$ have no common roots and $\beta(z) \neq 0$ in the closed disk $\{z : |z| \leq 1\}$. Then, the coefficients $\lambda_{d,k}$, for all $k \in \mathbb{N}$, are given by*

$$\lambda_{d,k} = \begin{cases} 1, & \text{if} \quad k = 0, \\ -\alpha_k + \displaystyle\sum_{i=0}^{k-1} \lambda_{d,i}\Big(\sum_{j=0}^{(k-i)\wedge q} \delta^*_{d,\frac{k-i-j}{s}}\beta_j\Big), & \text{if} \quad k \leq p; \\ \displaystyle\sum_{i=0}^{k-1} \lambda_{d,i}\Big(\sum_{j=0}^{(k-i)\wedge q} \delta^*_{d,\frac{k-i-j}{s}}\beta_j\Big), & \text{if} \quad k > p, \end{cases}$$

where $(k-i)\wedge q = \min\{k-i, q\}$ and, by definition,

$$\delta^*_{d,m} = \begin{cases} \delta_{d,m}, & \text{if} \quad m \in \mathbb{N}, \\ 0, & \text{if} \quad m \notin \mathbb{N}, \end{cases} \tag{1.13}$$

with $\delta_{d,m}$, for all $m \in \mathbb{N}$, defined in (1.5).

The following proposition presents some properties of the stochastic process $\{g(Z_t)\}_{t\in\mathbb{Z}}$. Although the proof is straightforward and follows immediately from the fact that $\{Z_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables, the proposition is fundamental to establish the result in Lemma 1.1, Corollary 1.1 and Theorem 1.3.

Proposition 1.2. *Let $g(\cdot)$ be the function defined by (1.3) and $\{Z_t\}_{t\in\mathbb{Z}}$ be a sequence of i.i.d. random variables, with zero mean and variance equal to one. Then, $\{g(Z_t)\}_{t\in\mathbb{Z}}$ is a white noise process with i.i.d. random variables and its variance σ_g^2 is given by*

$$\sigma_g^2 = \theta^2 + \gamma^2 - [\gamma\mathbb{E}(|Z_0|)]^2 + 2\theta\gamma\mathbb{E}(Z_0|Z_0|). \tag{1.14}$$

Moreover, the stochastic process $\{g(Z_t)\}_{t\in\mathbb{Z}}$ is stationary (weak and strictly) and ergodic.

Remark 1.6. Henceforth, GED shall stands for the so-called Generalized Error Distribution (see Nelson, 1991). 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.

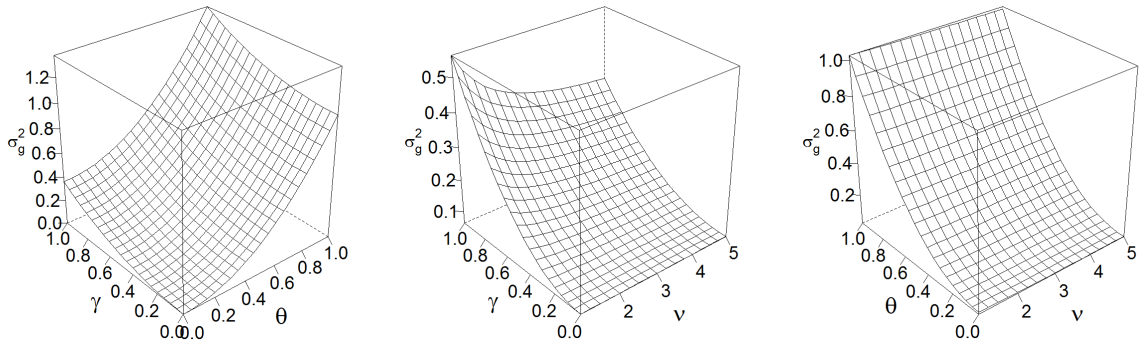
Remark 1.7. If the random variable Z_0 is symmetric, then $\mathbb{E}(Z_0|Z_0|) = 0$ and (1.14) is replaced by $\sigma_g^2 = \theta^2 + \gamma^2(1 - \mathbb{E}(|Z_0|)^2)$. Besides, if $Z_0 \sim \text{GED}(\nu)$, then

$$\sigma_g^2 = \theta^2 + \gamma^2 \left(1 - \frac{\Gamma(2/\nu)}{\sqrt{\Gamma(1/\nu)\Gamma(3/\nu)}}\right), \quad \text{for any} \quad \nu > 1.$$

If $\nu = 2$, one has the Gaussian case, that is, $Z_0 \sim \mathcal{N}(0,1)$ and $\sigma_g^2 = \theta^2 + \gamma^2(1 - 2/\pi)$.

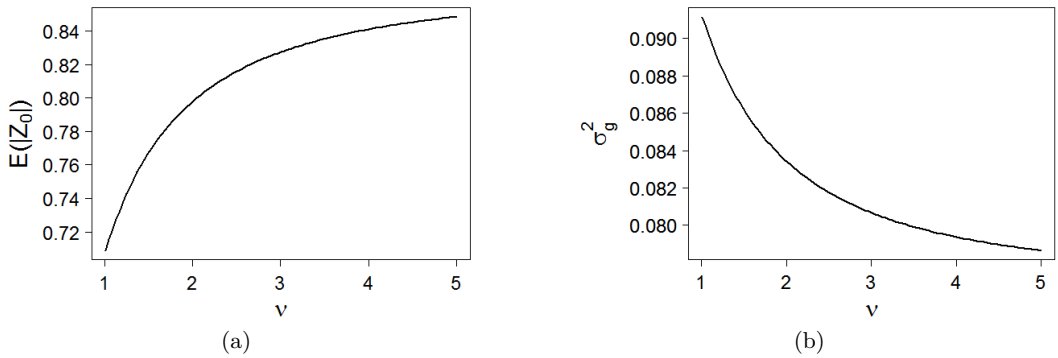
Example 1.2. Figure 1.2 (a) shows the graphs of σ_g^2 as a function of θ and γ, when $Z_0 \sim \mathcal{N}(0,1)$. Figures 1.2 (b) and (c) consider $Z_0 \sim \text{GED}(\nu)$, with $\nu > 1$ ($\nu = 2$ corresponds to the Gaussian case), and present the graphs of σ_g^2, respectively, as a function of γ and ν, for $\theta = 0.25$, and as a function of θ and ν, for $\gamma = 0.24$. Notice that for these graphs, Z_0 is a symmetric random variable, so we only consider positive values of θ and γ.

From Figure 1.2 one observes that, although σ_g^2 is increasing in both θ and γ, for any $\nu > 1$, it varies faster as θ increases than when γ does (notice the scales for the ordinate axis). Moreover, for each γ and θ fixed, σ_g^2 is decreasing in ν. This is expected since $\mathbb{E}(|Z_0|)$ is increasing for $\nu \in (1, 5.00]$. This fact is illustrated in Figure 1.3 where the graphs of $\mathbb{E}(|Z_0|)$ and σ_g^2 as functions of ν are presented. In Figure 1.3 (b) we fixed $\theta = 0.25$ and $\gamma = 0.24$. From this figure it is easy to see that σ_g^2 is indeed decreasing in ν.



(a) (b) (c)

Figure 1.2: This figure presents the behavior of σ_g^2, the variance of the process $\{g(Z_t)\}_{t\in\mathbb{Z}}$, as a function of parameters θ, γ and ν, when $Z_0 \sim \text{GED}(\nu)$. Panel (a) considers σ_g^2 as a function of θ and γ when $\nu = 2$, that is, when $Z_0 \sim \mathcal{N}(0,1)$. Panel (b) shows σ_g^2 as a function of γ and ν when $\theta = 0.25$. Panel (c) considers σ_g^2 as a function of θ and ν when $\gamma = 0.24$.



(a) (b)

Figure 1.3: This figure considers $Z_0 \sim \text{GED}(\nu)$. Panel (a) presents the graph of $\mathbb{E}(|Z_0|)$ as a function of ν. Panel (b) shows the the graph of $\mathbb{E}(g(Z_0)^2) = \sigma_g^2$, the variance of the process $\{g(Z_t)\}_{t\in\mathbb{Z}}$, as a function of ν when $\theta = 0.25$ and $\gamma = 0.24$.

Lemma 1.1 provides the necessary and sufficient conditions for the existence of the process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$.

Lemma 1.1. *Suppose that $\{Z_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables with zero mean and variance equal to one. Let $\{g(Z_t)\}_{t\in\mathbb{Z}}$ be the process defined by (1.3), ω be a real constant and $\lambda(\cdot)$ be the operator defined by (1.9). Define*

$$\ln(\sigma_t^2) - \omega = \sum_{k=0}^{\infty} \lambda_{d,k} g(Z_{t-1-k}), \quad \text{for all } t \in \mathbb{Z}. \tag{1.15}$$

Thus, the series (1.15) is well defined and converges a.s. if and only if $d < 0.5$. Moreover, the series (1.15) converges absolutely a.s. for $d \leq 0$.

Corollaries 1.1 and 1.2 follow immediately from Lemma 1.1 and show, respectively, that the process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is a causal SARFIMA process and that X_t is finite with probability one, for all $t \in \mathbb{Z}$. We emphasize that, causality and invertibility are defined in terms of convergence in the linear space L^2 (see Palma, 2007) and not in the linear space L^1 (as in Brockwell and Davis, 1991). The same approach is considered in Bloomfield (1985) and Bondon and Palma (2007).

Corollary 1.1. *Let $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ be the stochastic process defined by expression (1.2), with $d < 0.5$. Then, $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is a causal* SARFIMA$(p, 0, q) \times (0, D, 0)_s$ *process, with $D = d$.*

Corollary 1.2. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be an* SFIEGARCH$(p, d, q)_s$ *process, with $d < 0.5$. Then, the random variable X_t is finite with probability one, for all $t \in \mathbb{Z}$.*

Bisognin and Lopes (2009) show that a SARFIMA$(p, d, q) \times (P, D, Q)_s$ process is invertible whenever $|d + D| < 0.5$. Moreover, it is usually stated that an ARFIMA(p, d, q) process is invertible for $|d| < 0.5$ (see for instance Hosking, 1981; Brockwell and Davis, 1991). However, Bloomfield (1985) proves that, for an ARFIMA$(0, d, 0)$, this range can be extended to $d \in (-1, 0.5)$. Bondon and Palma (2007) show that this result actually holds for any ARFIMA(p, d, q). Although the spectral density function of $\{\ln(\sigma_t^2) - \omega\}_{t\in\mathbb{Z}}$ does not satisfy all conditions imposed in Bondon and Palma (2007), with some modifications in their proof, we show here that the results still holds for a SARFIMA$(p, 0, q) \times (0, D, 0)_s$ process (see Theorem 1.3).

Theorem 1.3. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be an* SFIEGARCH$(p, d, q)_s$*, defined by (1.1) and (1.2), with γ and θ, not both equal to zero. Assume that $\alpha(z) \neq 0$, for all $|z| \leq 1$. Let $Y_t := \ln(\sigma_t^2) - \omega$, for all $t \in \mathbb{Z}$. Then,*

$$\lim_{m\to\infty} \mathbb{E}\left(\left| \sum_{k=0}^{m} \tilde{\lambda}_{d,k} Y_{t-k} - g(Z_{t-1}) \right|^{\mathfrak{p}} \right) = 0, \quad \text{for all} \quad 0 < \mathfrak{p} \leq 2,$$

if and only if $d \in (-1, 0.5)$, where $\{\tilde{\lambda}_{d,k}\}_{k\in\mathbb{N}}$ is the sequence of coefficients in the series expansion of $\tilde{\lambda}(z) := \lambda^{-1}(z)$, for $|z| < 1$, that is,

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

2 Stationarity and Ergodicity

Here we show that for any SFIEGARCH$(p, d, q)_s$, with θ and γ not both equal to zero and $d < 0.5$, the processes $\{X_t\}_{t\in\mathbb{Z}}$ and $\{\sigma_t^2\}_{t\in\mathbb{Z}}$ are strictly stationary and ergodic processes. We also prove that if $\mathbb{E}([\ln(Z_0^2)]^2) < \infty$, the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is well defined and it is stationary (weakly and strictly) and ergodic. Weakly stationarity of the processes $\{X_t\}_{t\in\mathbb{Z}}$ and $\{\sigma_t^2\}_{t\in\mathbb{Z}}$ is also discussed. For any stationary SFIEGARCH process, we give the expressions for the autocovariance and autocorrelation functions of $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ and $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ and study their relation and asymptotic behavior. We also provide expression for the asymmetry (also known as skewness) and kurtosis measures for any stationary SFIEGARCH process $\{X_t\}_{t\in\mathbb{Z}}$.

Lemma 2.1 presents the conditions for the stationarity of the SARFIMA process $\{\ln(\sigma_t^2)-\omega\}_{t\in\mathbb{Z}}$. This lemma is useful to prove Theorem 2.1 that presents results on the stationarity of the processes $\{X_t\}_{t\in\mathbb{Z}}$ and $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$.

Lemma 2.1. *Let $\{\ln(\sigma_t^2)-\omega\}_{t\in\mathbb{Z}}$ be defined by (1.2), with γ and θ not both equal to zero. If $d<0.5$, the stochastic process $\{\ln(\sigma_t^2)-\omega\}_{t\in\mathbb{Z}}$ is stationary (strictly and weakly) and ergodic.*

Corollary 2.1. *If $d<0.5$, the stochastic process $\{\sigma_t^2\}_{t\in\mathbb{Z}}$ is strictly stationary and ergodic.*

Theorem 2.1 shows that both processes $\{X_t\}_{t\in\mathbb{Z}}$ and $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ are strictly stationary and ergodic, whenever $d<0.5$ and $\mathbb{E}(|\ln(Z_0)|)<\infty$, regardless the distribution of the random variable Z_0. This theorem also provides the necessary condition for $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ to be a weakly stationary process.

Theorem 2.1. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be an* SFIEGARCH$(p,d,q)_s$*, defined by (1.1) and (1.2). Suppose that γ and θ, given in (1.3), are not both equal to zero. If $d<0.5$*

i) *the stochastic process $\{X_t\}_{t\in\mathbb{Z}}$ is strictly stationary and ergodic.*

ii) *if $\mathbb{E}(|\ln(Z_0^2)|)<\infty$, the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is well defined and it is strictly stationary and ergodic. Moreover, if $\mathbb{E}([\ln(Z_0^2)]^2)<\infty$ then it is also weakly stationary.*

Although both processes $\{X_t\}_{t\in\mathbb{Z}}$ and $\{\sigma_t^2\}_{t\in\mathbb{Z}}$ are strictly stationary, whenever $d<0.5$, they are not necessarily weakly stationary. This property depends on the distribution of Z_0 and not only on the existence of the second moment for this random variable. Theorem 2.2 gives the condition for the existence of the r-th moment, for any $\mathsf{r}>0$, for both processes $\{X_t\}_{t\in\mathbb{Z}}$ and $\{\sigma_t^2\}_{t\in\mathbb{Z}}$.

Theorem 2.2. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be an* SFIEGARCH$(p,d,q)_s$ *process, with $d<0.5$. Assume that θ and γ are not both equal to zero. If there exists $\mathsf{r}>2$ such that*

$$\mathbb{E}(|Z_0|^{\mathsf{r}})<\infty \quad \textit{and} \quad \sum_{k=0}^{\infty}\left|\mathbb{E}\Big(\exp\Big\{\frac{\mathsf{r}}{2}\lambda_{d,k}g(Z_0)\Big\}\Big)-1\right|<\infty, \tag{2.1}$$

then, $\mathbb{E}(|X_t|^{\mathsf{m}})<\infty$ and $\mathbb{E}(|\sigma_t|^{\mathsf{m}})<\infty$, for all $t\in\mathbb{Z}$ and $0<\mathsf{m}\leq\mathsf{r}$.

Assume that $d<0.5$ and $Z_0\sim\mathcal{N}(0,1)$. Let $\Phi(\cdot)$ and $\mathrm{erf}(\cdot)$ be, respectively, the standard Gaussian distribution and the error function, that is,

$$\Phi(z)=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\mathrm{d}t \quad \text{and} \quad \mathrm{erf}(z)=\frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}\mathrm{d}t=2\Phi(z\sqrt{2})-1, \quad \text{for all } z\in\mathbb{R}.$$

It follows that,

$$\begin{aligned}
\mathbb{E}\Big(\exp\Big\{\mathsf{r}\lambda_{d,k}g(Z_0)\Big\}\Big) &= \int_{-\infty}^{\infty}\exp\Big\{\mathsf{r}\lambda_{d,k}\Big[\theta|z|+\gamma\Big(|z|-\sqrt{2/\pi}\Big)\Big]\Big\}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\mathrm{d}z\\
&= e^{-\mathsf{r}\lambda_{d,k}\gamma\sqrt{2/\pi}}\Bigg[\exp\Big\{\frac{(\mathsf{r}\lambda_{d,k})^2(\theta-\gamma)^2}{2}\Big\}\Phi\big(\mathsf{r}\lambda_{d,k}(\gamma-\theta)\big)\\
&\qquad + \exp\Big\{\frac{(\mathsf{r}\lambda_{d,k})^2(\theta+\gamma)^2}{2}\Big\}\Phi\big(\mathsf{r}\lambda_{d,k}(\gamma+\theta)\big)\Bigg]\\
&= \frac{1}{2}e^{-\mathsf{r}\lambda_{d,k}\gamma\sqrt{2/\pi}}\Bigg(e^{\frac{(\mathsf{r}\lambda_{d,k})^2(\gamma-\theta)^2}{2}}\Bigg[1+\mathrm{erf}\Big(\frac{\mathsf{r}\lambda_{d,k}(\gamma-\theta)}{\sqrt{2}}\Big)\Bigg]\\
&\qquad + e^{\frac{(\mathsf{r}\lambda_{d,k})^2(\gamma+\theta)^2}{2}}\Bigg[1+\mathrm{erf}\Big(\frac{\mathsf{r}\lambda_{d,k}(\gamma+\theta)}{\sqrt{2}}\Big)\Bigg]\Bigg),
\end{aligned} \tag{2.2}$$

for all $k \in \mathbb{N}$ and all $\mathsf{r} > 0$. Since $e^x = 1 + x + O(x^2)$ and $\text{erf}(x) = \frac{2}{\sqrt{\pi}}x + O(x^3)$, as $x \to 0$, one can rewrite (2.2) as,

$$\mathbb{E}\Big(\exp\Big\{\mathsf{r}\lambda_{d,k}g(Z_0)\Big\}\Big) = \frac{1}{2}\left[1 - \mathsf{r}\lambda_{d,k}\gamma\sqrt{\frac{2}{\pi}} + O(\lambda_{d,k}^2)\right]\left[2 + 2\mathsf{r}\lambda_{d,k}\gamma\sqrt{\frac{2}{\pi}} + O(\lambda_{d,k}^2)\right] = 1 + O(\lambda_{d,k}^2),$$

as $k \to \infty$. Thus, condition (2.1) holds and hence, $\mathbb{E}(|X_t|^{\mathsf{r}}) < \infty$ and $\mathbb{E}(|\sigma_t|^{\mathsf{r}}) < \infty$, for all $\mathsf{r} > 0$. Corollary 2.2 shows that this result also holds if $Z_0 \sim \text{GED}(\nu)$, for any $\nu > 0$.

Corollary 2.2. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be an* SFIEGARCH$(p, d, q)_s$ *process, with $d < 0.5$. Assume that θ and γ are not both equal to zero. Let $\{Z_t\}_{t\in\mathbb{Z}}$ be i.i.d.* GED *with zero mean, variance equal to one, and tail-thickness parameter $\nu > 1$. Then, $\mathbb{E}(X_t^{\mathsf{r}}) < \infty$ and $\mathbb{E}([\sigma_t^2]^{\mathsf{r}}) < \infty$, for all $t \in \mathbb{Z}$ and $\mathsf{r} > 0$.*

From expression (A.1.5) in Nelson (1991), if $Z_0 \sim \text{GED}(\nu)$, with $\nu > 1$, and $b_k := \frac{r}{2}\lambda_{d,k}$, for all $k \in \mathbb{N}$ and any $r > 0$, then

$$\mathbb{E}(e^{b_k g(Z_0)}) = \exp\left\{-b_k\gamma\lambda 2^{1/\nu}\frac{\Gamma\left(\frac{2}{\nu}\right)}{\Gamma\left(\frac{1}{\nu}\right)}\right\}\sum_{j=0}^{\infty}\left(b_k\lambda 2^{1/\nu}\right)^j\left[(\gamma+\theta)^j + (\gamma-\theta)^j\right]\frac{\Gamma\left(\frac{j+1}{\nu}\right)}{2\Gamma\left(\frac{1}{\nu}\right)\Gamma(j+1)}, \tag{2.3}$$

where $\lambda = \left[2^{1/\nu}\Gamma(1/\nu)/\Gamma(3/\nu)\right]^{1/2}$, for all $k \in \mathbb{N}$. From expression (2.3), it is easy to see that $\mathbb{E}(X_t^r)$ is symmetric in θ, for any $r > 0$ and $\nu > 1$.

Example 2.1. Figures 2.1 - 2.3 consider SFIEGARCH$(0, d, 0)_s$ processes, with $s = 2$. In these figures we analyze the behavior of $\mathbb{E}(X_t^2)$, with respect to the parameters θ, γ and d. We also study the behavior of $\mathbb{E}(X_t^2)$ with respect to the parameter ν, when $Z_0 \sim \text{GED}(\nu)$. From expression (2.3) one observes that $\mathbb{E}(X_t^2)$ is symmetric in θ, whenever $Z_0 \sim \text{GED}(\nu)$, for any $\nu > 1$. Therefore, for these figures we only consider positive values of θ. Figure 2.1 (a) shows the behavior of $\mathbb{E}(X_t^2)$ as a function of θ and γ, for $d = 0.25$. Figure 2.1 (b) presents $\mathbb{E}(X_t^2)$ as a function of θ and d, for $\gamma = 0.24$. Figure 2.1 (c) shows $\mathbb{E}(X_t^2)$ as a function of γ and d, for $\theta = 0.25$. For all graphs in Figure 2.1, $s = 2$, $\omega = 0$ and $Z_0 \sim \mathcal{N}(0, 1)$.

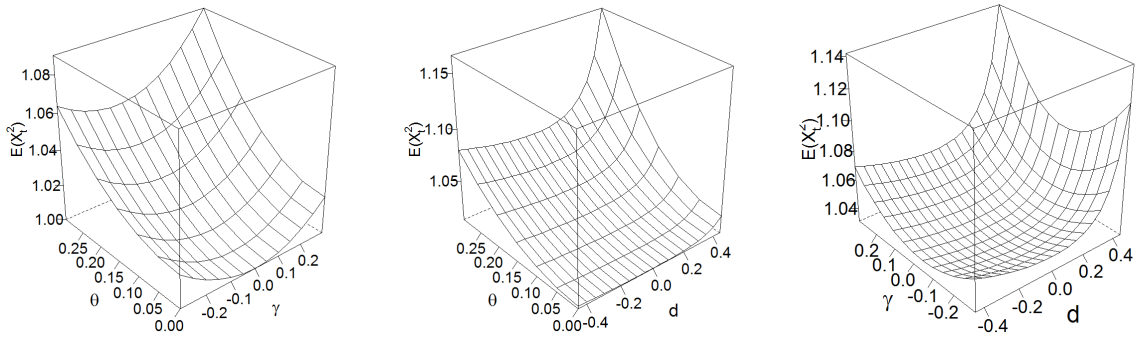
From Figure 2.1, one observes that for θ or γ fixed, $\mathbb{E}(X_t^2)$ slowly decreases for $d \in [-0.45, 0]$ and increases for $d \in [0, 0.45]$. This behavior can be better observed in Figures 2.2 (a) and (b) where the values of $\mathbb{E}(X_t^2)$, as a function of d, are plotted for $\theta \in \{0, 0.15, 0.30\}$ and $\gamma \in \{-0.30, 0, 0.30\}$, respectively. Similarly, for each d fixed, the function $\mathbb{E}(X_t^2)$ is decreasing for $\theta, \gamma \in [-0.3, 0.0]$ (the function is symmetric in θ) and it is increasing for $\theta, \gamma \in [0.0, 0.3]$. This behavior can be observed in Figures 2.2 (c) and (d) where $\mathbb{E}(X_t^2)$ is given, respectively, as a function of θ and γ, for $d \in \{-0.40, 0, 0.45\}$.

Example 2.2. Figure 2.3 (a) shows the graph of $\mathbb{E}(X_t^2)$, as a function of ν and d, when $Z_0 \sim \text{GED}(\nu)$, with $\nu > 1$. Figures 2.3 (b) and (c) present the graph of $\mathbb{E}(X_t^2)$, respectively, as a function of ν, for $d \in \{-0.45, 0, 0.45\}$ and as a function of d, for $\nu \in \{1.01, 3, 5\}$. For all graphs, $s = 2$, $\omega = 0$, $\theta = 0.25$ and $\gamma = 0.24$. From Figure 2.3 one concludes that $\mathbb{E}(X_t^2)$ is a decreasing function of ν and, as a function of d, $\mathbb{E}(X_t^2)$ presents the same behavior as in the Gaussian case, that is, it is decreasing for $d \in [-0.45, 0]$ and increasing for $d \in [0, 0.45]$.

The following proposition presents the kurtosis and the asymmetry measures for any stationary SFIEGARCH process.

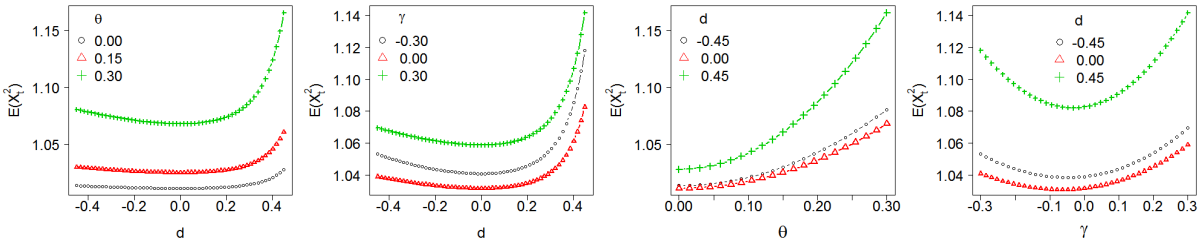
Proposition 2.1. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a stationary* SFIEGARCH$(p, d, q)_s$ *process with $\mathbb{E}(|X_t^4|) < \infty$. The asymmetry and kurtosis measures of $\{X_t\}_{t\in\mathbb{Z}}$ are given, respectively, by*

$$A_X = \mathbb{E}(Z_0^3)\,\frac{\prod_{k=0}^{\infty}\mathbb{E}\left(\exp\left\{\frac{3}{2}\lambda_{d,k}g(Z_t)\right\}\right)}{\prod_{k=0}^{\infty}\left[\mathbb{E}\left(\exp\left\{\lambda_{d,k}g(Z_t)\right\}\right)\right]^{3/2}} \quad \text{and} \quad K_X = \mathbb{E}(Z_0^4)\,\frac{\prod_{k=0}^{\infty}\mathbb{E}\left(\exp\left\{2\lambda_{d,k}g(Z_t)\right\}\right)}{\prod_{k=0}^{\infty}\left[\mathbb{E}\left(\exp\left\{\lambda_{d,k}g(Z_t)\right\}\right)\right]^2}.$$



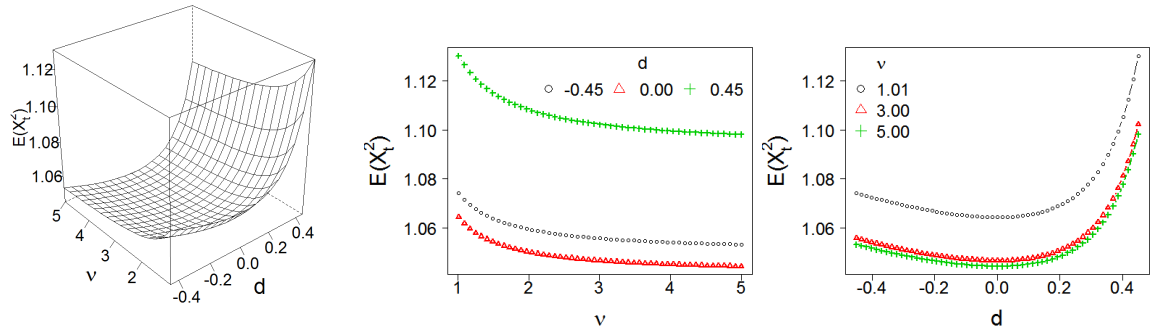
(a) (b) (c)

Figure 2.1: This figure illustrates the behavior of $\mathbb{E}(X_t^2)$ with respect to the parameters θ, γ and d when $\{X_t\}_{t\in\mathbb{Z}}$ is an SFIEGARCH$(0,d,0)_s$ processes, with $s=2$, $\omega=0$ and $Z_0 \sim \mathcal{N}(0,1)$. Panel (a) fixes $d=0.25$ and shows $\mathbb{E}(X_t^2)$ as a function of θ and γ. Panel (b) fixes $\gamma=0.24$ and presents $\mathbb{E}(X_t^2)$ as a function of θ and d. Panel (c) fixes $\theta=0.25$ and considers $\mathbb{E}(X_t^2)$ as a function of γ and d.



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

Figure 2.2: This figure shows the behavior of $\mathbb{E}(X_t^2)$ as a function of d, θ or γ, when $\{X_t\}_{t\in\mathbb{Z}}$ is an SFIEGARCH$(0,d,0)_s$ processes with $s=2$, $\omega=0$ and $Z_0 \sim \mathcal{N}(0,1)$. Panel (a) shows $\mathbb{E}(X_t^2)$ as a function of d, for $\theta \in \{0, 0.15, 0.30\}$ and $\gamma=0.24$. Panel (b) considers $\mathbb{E}(X_t^2)$ as a function of d, for $\gamma \in \{-0.30, 0.15, 0.30\}$ and $\theta=0.25$. Panel (c) gives $\mathbb{E}(X_t^2)$ as a function of θ, for $d \in \{-0.45, 0, 0.45\}$ and $\gamma=0.24$. Panel (d) presents $\mathbb{E}(X_t^2)$ as a function of γ, for $d \in \{-0.45, 0, 0.45\}$ and $\theta=0.25$.

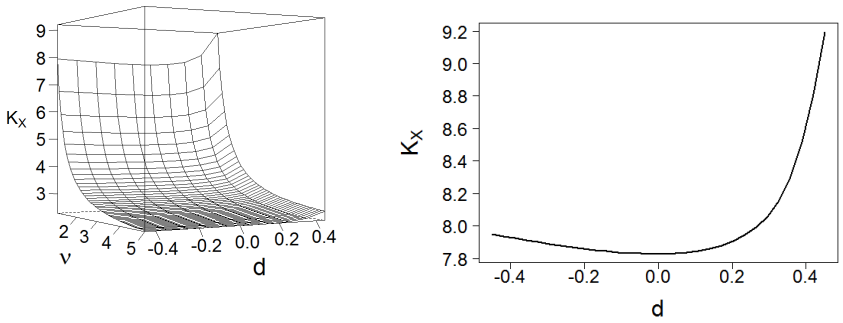


(a) (b) (c)

Figure 2.3: This figure illustrates the behavior of $\mathbb{E}(X_t^2)$ as a function of d and ν, when $\{X_t\}_{t\in\mathbb{Z}}$ is an SFIEGARCH$(0,d,0)_s$ processes with $s=2$, $\theta=0.25$, $\gamma=0.24$, $\omega=0$ and $Z_0 \sim \text{GED}(\nu)$. Panel (a) shows $\mathbb{E}(X_t^2)$ as a function of ν and d. Panel (b) gives $\mathbb{E}(X_t^2)$ as a function of ν, for $d \in \{-0.45, 0, 0.45\}$. Panel (c) considers $\mathbb{E}(X_t^2)$ as a function of d, for $\nu \in \{1.01, 3, 5\}$.

Example 2.3. From expression (2.3), one easily concludes that the kurtosis measure K_X is symmetric in θ, whenever $Z_0 \sim \text{GED}(\nu)$, for any $\nu > 1$. Figure 2.4 (a) considers SFIEGARCH$(0,d,0)_s$ processes and shows the behavior of K_X as a function of d and ν, for $s=2$, $\omega=0$, $\theta=0.25$ and $\gamma=0.24$. Figure 2.4 (b) presents the graph of K_X as a function of d, for $\nu=2$. From the graphs in

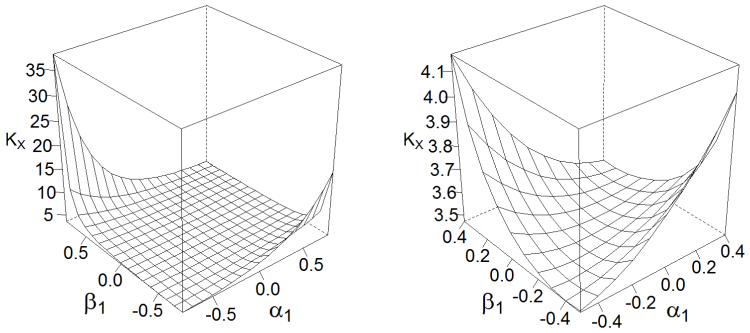
Figure 2.4 one concludes that the kurtosis measure is a decreasing function of ν, for each d fixed. Moreover, for each ν fixed, it is decreasing for $d \in [-0.45, 0]$ and increasing for $d \in [0, 0.45]$.



(a) (b)

Figure 2.4: The kurtosis measure of an SFIEGARCH$(0, d, 0)_s$ process with $s = 2$, $\omega = 0$, $\theta = 0.25$, $\gamma = 0.24$ and $Z_0 \sim \text{GED}(\nu)$. Panel (a) shows the graph of the kurtosis measure as a function of d and ν. Panel (b) presents the kurtosis measure as a function of d, for $\nu = 2$ fixed.

Example 2.4. Figure 2.5 considers SFIEGARCH$(p, d, q)_s$ processes, with $p, q \in \{0, 1\}$, $d = 0.25$ and the same values of s, ω, θ and γ as in Figure 2.4. This figure presents the behavior of K_X as a function of α_1 and β_1. The cases $\alpha_1 = \beta_1$ (the polynomials have a common root) are actually equivalent to the case $\alpha_1 = 0 = \beta_1$ and, in this case, one has an SFIEGARCH$(0, d, 0)_s$ process. While Figure 2.5 (a) shows the graphs of K_X for $\alpha_1, \beta_1 \in [-0.8, 0.8]$, Figure 2.5 (b) considers only the interval $[-0.4, 0.4]$.



(a) (b)

Figure 2.5: The kurtosis measure of an SFIEGARCH$(p, d, q)_s$ process with $p, q \in \{0, 1\}$, $s = 2$, $\omega = 0$, $\theta = 0.25$, $\gamma = 0.24$ and $Z_0 \sim \text{GED}(\nu)$. Panel (a) shows the kurtosis measure as a function of α_1 and β_1 for $\alpha_1, \beta_1 \in [-0.8, 0.8]$. Panel (b) considers α_1 and β_1 only in the interval $[-0.4, 0.4]$ (for a better visualization).

From Figure 2.5 one observes that the behavior of K_X depends on the sign of both α_1 and β_1. Also, it goes from increasing (when $\beta_1 = -0.8$) to decreasing (when $\beta_1 = 0.8$) in α_1. Also, by comparing the graphs in Figures 2.5 (a) and (b), it is easy to see that the kurtosis measure presents small variation on its value for $(\alpha_1, \beta_1) \in [-0.4, 0.4] \times [-0.4, 0.4]$ (in this region $3 < K_X < 4.5$). Moreover, for $\alpha_1 \in [-0.8, -0.4]$ and $\beta_1 \in [-0.8, 0.8]$ the kurtosis measure varies faster than in the case $\beta_1 \in [-0.8, -0.4]$ and $\alpha_1 \in [-0.8, 0.8]$.

Lemma 2.2 and Corollary 2.3 present the autocovariance function of the process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ and its asymptotic behavior. Although, in practice, this stochastic process cannot be observed and hence, the sample autocovariance structure cannot be analysed, the results presented in these theorems are necessary to prove Theorem 2.3.

Lemma 2.2. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be an* SFIEGARCH$(p,d,q)_s$ *process, defined by (1.1) and (1.2), with $d<0.5$. Suppose that γ and θ are not both equal to zero. Then, the autocovariance function $\gamma_{\ln(\sigma_t^2)}(\cdot)$ of the process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is given by*

$$\gamma_{\ln(\sigma_t^2)}(sh+r)=\sum_{k\in\mathbb{Z}}\gamma_A(sk+r)\gamma_V(sh-sk),\quad \text{for all } h\in\mathbb{Z}\quad \text{and } r\in\{0,\cdots,s-1\},\tag{2.4}$$

where $\gamma_A(\cdot)$ and $\gamma_V(\cdot)$ are given, respectively, by

$$\gamma_A(h)=\sum_{i=0}^{\infty}f_if_{i+|h|},\quad \text{for all } h\in\mathbb{Z},\quad \text{with } f(z):=\frac{\alpha(z)}{\beta(z)}=\sum_{k=0}^{\infty}f_kz^k,\tag{2.5}$$

and

$$\gamma_V(hs)=\sigma_g^2\frac{(-1)^h\Gamma(1-2d)}{\Gamma(1-d+h)\Gamma(1-d-h)}\quad \text{and}\quad \gamma_V(hs+r)=0,\tag{2.6}$$

for all $h\in\mathbb{Z}$ and $r\in\{0,\cdots,s-1\}$, with σ_g^2 given by (1.14).

From Lemma 2.2 one easily concludes that, if $p=0=q$, then $\gamma_{\ln(\sigma_t^2)}(sh+r)\neq 0$ if an only if $r=0$. Moreover, if $p>0$ and $q=0$, $\gamma_{\ln(\sigma_t^2)}(sh+r)=0$, for all $h\in\mathbb{Z}$, whenever $\lceil-\frac{p+r}{s}\rceil>\lfloor\frac{p-r}{s}\rfloor$. This is so because $\gamma_A(sk+r)\neq 0$ if and only if $|sk+r|\leq p$, that is, $-\frac{p+r}{s}\leq k\leq\frac{p-r}{s}$ and $k\in\mathbb{Z}$. Thus, for $p\geq 0$ and $q=0$, one can rewrite (2.4) as

$$\gamma_{\ln(\sigma_t^2)}(sh+r)=\begin{cases}\displaystyle\sum_{\lceil-\frac{p+r}{s}\rceil\leq k\leq\lfloor\frac{p-r}{s}\rfloor}\gamma_A(sk+r)\gamma_V(sh-sk), & \text{if}\quad \lceil-\frac{p+r}{s}\rceil\leq\lfloor\frac{p-r}{s}\rfloor;\\ 0, & \text{otherwise,}\end{cases}$$

for all $r\in\{0,\cdots,s-1\}$, and $h\in\mathbb{Z}$. In this case, it is obvious that $\sum_{h=0}^{\infty}|\rho_{\ln(\sigma_t^2)}(h)|<\infty$ if and only if, $\sum_{h=0}^{\infty}|\rho_V(h)|<\infty$. Corollary 2.3 presents the asymptotic behavior of $\gamma_{\ln(\sigma_t^2)}(h)$, as $h\to\infty$, which leads to the conclusion that this result actually holds for any p and q.

Corollary 2.3. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be an* SFIEGARCH$(p,d,q)_s$ *process, defined by (1.1) and (1.2), with $d<0.5$. Suppose that γ and θ are not both equal to zero. Let $\gamma_{\ln(\sigma_t^2)}(\cdot)$ be the autocovariance function of the process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$. Then, for all $r\in\{0,\cdots,s-1\}$,*

$$\gamma_{\ln(\sigma_t^2)}(sh+r)=\gamma_V(sh)\mathscr{G}(sh+r)+o(h^{-\nu}),\quad \text{as}\quad h\to\infty,\tag{2.7}$$

where ν is any positive number and $\mathscr{G}(\cdot)$ is a real function satisfying

$$\lim_{h\to\infty}\sum_{r=0}^{s-1}\mathscr{G}(sh+r)=\sum_{k\in\mathbb{Z}}\gamma_A(k),\quad \text{with } \gamma_A(\cdot) \text{ given by (2.5).}$$

Theorem 2.3 presents the autocovariance function $\gamma_{\ln(X_t^2)}(\cdot)$ of the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, where $\{X_t\}_{t\in\mathbb{Z}}$ is an SFIEGARCH process. This theorem also gives the asymptotic behavior of $\gamma_{\ln(X_t^2)}(sh+r)$, for all $r\in\{0,\cdots,s-1\}$, as h goes to infinity.

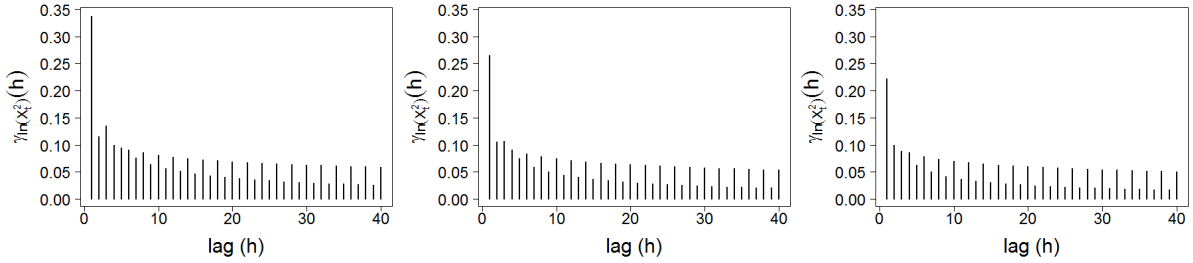
Theorem 2.3. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be an* SFIEGARCH$(p,d,q)_s$ *process, defined by (1.1) and (1.2), with $d<0.5$. Suppose that γ and θ are not both equal to zero and $\mathrm{Var}(\ln(Z_0^2)):=\sigma^2_{\ln(Z_t^2)}<\infty$. Then, the autocovariance function of the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is given by*

$$\gamma_{\ln(X_t^2)}(h)=\gamma_{\ln(\sigma_t^2)}(h)+C_1\lambda_{d,|h|-1}\mathbb{I}_{\mathbb{Z}^*}(h)+\sigma^2_{\ln(Z_t^2)}\mathbb{I}_{\{0\}}(h),\quad \text{for all } h\in\mathbb{Z},\tag{2.8}$$

where $C_1=\mathrm{Cov}(g(Z_0),\ln(Z_0^2))$ and $\gamma_{\ln(\sigma_t^2)}(\cdot)$ is given in Lemma 2.2. Thus, for all $\nu>0$,

$$\gamma_{\ln(X_t^2)}(sh+r)=\gamma_V(sh)\mathscr{G}(sh+r)+\pi_{d,s\lfloor\frac{sh+r-1}{s}\rfloor}\mathscr{K}(sh+r-1)+o(h^{-\nu}),\quad \text{as } h\to\infty,\tag{2.9}$$

for all $r\in\{0,\cdots,s-1\}$, where $\mathscr{G}(\cdot)$ and $\mathscr{K}(\cdot)$ are given, respectively, in Corollary (2.3) and Theorem 1.1.

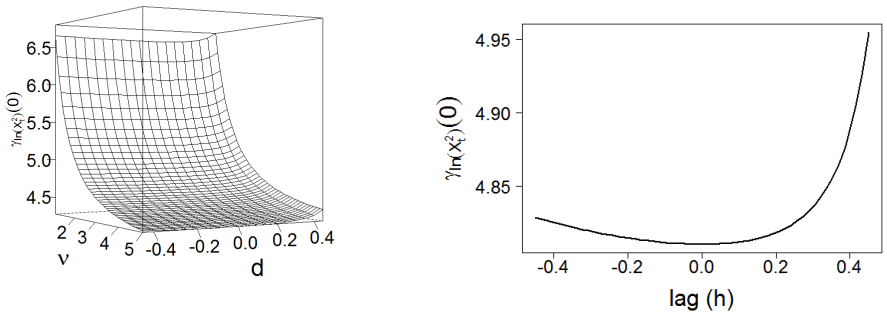


(a) (b) (c)

Figure 2.6: Theoretical autocovariance function $\gamma_{\ln(X_t^2)}(\cdot)$, for $h \in \{1, \cdots, 100\}$, corresponding to the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, when $\{X_t\}_{t\in\mathbb{Z}}$ is an SFIEGARCH$(0, d, 0)_s$ process with $d = 0.4$, $s = 2$, $\omega = 0$, $\theta = 0.25$, $\gamma = 0.24$ and $Z_0 \sim \text{GED}(\nu)$. Panel (a) considers $\nu = 1.01$. Panel (b) assumes $\nu = 2$ (Gaussian case). Panel (c) fixes $\nu = 5$.

Example 2.5. From expressions (2.4) and (2.8), one concludes that, if Z_0 is a symmetric random variable, then $\gamma_{\ln(X_t^2)}(\cdot)$ is symmetric in θ. Figures 2.6 (a) - (d) show the graphs of $\gamma_{\ln(X_t^2)}(h)$, for $h \in \{1, \cdots, 100\}$, where $\{X_t\}_{t\in\mathbb{Z}}$ is an SFIEGARCH$(0, d, 0)_s$ process, with $d = 0.4$, $s = 2$, $\omega = 0$, $\theta = 0.25$, $\gamma = 0.24$ and $Z_0 \sim \text{GED}(\nu)$, for $\nu \in \{1.01, 2, 3, 5\}$, respectively. All graphs are presented in the same scale for a better visualization. The corresponding values of $\gamma_{\ln(X_t^2)}(0)$ are, respectively, 6.7228, 5.0978, 4.6445 and 4.3556. From Figure 2.6, one observes that, for each fixed h, $\gamma_{\ln(X_t^2)}(h)$ decreases as ν increases.

Example 2.6. Figure 2.7 (a) presents $\gamma_{\ln(X_t^2)}(0) = \text{Var}(\ln(X_t^2))$ as a function of ν and d, where $\{X_t\}_{t\in\mathbb{Z}}$ is an SFIEGARCH$(0, d, 0)_s$ process, with $s = 2$, $\omega = 0$, $\theta = 0.25$ and $\gamma = 0.24$. Figure 2.7 (b) presents $\gamma_{\ln(X_t^2)}(0)$ as a function of d, for $\nu = 2$. From the graphs in Figure 2.7 one observes that the variance of $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ decreases with ν. For each ν fixed, $\gamma_{\ln(X_t^2)}(0)$ is decreasing for $d \in [-0.45, 0]$ and increasing for $d \in [0, 0.45]$.



(a) (b)

Figure 2.7: This figure shows the graphs of $\sigma^2_{\ln(X_t^2)} := \gamma_{\ln(X_t^2)}(0)$, the variance of the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, when $\{X_t\}_{t\in\mathbb{Z}}$ is an SFIEGARCH$(0, d, 0)_s$ process with $s = 2$, $\omega = 0$, $\theta = 0.25$, $\gamma = 0.24$ and $Z_0 \sim \text{GED}(\nu)$. Panel (a) considers the variance as a function of d and ν. Panel (b) shows the variance as a function of d, for $\nu = 2$ fixed.

The following corollary compares the asymptotic behavior of $\sum_{r=0}^{s-1} \gamma_{\ln(X_t^2)}(sh + r)$ when $d < 0$ and $d > 0$.

Corollary 2.4. *Let $\gamma_{\ln(X_t^2)}(\cdot)$ be the autocovariance function of the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, given in Theorem 2.3. Then,*

$$\sum_{r=0}^{s-1} \gamma_{\ln(X_t^2)}(sh + r) \sim \begin{cases} \mathscr{C}_1(h)h^{d-1}, & \text{if } d < 0; \\ \mathscr{C}_2(h)h^{2d-1}, & \text{if } d > 0, \end{cases} \tag{2.10}$$

with

$$\mathscr{C}_1(h) := \sigma_g^2 \frac{\Gamma(1-2d)}{\Gamma(1-d)\Gamma(d)} \sum_{r=0}^{s-1} \mathscr{G}(sh+r) \quad \textit{and} \quad \mathscr{C}_2(h) := \frac{1}{\Gamma(d)} \sum_{r=0}^{s-1} \mathscr{K}(sh+r-1), \tag{2.11}$$

where $\mathscr{G}(\cdot)$ and $\mathscr{K}(\cdot)$ are given, respectively, in Corollary 2.3 and Theorem 1.1 and σ_g^2 is given by (1.14).

3 Spectral Representation

Recently economists have noticed that volatility of high frequency financial time series shows long range dependence merged with periodic behavior due to some operating features of financial markets. Periodic components are represented as marked peaks at some frequencies in the periodogram function. It is a well known result that the periodogram function is an estimator of the spectral density function. Therefore, in order to choose the best model for a time series, one should know how does the spectral density function behaves in order to gather information from the periodogram function.

Here we present the spectral density function of both processes $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ and $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$. It is easy to see that expression (3.1), in Theorem 3.1, is similar to the expression (2.5) from Hurvich et al. (2005). In this paper, the authors present the asymptotic properties of some semiparametric estimators for the long-memory parameter for a class of stochastic process which includes LMSV (*Long Memory Stochastic Volatility*) and FIEGARCH models.

Theorem 3.1. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be an* SFIEGARCH$(p,d,q)_s$ *process, defined by (1.1) and (1.2), with $d < 0.5$. Suppose that γ and θ, given in (1.3), are not both equal to zero and that $\alpha(z) \neq 0$ in the closed disk $\{z : |z| \leq 1\}$. If $\mathrm{Var}(\ln(Z_t^2)) := \sigma^2_{\ln(Z_t^2)} < \infty$, for all $t \in \mathbb{Z}$, the spectral density function of $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is given by*

$$f_{\ln(X_t^2)}(\lambda) = f_{\ln(\sigma_t^2)}(\lambda) + \frac{C_1}{\pi} \Re\big(e^{-\mathrm{i}\lambda}\Lambda(\lambda)\big) + f_{\ln(Z_t^2)}(\lambda), \quad \textit{for all } \lambda \in [0,\pi], \tag{3.1}$$

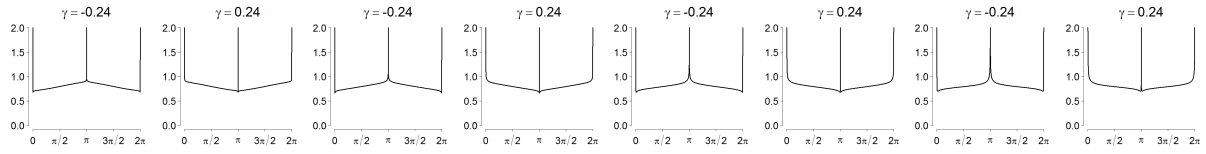
where $f_{\ln(\sigma_t^2)}(\cdot)$ is given in (A.6), $f_{\ln(Z_t^2)}(\lambda) = \frac{\sigma^2_{\ln(Z_t^2)}}{2\pi}$ is the spectral density function of $\{\ln(Z_t^2)\}_{t\in\mathbb{Z}}$, $C_1 = \mathrm{Cov}\big(g(Z_0), \ln(Z_0^2)\big)$, $\Lambda(z) := \lambda(e^{-\mathrm{i}z})$ and $\lambda(\cdot)$ is defined in (1.9).

Notice that the spectral density function is symmetric around π. Hence, in what follows, although the graphs consider the interval $[0, 2\pi]$, one only needs to pay attention to the interval $[0, \pi]$. Moreover, all graphs are presented in the same scale and they are truncated in the y-axis for a better visualization.

Example 3.1. Figures 3.1 and 3.2 show the spectral density function of the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, where $\{X_t\}_{t\in\mathbb{Z}}$ is an SFIEGARCH$(0,d,0)_s$ with different parameter values and $Z_0 \sim \mathcal{N}(0,1)$. Since Z_0 is a symmetric random variable, $\mathbb{E}(Z_0|Z_0|) = 0$ and the function $f_{\ln(X_t^2)}(\cdot)$ is symmetric in θ. Thus, in both figures, we fixed $\theta = 0.25$. In Figure 3.1 we fix $s = 2$ and, in Figure 3.2, we consider $s = 6$. For each figure, $\gamma \in \{-0.24, 0.24\}$ and $d \in \{0.1, 0.2, 0.3, 0.4\}$. From Figures 3.1 and 3.2, one observes that, for each fixed d and s, the behavior of the function completely changes as γ changes from -0.24 to 0.24 (left to right). While for $\gamma = -0.24$ the function attains its minimum in the region close to zero, for $\gamma = 0.24$ the minimum is attained close to π.

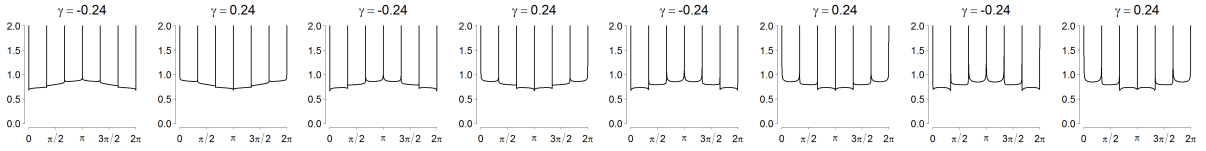
Example 3.2. Figures 3.3 - 3.5 present the spectral density function of $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, where $\{X_t\}_{t\in\mathbb{Z}}$ is an SFIEGARCH$(p,d,q)_s$, with $Z_0 \sim \mathcal{N}(0,1)$, $p, q \in \{0,1\}$ (not both equal to zero), $s = 4$, $d = 0.25$, $\theta = 0.25$, $\gamma \in \{-0.24, 0.24\}$ and $\alpha_1, \beta_1 \in \{-0.9, -0.5, -0.1, 0.1, 0.5, 0.9\}$. For these figures the parameters values increase from left to right and from top to bottom.

Although Figure 3.5 presents only the graphs for $\alpha_1 \in \{-0.5, 0.5\}$, in the sequel we discuss the behavior of $f_{\ln(X_t^2)}(\cdot)$, for all $\alpha_1 \in \{-0.9, -0.5, -0.1, 0.1, 0.5, 0.9\}$. The remaining graphs are available upon request. From Figures 3.3 - 3.5, one concludes the following:



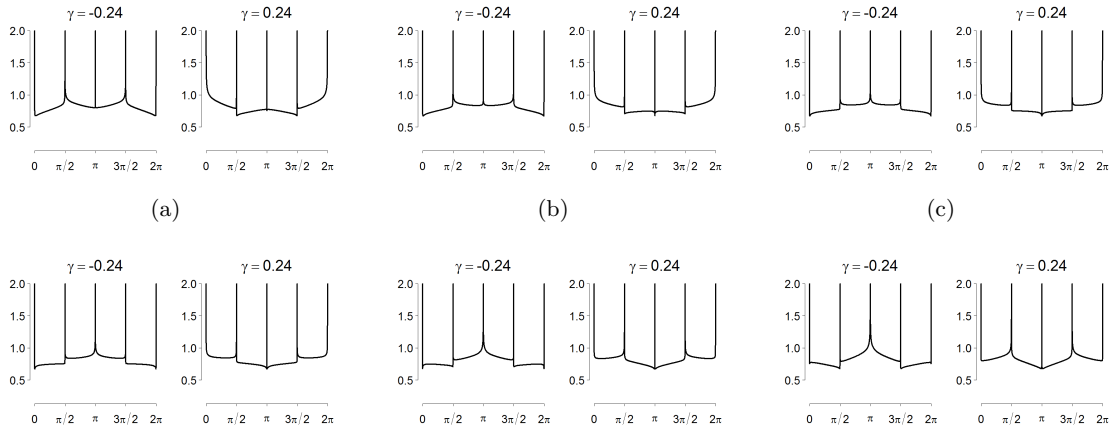
(a) (b) (c) (d)

Figure 3.1: Theoretical spectral density function of the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, when $\{X_t\}_{t\in\mathbb{Z}}$ is an SFIEGARCH$(0,d,0)_s$ process with $s=2$, $\theta=0.25$ and $\gamma\in\{-0.24,0.24\}$ (in each panel, from left to right). The parameter d is set as follows: in (a) $d=0.1$, in (b) $d=0.2$, in (c) $d=0.3$ and in (d) $d=0.4$.



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

Figure 3.2: Theoretical spectral density function of the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, when $\{X_t\}_{t\in\mathbb{Z}}$ is an SFIEGARCH$(0,d,0)_s$ process with $s=6$, $\theta=0.25$ and $\gamma\in\{-0.24,0.24\}$ (in each panel, from left to right). The parameter d is set as follows: in (a) $d=0.1$, in (b) $d=0.2$, in (c) $d=0.3$ and in (d) $d=0.4$.



(a) (b) (c)

(d) (e) (f)

Figure 3.3: Theoretical spectral density function of $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, when $\{X_t\}_{t\in\mathbb{Z}}$ is an SFIEGARCH$(1,d,0)_s$, with $s=4$, $\omega=0$, $d=0.25$, $\theta=0.25$, $\gamma\in\{-0.24,0.24\}$ (in each panel, from left to right). The parameter α_1 is set as follows: in (a) $\alpha_1=-0.9$, in (b) $\alpha_1=-0.5$, in (c) $\alpha_1=-0.1$, in (d) $\alpha_1=0.1$, in (e) $\alpha_1=0.5$ and in (f) $\alpha_1=0.9$.

- if $p=1$ and $q=0$
  - the region where $f_{\ln(X_t^2)}(\cdot)$ attains its minimum depends not only on the sign of γ, but also on the sign of α_1. If $\gamma<0$, the minimum is attained either close to zero or close to $\pi/2$. If $\gamma>0$, the minimum is attained either close to $\pi/2$ or close to π;
  - for γ fixed, the graph of $f_{\ln(X_t^2)}(\cdot)$ slowly changes its behavior in the regions around the seasonal frequencies, as α_1 increases;
  - almost no difference is observed in the graphs of $f_{\ln(X_t^2)}(\cdot)$ when α_1 changes from -0.1 to 0.1. Actually, for these values of α_1, the function behaves as in the case $p=0=q$ (the

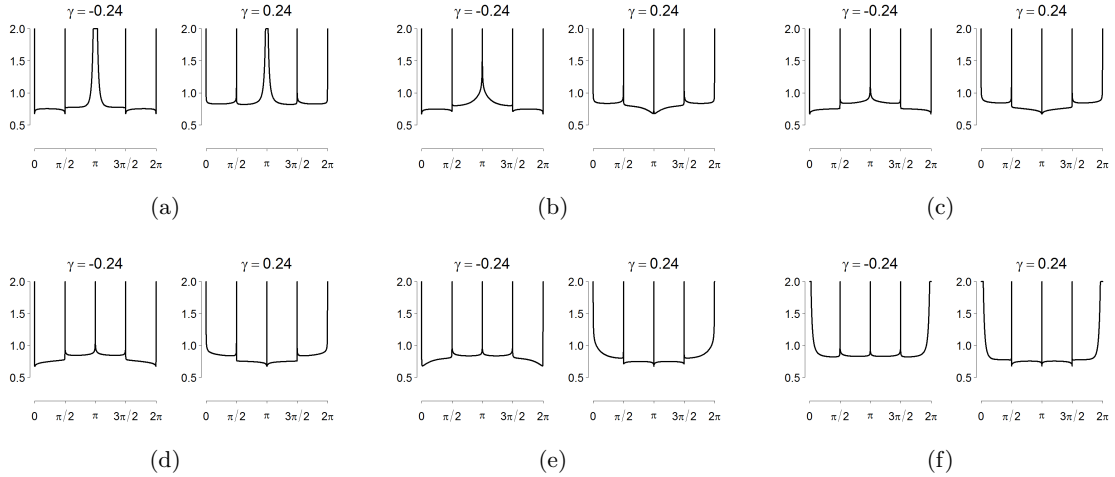


Figure 3.4: Theoretical spectral density function of $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, when $\{X_t\}_{t\in\mathbb{Z}}$ is an SFIEGARCH$(1,d,0)_s$, with $s=4$, $\omega=0$, $d=0.25$, $\theta=0.25$, $\gamma\in\{-0.24,0.24\}$ (in each panel, from left to right). The parameter β_1 is set as follows: in (a) $\beta_1=-0.9$, in (b) $\beta_1=-0.5$, in (c) $\beta_1=-0.1$, in (d) $\beta_1=0.1$, in (e) $\beta_1=0.5$ and in (f) $\beta_1=0.9$.

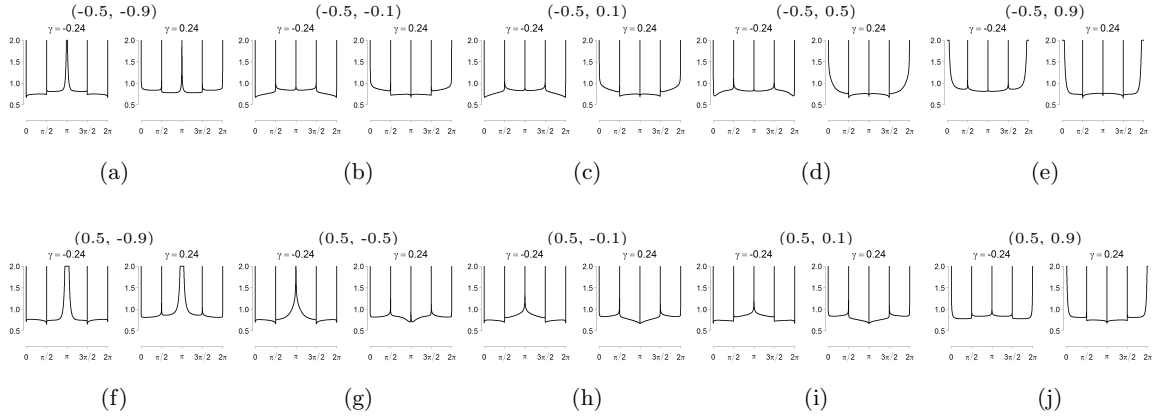


Figure 3.5: Theoretical spectral density function of $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, when $\{X_t\}_{t\in\mathbb{Z}}$ is an SFIEGARCH$(1,d,1)_s$, with $s=4$, $\omega=0$, $d=0.25$, $\theta=0.25$. For each panel $\gamma\in\{-0.24,0.24\}$ (from left to right). For all panels in the first row $\alpha_1=-0.5$ and for all panels in the second row $\alpha=0.5$. For each row $\beta_1\in\{-0.9,-0.5,-0.1,0.1,0.5,0.9\}$ (from left to right panel).

graphs can be obtained upon request);

- the graph of $f_{\ln(X_t^2)}(\cdot)$ for $\alpha_1=-0.9$ and $\gamma=-0.24$ is very similar to the graph of the same function for $\alpha_1=0.9$, $\gamma=0.24$.

- if $p=0$ and $q=1$
  - the region where $f_{\ln(X_t^2)}(\cdot)$ attains its minimum depends on the sign of γ and on the sign of β_1. The behavior is similar to the case $p=1$ and $q=0$;
  - for γ fixed, the changes in the graph of $f_{\ln(X_t^2)}(\cdot)$, in the regions around the seasonal frequencies, are much more visible than in the case $p=1$ and $q=0$;
  - for $\beta_1\in\{-0.1,0.1\}$ the graphs are very similar to the case $p=0$ and $q=1$, with $\alpha_1\in\{-0.1,0.1\}$.
  - The graphs of $f_{\ln(X_t^2)}(\cdot)$ for $\beta_1=-0.5$ are almost identical to the graphs of this function in the case $p=1$ and $q=0$ with $\alpha_1=0.5$. The same similarity is observed between the graphs of this function for $\beta_1=0.5$ and $\alpha_1=-0.5$.

- if $p = 1$ and $q = 1$
  - for γ and θ fixed, the changes in the graph of $f_{\ln(X_t^2)}(\cdot)$ are more visible as β_1 changes than when α_1 does;
  - for $\alpha_1, \beta_1 \in \{-0.1, 0.1\}$ the graphs are very similar to the case $p = 0$ and $q = 1$ (same occurs with $p = 1$ and $q = 0$);
  - generally, the graphs do not show any peculiar characteristic that is not present in the cases $p = 0$ or $q = 0$.

4 Forecasting

Let $\{X_t\}_{t\in\mathbb{Z}}$ be an SFIEGARCH$(p, d, q)_s$ process and $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$ be the filtration defined by $\mathcal{F}_t := \sigma(\{Z_s\}_{s\leq t})$. Notice that, by considering the same argument as in the proof of lemma 1 in Lopes and Prass (2013), one can show that $\{X_t\}_{t\in\mathbb{Z}}$ is a martingale difference with respect to $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$. In this case, the best predictor (in terms of the mean square error measure) for X_{t+h}, given $\mathcal{F}_t$, is $\mathbb{E}(X_{t+h}|\mathcal{F}_t) = 0$, for all $h > 0$ and $t \in \mathbb{Z}$.

Since the h-step ahead predictor for $\{X_t\}_{t\in\mathbb{Z}}$ is always zero (the mean of the process), the aim of this section is to derive expressions for the h-step ahead forecast for the processes $\{\sigma_t^2\}_{t\in\mathbb{Z}}$, $\{X_t^2\}_{t\in\mathbb{Z}}$, $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ and $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, for any $h > 0$. The approach considered in this work is slightly different from Lopes and Prass (2013). In Lopes and Prass (2013) two different h-step ahead predictors for σ_{t+h}^2 (consequently, for X_{t+h}^2) were proposed, both based on the h-step ahead predictor for $\ln(\sigma_{t+h}^2)$. Here we provide the exact formula for $\mathbb{E}(\sigma_{t+h}^2|\mathcal{F}_t)$, for any $h > 0$ and $t \in \mathbb{Z}$, and the relation between this expression and the ones in Lopes and Prass (2013).

Remark 4.1. In the sequel we consider the following notation, which is the same as in Lopes and Prass (2013). Let Y_t, for $t \in \mathbb{Z}$, denote any random variable. Then

- the symbol "$\hat{}$" denotes 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 (or non-linear) predictor in terms of mean square error value;
- the symbols "$\tilde{}$" (e.g. $\tilde{Y}_{t+h}$) and "$\check{}$" (e.g. $\check{Y}_{t+h}$) denote alternative estimators;
- $\hat{\ln}(Y_{t+h})$ denotes the h-step ahead forecast of $\ln(Y_{t+h})$ (analogously, for "$\tilde{}$" and "$\check{}$");
- we follow the approach usually considered in the literature and denote the h-step ahead forecast of Y_{t+h}^2 as $\hat{Y}_{t+h}^2$ instead of $\widehat{Y_{t+h}^2}$ and, to avoid confusion, we will denote the square of $\hat{Y}_{t+h}$ as $(\hat{Y}_{t+h})^2$ (analogously, for "$\tilde{}$" and "$\check{}$").

To obtain the predictors for the processes $\{\sigma_t^2\}_{t\in\mathbb{Z}}$ and $\{X_t^2\}_{t\in\mathbb{Z}}$ observe that the i.i.d. property of $\{Z_t\}_{t\in\mathbb{Z}}$ implies that $\mathbb{E}(Z_{n+h}^2|\mathcal{F}_n) = \mathbb{E}(Z_{n+h}^2) = 1$, σ_{n+1}^2 is $\mathcal{F}_n$-measurable and σ_n^2 and Z_n^2 are independent, for all $n \in \mathbb{Z}$ and $h > 0$. Therefore, the h-step ahead forecast for X_{n+h}^2 given $\mathcal{F}_n$ is given by

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

In particular, $\hat{\sigma}_{n+1}^2 = \sigma_{n+1}^2$ and hence the 1-step ahead forecast of X_{n+1}^2, given $\mathcal{F}_n$, is simply $\hat{X}_{n+1}^2 = \sigma_{n+1}^2$. Theorem 4.1 provides the general formula for $\hat{\sigma}_{n+h}^2$, when $h > 1$.

Theorem 4.1. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a stationary* SFIEGARCH$(p, d, q)_s$ *process with $\mathbb{E}(\sigma_t^2) < \infty$. Then, for any fixed $n \in \mathbb{Z}$, the h-step ahead forecast of σ_{n+h}^2 (consequently, for X_{n+h}^2), given $\mathcal{F}_n$, can be expressed as*

$$\hat{\sigma}_{n+h}^2 = e^{\omega} \prod_{k=h-1}^{\infty} \exp\left\{\lambda_{d,k} g(Z_{n+h-1-k})\right\} \prod_{\ell=0}^{h-2} \mathbb{E}\big(\exp\left\{\lambda_{d,\ell} g(Z_0)\right\}\big), \quad \textit{for all } h > 1. \tag{4.1}$$

Moreover, if $\mathbb{E}(\sigma_{n+h}^4) < \infty$, *the mean square errors of forecast for* σ_{n+h}^2 *and* X_{n+h}^2, *respectively denoted as* $mse(\sigma_{n+h}^2)$ *and* $mse(X_{n+h}^2)$, *are given by*

$$mse(\sigma_{n+h}^2) = \left[\prod_{k=0}^{h-2}\mathbb{E}\big(\exp\{2\lambda_{d,k}g(Z_0)\}\big) - \prod_{\ell=0}^{h-2}\Big[\mathbb{E}\big(\exp\{\lambda_{d,\ell}g(Z_0)\}\big)\Big]^2\right]\prod_{j=h-1}^{\infty}\mathbb{E}\big(\exp\{2\lambda_{d,j}g(Z_0)\}\big) \tag{4.2}$$

and

$$mse(X_{n+h}^2) = \mathbb{E}(\sigma_0^4)\big[\mathbb{E}(Z_0^4) - 1\big] + mse(\sigma_{n+h}^2), \quad \textit{for all } h > 1.$$

Lopes and Prass (2013) propose two h-step ahead predictors for the process $\{\sigma_t^2\}_{t\in\mathbb{Z}}$ in the context of FIEGARCH processes. The first one, denoted by $\check{\sigma}_{n+h}^2$, was obtained through the relation $\check{\sigma}_{n+h}^2 := \exp\{\hat{\ln}(\sigma_{n+h}^2)\}$, where $\hat{\ln}(\sigma_{n+h}^2)$ is the h-step ahead predictor for $\ln(\sigma_{n+h}^2)$. The second predictor, denoted by $\tilde{\sigma}_{n+h}^2$, was derived upon considering an order 2 Taylor's expansion of the exponential function. The authors also showed that $\check{\sigma}_{n+h}^2$ and $\tilde{\sigma}_{n+h}^2$ satisfy

$$\tilde{\sigma}_{n+h}^2 = \begin{cases} \exp\{\hat{\ln}(\sigma_{n+h}^2)\} = \check{\sigma}_{n+h}^2, & \text{if} \quad h = 1; \\ \exp\{\hat{\ln}(\sigma_{n+h}^2)\}\left[1 + \frac{1}{2}\sigma_g^2\sum_{k=0}^{h-2}\lambda_{d,k}^2\right] = \check{\sigma}_{n+h}^2\left[1 + \frac{1}{2}\sigma_g^2\sum_{k=0}^{h-2}\lambda_{d,k}^2\right], & \text{if} \quad h > 1. \end{cases} \tag{4.3}$$

With obvious identifications, the same predictors $\check{\sigma}_{n+h}^2$ and $\tilde{\sigma}_{n+h}^2$ can be defined for SFIEGARCH processes. However, by following the same steps as in Lopes and Prass (2013), it can be shown that both $\check{\sigma}_{n+h}^2$ and $\tilde{\sigma}_{n+h}^2$ are biased estimators for σ_{n+h}^2. On the other hand $\hat{\sigma}_{n+h}^2$, given in (4.1), not only is an unbiased estimator but also it is the best predictor for σ_{n+h}^2 in terms of the mean square error measure.

Now, to obtain the h-step ahead predictor for $\ln(X_{n+h}^2)$ observe that, from Definition 1.1 and from the i.i.d. property of $\{Z_t\}_{t\in\mathbb{Z}}$,

$$\hat{\ln}(X_{n+h}^2) := \mathbb{E}\big(\ln(X_{n+h}^2)|\mathcal{F}_n\big) = \mathbb{E}\big(\ln(\sigma_{n+h}^2)|\mathcal{F}_n\big) + \mathbb{E}\big(\ln(Z_{n+h}^2)|\mathcal{F}_n\big) := \hat{\ln}(\sigma_{n+h}^2) + \mathbb{E}(\ln(Z_0^2)),$$

for all $n \in \mathbb{Z}$ and $h > 0$. The expressions for $\hat{\ln}(\sigma_{n+h}^2) := \mathbb{E}(\ln(\sigma_{n+h}^2)|\mathcal{F}_t)$ and for the mean square errors of forecast for $\ln(\sigma_{n+h}^2)$ and $\ln(X_{n+h}^2)$ are given in Theorem 4.2.

Theorem 4.2. *Let* $\{X_t\}_{t\in\mathbb{Z}}$ *be an* SFIEGARCH$(p,d,q)_s$ *process. Then, for any fixed* $n \in \mathbb{Z}$, *the* h*-step ahead forecast* $\hat{\ln}(\sigma_{n+h}^2)$ *of* $\ln(\sigma_{n+h}^2)$, *given* $\mathcal{F}_n$, *can be expressed as*

$$\hat{\ln}(\sigma_{n+h}^2) = \omega + \sum_{k=0}^{\infty}\lambda_{d,k+h-1}\,g(Z_{n-k}), \quad \textit{for all } h > 0. \tag{4.4}$$

Moreover, if $h = 1$, *the mean square errors of forecast for* $\ln(\sigma_{n+h}^2)$ *and* $\ln(X_{n+h}^2)$ *are both equal to zero and, for any* $h > 1$, *are given, respectively, by*

$$mse(\ln(\sigma_{n+h}^2)) = \sigma_g^2\sum_{k=0}^{h-2}\lambda_{d,k}^2 \quad \textit{and} \quad mse(\ln(X_{n+h}^2)) = mse(\ln(\sigma_{n+h}^2)), \tag{4.5}$$

where $\sigma_g^2 = \mathbb{E}\big([g(Z_0)]^2\big)$.

Remark 4.2. Lopes and Prass (2013) consider the h-step ahead predictor for $\ln(X_{n+h}^2)$ defined as $\check{\ln}(X_{n+h}^2) = \hat{\ln}(\sigma_{n+h}^2)$, for any $n \in \mathbb{Z}$ and $h > 0$. This is an unbiased estimator if and only if $\mathbb{E}(\ln(Z_0^2)) = 0$.

From (4.1) and (4.4), one concludes that, $\hat{\sigma}_{n+1}^2 = \check{\sigma}_{n+1}^2 = \tilde{\sigma}_{n+1}^2$ and

$$\hat{\sigma}_{n+h}^2 = \check{\sigma}_{n+h}^2\prod_{\ell=0}^{h-2}\mathbb{E}\big(\exp\{\lambda_{d,\ell}g(Z_0)\}\big) = \tilde{\sigma}_{n+h}^2\left[1 + \frac{1}{2}\sigma_g^2\sum_{k=0}^{h-2}\lambda_{d,k}^2\right]^{-1}\prod_{\ell=0}^{h-2}\mathbb{E}\big(\exp\{\lambda_{d,\ell}g(Z_0)\}\big), \tag{4.6}$$

for any fixed $n \in \mathbb{Z}$ and $h > 1$.

Now, let $\{r_t\}_{t\in\mathbb{Z}}$ be the stochastic process defined by

$$r_t = \mu + \sum_{k=0}^{\infty} \psi_k X_{t-k} := \mu + \psi(\mathcal{B})X_t, \quad \text{for all } t \in \mathbb{Z}, \tag{4.7}$$

where $\mu \in \mathbb{R}$, $\{\psi_k\}_{t\in\mathbb{Z}}$ is a sequence of real numbers satisfying $\sum_{k=0}^{\infty} \psi_k^2 < \infty$ and $\{X_t\}_{t\in\mathbb{Z}}$ is an SFIEGARCH$(p,d,q)_s$ with $\sup_{t\in\mathbb{Z}}\{\mathrm{Var}(X_t^2)\} < \infty$. Notice that, $\sum_{k=0}^{\infty} \psi_k^2 < \infty$ and $\sup_{t\in\mathbb{Z}}\{\mathrm{Var}(X_t^2)\} < \infty$ imply $\mathrm{Cov}(X_k, X_j) = 0$, for all $k \neq j$, and

$$\mathbb{E}\left(\left|\sum_{k=m}^{n} \psi_k X_{t-k}\right|^2\right) \leq \sup_{t\in\mathbb{Z}}\{\mathrm{Var}(X_t^2)\} \sum_{k=m}^{n} \psi_k^2, \quad \text{for all } m \leq n.$$

Therefore, (4.7) converges in L^2 (Cauchy convergence criterion) and hence $\{r_t\}_{t\in\mathbb{Z}}$ is well defined.

Example 4.1. Let $\phi(\cdot)$ and $\varphi(\cdot)$ be the polynomials of order p and q, with no common roots, respectively defined by $\phi(z) := \sum_{k=0}^{p_1}(-\phi_k)z^k$ and $\varphi(z) := \sum_{j=0}^{q_1}(-\varphi_j)z^j$, with $\phi_0 = \varphi_0 = -1$. Given a weakly stationary SFIEGARCH$(p,d,q)_s$ process $\{X_t\}_{t\in\mathbb{Z}}$, define $\{r_t\}_{t\in\mathbb{Z}}$ by

$$\phi(\mathcal{B})(r_t - \mu) = \varphi(\mathcal{B})X_t, \quad \text{for all } t \in \mathbb{Z}.$$

Observe that $\{r_t\}_{t\in\mathbb{Z}}$ is an ARMA(p_1, q_1) process and hence it can be rewritten as in equation (4.7), whenever $\phi(z) \neq 0$ in the closed disk $\{z : |z| \leq 1\}$, with $\{\psi_k\}_{k\in\mathbb{Z}}$ uniquely defined through

$$\sum_{k=0}^{\infty} \psi_k z^k = \psi(z) := \frac{\varphi(z)}{\phi(z)}, \quad |z| \leq 1.$$

Theorem 4.3 provides the h-step ahead forecast for the process $\{r_t^2\}_{t\in\mathbb{Z}}$, with r_t defined in (4.7). Similar equations can be derived if the assumption that $\{X_t\}_{t\in\mathbb{Z}}$ is an SFIEGARCH$(p,d,q)_s$ is replaced by any other ARCH-type model. This result is applied in Section 5 to compare the forecasting performance of the different models considered in the time series analysis.

Theorem 4.3. *Let $\{r_t\}_{t\in\mathbb{Z}}$ be defined by (4.7). Then, for any fixed $n \in Z$, the h-step ahead forecast $\hat{r}_{n+h}^2$ of r_{n+h}^2, given $\mathcal{F}_n$, can be expressed as*

$$\hat{r}_{n+h}^2 = \mu^2 + \sum_{k=0}^{h-1} \psi_k^2 \hat{\sigma}_{n+h-k}^2 + \sum_{j=h}^{\infty}\sum_{\ell=h}^{\infty} \psi_j\psi_\ell X_{n+h-j}X_{n+h-\ell} + 2\mu\sum_{i=h}^{\infty} \psi_i X_{n+h-i}, \quad \textit{for any } h > 0, \tag{4.8}$$

where $\hat{\sigma}_{n+1}^2 = \sigma_{n+1}^2$ is given by (1.2) and $\hat{\sigma}_{n+h}^2$ is the h-step ahead forecast of X_{n+h}^2 given in (4.1), for all $h > 1$.

5 An Application

In this section we analyze the behavior of the intraday volatility of the S&P500 US stock index log-return time series in the period from December 13, 2004 to October 10, 2009. The trading hours are from 8:30 am to 3:15 pm (Chicago time) and the intraday frequency of the original index time series is 15 minutes, which gives a total of 33993 observations (1259 days). The fifteen-minute log-returns (see Remark 5.1) are aggregated to obtain a one-hour[1] log-return time series $\{R_t\}_{t=1}^{8498}$.

[1] Up to this day, we only had access to the stock index time series with sampling frequency equal to 15 minutes. Therefore, a thorough study on the existence of microstructural noise could not be performed. With the data set available we create a signature volatility plot (see, for instance, Andersen et al., 2000) by considering only sample frequencies which are multiples of 15 minutes (e.g., 15, 30 and 45 minutes). The minimum and maximum sample frequency values considered were, respectively, 15 minutes (which is the original sampling frequency) and 405 minutes (which corresponds to one-day log-returns). Under this scenario, the average realized volatilities were all close to each other regardless the sampling frequencies considered.

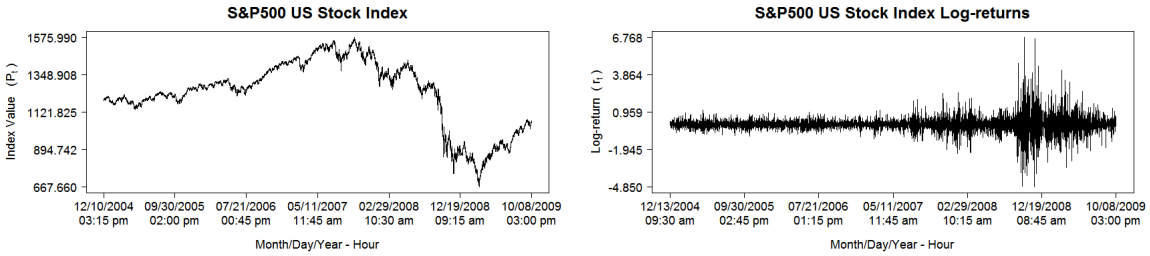
Remark 5.1. The fifteen-minutes log-return time series $\{R_t^{(15)}\}_{t=1}^n$ is defined as

$$R_t^{(15)} = 100 \times \ln\left(\frac{P_t}{P_{t-1}}\right), \quad \text{for all } t \in \{1, \cdots, n\},$$

where $\{P_t\}_{t=0}^n$ is the index time series, with $n = 33992$ ($t = 0$ corresponds to the first available observation). The one-hour log-return time series $\{R_t\}_{t=1}^{8498}$ is obtained by letting $R_t = R_{\tau-3}^{(15)} + \cdots + R_\tau^{(15)}$, with $\tau = 4t$, for all $t \in \{1, \cdots, 8498\}$.

Figure 5.1 (a) and (b) show, respectively, the S&P500 US stock index log-return time series and the one-hour log-return time series $\{R_t\}_{t=1}^{8498}$, in the studied period. Notice that, for any $t \in \{1, \cdots, 1259\}$ the times 3:15 pm (closing time) from day t and 8:30 am (opening time) from day $t+1$ are equivalent (there is no trading between these two periods). Therefore, there are 27 index values and, consequently, 27 available 15-minutes log-returns for each trading day. It is easy to see that, by applying the aggregation equation described in Remark 5.1, the trading time associated to one-hour log-returns for two consecutive days are not necessarily the same. In particular, the following holds

- the first available index value corresponds to December 13, 2004 8:30 am (or equivalently, December 12, 2004 3:15 pm). Consequently, the first one-hour log-return for day 1 corresponds to 9:30 am;
- the last one-hour log-return for day 1 corresponds to 2:30 pm and, since the trading day ends at 3:15 pm, the next one-hour log-return will be the aggregation of 15-minutes log-returns for 2:30 pm, 2:45 pm, 3:00 pm, 3:15 pm (or equivalently, 8:30 am from day 2) and 8:45 am from day 2. Consequently, the first one-hour log-return for day 2 corresponds to 8:45 am;
- whenever the first one-hour log-return corresponds to 9:30 am, there are only 6 one-hour log-returns for the corresponding day;
- for every four days there is one day with 6 one-hour log-returns followed by 3 consecutive days with 7 one-hour log-returns. This fact may or may not induce a cyclical behavior.



(a) (b)

Figure 5.1: This figure considers the S&P500 US stock index time series in the period from December 13, 2004 to October 10, 2009. The intraday frequency of the index time series is 15 minutes, which gives a total of 33993 observations (1259 days). Panel (a) shows the original index time series. Panel (b) presents the one-hour log-return time series, with $n = 8498$ observations, obtained by aggregating the fifteen-minute returns.

From Figure 5.1 (b) it is clear that the volatility in the period from 2004 to 2007 is much lower[2] than in the period from 2007 to 2009. Given that high volatility is more concerning and more difficult to model than low volatility we shall discard the first 3996 observations (592 days). We also reserve the last 40 days of data (270 observations) to analyze the out-of-sample forecasting

[2]This fact is already known in the literature and it is beyond the scope of this work to discuss possible causes for this behavior.

performance of the fitted models. The remaining time series corresponds to the period from March 21, 2007 to August 13, 2008 and has 4232 observations. This time series shall be denoted by $\{r_t\}_{t=1}^{4232}$, where $r_t := R_{t+3996}$, for all $t = 1, \cdots, 4232$.

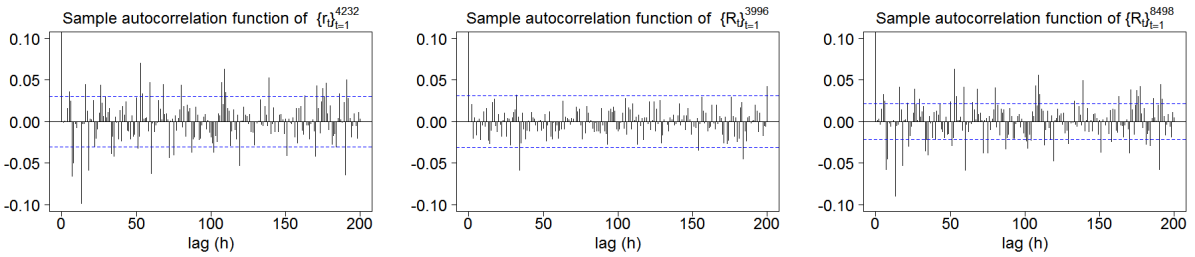
Remark 5.2. The two highest peaks in Figure 5.1 (a) correspond, respectively, to October 10, 2008 and November 21, 2008. The two lowest values in Figure 5.1 (a) correspond, respectively, to October 06, 2008 and November 20, 2008. October 10, 2008 is the day with the highest trading volume ever for the S&P 500 index. In this day, the trading volume for the SPY SPDR surpassed 871 million shares (see, for instance, AMEX:SPY daily prices for October 2008 from Yahoo! Finance). On November 20, 2008 the S&P 500 index closed at 752.44, its lowest since early 1997.

The descriptive statistics for the log-return time series $\{r_t\}_{t=1}^{4232}$ are given in Table 5.1. For comparison, this table also shows the descriptive statistics for time series $\{R_t\}_{t=1}^{8498}$. From Table 5.1 one observes that both time series $\{R_t\}_{t=1}^{8498}$ and $\{r_t\}_{t=1}^{4232}$ have mean approximately equal to zero but high skewness values, which usually indicates a non-symmetric distribution. However, we shall notice that, for these time series, the high skewness values could be due to the presence of some outliers instead of non-symmetry. In fact, upon replacing all values higher than eight standard deviations by the sample mean of the corresponding time series, the skewness values for $\{R_t\}_{t=1}^{8498}$ and $\{r_t\}_{t=1}^{4232}$ are, respectively, 0.0398 and -0.0018, which reinforces our claim. Nevertheless, the possible outliers are not removed in the analysis to be performed in the sequel. The aim of this approach is to observe whether the SFIEGARCH model captures or not this feature.

Table 5.1: Descriptive statistics for the S&P500 one-hour log-return time series $\{R_t\}_{t=1}^{8498}$ and for the time series $\{r_t\}_{t=1}^{4232}$, where $r_t := R_{t+3996}$, for all $t \in \mathbb{Z}$.

Period	n	mean	st. dev.	kurtosis	skewness
2007 - 2009	4232	-0.0078	0.6931	14.5986	0.4156
2004 - 2009	8498	-0.0013	0.5168	23.6251	0.4609

Note: st. dev. := standard deviation.



(a) (b) (c)

Figure 5.2: This figure shows the sample autocorrelation function for the S&P500 US stock index one-hour log-return time series $\{R_t\}_{t=1}^{8498}$ and its sub-samples. For $\{R_t\}_{t=1}^{8498}$ the time index $t = 1$ corresponds to December 13, 2004 at 09:30 am, $t = 3997$ corresponds to March 21, 2007 at 09:30 am and $t = 8498$ corresponds to October 10, 2009 at 03:00 pm. Panel (a) considers the time series $\{r_t\}_{t=1}^{4232}$ with $r_t := R_{t+3996}$. Panel (b) corresponds to the sub-sample $\{R_t\}_{t=1}^{3996}$. Panel (c) shows the sample autocorrelation function of $\{R_t\}_{t=1}^{8498}$. For better visualization the y-axis only shows the interval $[-0.10, 0.10]$. In all graphs the dashed lines correspond to $\pm 1.96/\sqrt{n}$, where n is the sample size of the corresponding time series.

Figure 5.2 (a), (b) and (c) shows, respectively, the sample autocorrelation functions for the log-return time series $\{r_t\}_{t=1}^{4232}$, $\{R_t\}_{t=1}^{3996}$ and $\{R_t\}_{t=1}^{8498}$. From this figure we observe that the log-return time series presents small (notice the scales) autocorrelations (but significatively different from zero) for some lags $h > 0$. Upon comparing Figure 5.2 (a), (b) and (c), one observes that while the sample autocorrelation functions for $\{r_t\}_{t=1}^{4232}$ and $\{R_t\}_{t=1}^{8498}$ seem identical, the difference is remarkable when

considering the sample autocorrelation functions for $\{R_t\}_{t=1}^{3996}$ and $\{R_t\}_{t=1}^{8498}$. This indicates that the correlation in the series is mainly due to the last observations of the time series $\{R_t\}_{t=1}^{8498}$.

From Figures 5.2 (a) and (c) one observes that the autocorrelation value with higher magnitude is associated to the lag 13 (roughly 2 days). Next, in order of magnitude (including the values not reported in Figure 5.2), are the autocorrelations associated to lags 53, 7, 263, 190, 109, 60, 18 (roughly 2 and a half days), 203, 218, 119, 139, 191 and 8. The remaining autocorrelation values are all smaller (in magnitude) than the one associated to lag 8 and, therefore, very close to the confidence limits (this includes the autocorrelation values with lag higher than 200, which are not reported in Figure 5.2). Recall that the aggregation rule considered implies that for every 4 days there is one day with only 6 one-hour log-returns followed by 3 days with 7 one-hour log-returns. Moreover, notice that 53, 109, 139, 190, 191, 218 are very close to multiples of 27 (total number of observations in 4 days). Furthermore, while 60 and 263 differ from multiples of 27 by approximately 7, 119 and 203 differ from approximately 13.

The facts just mentioned, indicate a short-memory cyclical behavior of length 27. On the other hand, there is also evidence that a single seasonal polynomial may not be enough to remove the correlation. For this reason we shall consider a constrained[3] ARMA(p_1, q_1) model for the log-return time series. Under this assumption we have

$$\phi(\mathcal{B})(r_t - \mu) = \varphi(\mathcal{B})X_t, \quad \text{for all } t \in \mathbb{Z}, \tag{5.1}$$

where $\mu \in \mathbb{R}$, $\phi(z) = \sum_{k=0}^{p_1}(-\phi_k)z^k$, $\varphi(z) = \sum_{j=0}^{q_1}(-\varphi_j)z^j$, with $\phi_0 = \varphi_0 = -1$, and $\{X_t\}_{t\in\mathbb{Z}}$ is a white noise process. Notice that, by letting p_1 and q_1 be large enough, equation (5.1) also covers the seasonal ARMA class of model, denoted by SARMA$(p_1, q_1) \times (P, Q)_s$ (see Remark 5.3).

Remark 5.3. For any $d, D \in \mathbb{N}$, let $(1-\mathcal{B})^d$ and $(1-\mathcal{B}^s)^D$ be, respectively, the non-seasonal and seasonal difference operators iterated, respectively, d and D times. Let $A(z) = \sum_{k=0}^{P}(-A_k)z^k$, $a(z) = \sum_{k=0}^{p}(-a_k)z^k$, $M(z) = \sum_{k=0}^{Q}(-M_k)z^k$ and $m(z) = \sum_{k=0}^{q}(-m_k)z^k$ be polynomials, respectively, of order P, p, Q and q, with $A_0 = a_0 = M_0 = m_0 = -1$. A seasonal autoregressive integrated moving average model, denoted by SARIMA$(p, d, q) \times (P, D, Q)_s$, is defined by (for more details and for the definition of a SARFIMA process, see Bisognin and Lopes, 2009)

$$A(\mathcal{B}^s)a(\mathcal{B})(1-\mathcal{B})^d(1-\mathcal{B}^s)^D(Y_t - \mu) = M(\mathcal{B}^s)m(\mathcal{B})\varepsilon_t, \quad \text{for all } t \in \mathbb{Z}, \tag{5.2}$$

where $\mu \in \mathbb{R}$ and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a white noise process with zero mean and variance σ_ε^2. In particular, when $d = D = 0$, (5.2) is called a SARMA$(p, q) \times (P, Q)_s$ model. It is immediate that by letting $d = D = 0$, $\phi(z) = A(\mathcal{B}^s)a(\mathcal{B})$ and $\varphi(z) = M(\mathcal{B}^s)m(\mathcal{B})$, (5.2) can be rewritten as an ARMA(p_1, q_1) model, given in (5.1), with $p_1 = P + p$ and $q_1 = Q + q$.

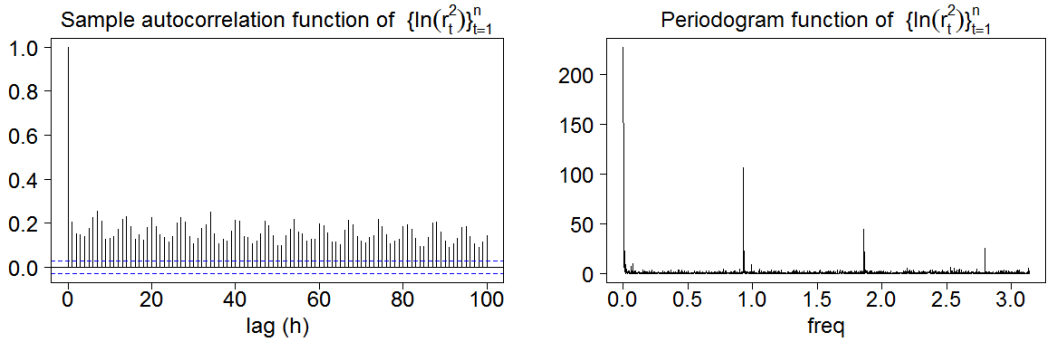
Figure 5.3 (a) and (b) present, respectively, the sample autocorrelation and the periodogram functions for the time series $\{\ln(r_t^2)\}_{t=1}^{4232}$. We observe that both functions indicate long-memory and cyclical behavior, with seasonal parameter $s = 7$ (one day cycle). To account for the long-memory cyclical behavior in the volatility, we shall consider an SFIEGARCH$(p_2, d, q_2)_s$, described in Section 1. To confirm the importance of including the seasonal effect associated to long-memory[4] we also consider a FIEGARCH(p_2, d, q_2) (Bollerslev and Mikkelsen, 1996), and an EGARCH(p_2, q_2) (Nelson, 1991) model.

Recall that, for all models just mentioned, $\{X_t\}_{t\in\mathbb{Z}}$ is written as $X_t = \sigma_t Z_t$, for any $t \in \mathbb{Z}$, where $\{Z_t\}_{t\in\mathbb{Z}}$ is an i.i.d. sequence of random variables with $\mathbb{E}(Z_0) = 0$ and $\mathrm{Var}(Z_0) = 1$. For the SFIEGARCH$(p_2, d, q_2)_s$ model one has

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

[3] By constrained we mean that some ϕ_i and θ_j will be fixed as zero.

[4] The FIEGARCH model captures non-seasonal long-memory and short-memory cyclical behaviors (if p_2 and q_2 are large enough). The EGARCH model is only able to describe short-memory cyclical behavior.



(a) (b)

Figure 5.3: This figure considers the time series $\{\ln(r_t^2)\}_{t=1}^{4232}$, where $\{r_t\}_{t=1}^{4232}$ is the S&P500 US stock index log-return time series corresponding to the period from March 21, 2007 at 09:30 am to August 13, 2009 at 03:00 pm. Panel (a) shows the sample autocorrelation function of $\{\ln(r_t^2)\}_{t=1}^{4232}$. Panel (b) presents its periodogram function.

with $\alpha(\cdot)$, $\beta(\cdot)$ and $(1-\mathcal{B}^s)^{-d}$ be defined as in (1.4) and (1.5), and $g(Z_t) = \theta Z_t + \gamma\big[|Z_t| - \mathbb{E}\big(|Z_t|\big)\big]$. The FIEGARCH and EGARCH models are particular cases of the SFIEGARCH model obtained from (1.2), respectively, when $s = 1$ and $d = 0$.

Remark 5.4. By comparing the sample kurtosis values given in Table 5.1 with the theoretical kurtosis values of an SFIEGARCH$(p, d, q)_s$ process (see Figures 2.4 and 2.5) we conclude that the best SFIEGARCH fit for the data more likely will have $p, q > 0$.

Remark 5.5. Our first intention was to compare the performance of the SFIEGARCH model with the PLM-GARCH (Bordignon et al., 2007, 2009), since both models are able to describe long-memory cyclical behavior. Analogously to the SFIEGARCH case, upon considering the PLM-GARCH we would also include a HYGARCH (Davidson, 2004) and a GARCH model (Bollerslev, 1986). It turns out that we were not able to fit any PLM-GARCH model for which the squared residuals time series $\{\hat{z}_t^2\}_{t=1}^n$ shows no correlation and at the same time the positivity criteria for $\{\sigma_t^2\}_{t=1}^n$ would be satisfied. The number of cases for which $\sigma_t^2 < 0$ were always too high to be replaced by a constant or by $|\sigma_t^2|$.

5.1 Model Selection Procedure

Parameter estimation is carried out by applying the so-called quasi-likelihood method. In this method, the estimate $\hat{\boldsymbol{\eta}}$ for the vector of unknown parameters $\boldsymbol{\eta}$ is the value that maximizes

$$\mathcal{L}_n(\boldsymbol{\eta}) := -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\left[\ln(\sigma_t^2) + \frac{(r_t-\mu_t)^2}{\sigma_t^2}\right],$$

where $\mu_t := \mu + \sum_{k=1}^{p_1}\phi_k(r_{t-k}-\mu) + \sum_{j=1}^{q_1}\varphi_j X_{t-j}$, for all $t \in \mathbb{Z}$. The recursion starts by setting $r_t = \bar{r}$, where $\bar{r} = \frac{1}{n}\sum_{t=1}^{n} r_t$ is the sample mean of the log-return time series, and $X_t = g(Z_t) = 0$ and $X_t^2 = \sigma_X^2$, whenever $t \leq 0$, where σ_X^2 is the sample variance of $\{X_t\}_{t=1}^n$.

Since we intent to compare the performance of the SFIEGARCH model with other ARCH-type models which incorporate or not seasonal long-memory cyclical behavior in the volatility, we shall consider a two step estimation procedure. First an ARMA(p_1, q_1) model for the one-hour log-return time series is selected. The second step consists on fitting an SFIEGARCH$(p_2, d, q_2)_s$ (or any other ARCH-type) model to the residuals of the ARMA model.

The information obtained through the analysis of the autocorrelation function of $\{r_t\}_{t=1}^{4232}$ is applied to select the orders p_1 and q_1. Given the large number of possible combinations of p_1 and

q_1 we restrict our attention to ARMA$(263,0)$ and ARMA$(0,263)$ models with $\phi_i = \theta_j = 0$, whenever $i,j \notin \{7,8,13,18,53,60,109,119,139,190,191,203,218,263\}$.

In order to select an SFIEGARCH$(p_2,d,q_2)_s$ model for the residuals we fix $s=7$ (estimated from the periodogram and sample autocorrelation functions) and consider different combinations of p_2 and q_2. Once the right combination of p_2 and q_2 is found, the same values are consider to select the FIEGARCH and the EGARCH models.

The following criteria applies to both estimation steps.

1. For any combination of p_1, q_1 or p_2, q_2, we start with the full model and remove the non-significant parameters (one at a time) until all p-values are smaller than 0.05.

2. The standard deviations for the model parameters were obtained by considering the robust covariance matrix given by $n^{-1}H^{-1}(\hat{\boldsymbol{\eta}})B(\hat{\boldsymbol{\eta}})H^{-1}(\hat{\boldsymbol{\eta}})$, where $n^{-1}H(\hat{\boldsymbol{\eta}})$ and $n^{-1}B(\hat{\boldsymbol{\eta}})$ are, respectively, the Hessian and the outer product of the gradients (see Bollerslev and Wooldridge, 1992).

3. A model is considered to fit the data well if $\{\hat{x}_t\}_{t=1}^n$ (the residual of the ARMA model), $\{\hat{z}_t\}_{t=1}^n$ (the residual of the SFIEGARCH model) and $\{\hat{z}_t^2\}_{t=1}^n$ show no significant correlation. To test for correlation we consider both the Box-Pierce and Ljung-Box hypothesis tests (see Remark 5.6).

4. When more than one model satisfy the criteria in Step 3, model selection is performed based on the values of the log-likelihood, AIC, BIC and HQC criteria, obtained in Step 2.

5. In case two or more models present similar AIC, BIC, HQC and/or log-likelihood criteria values, we chose the more parsimonious one.

Remark 5.6. When applying the Box-Pierce (or the Ljung-Box) hypothesis test, if the null hypothesis is rejected for $\{\hat{x}_t\}_{t=1}^n$ but it is not reject for both $\{\hat{z}_t\}_{t=1}^n$ and $\{\hat{z}_t^2\}_{t=1}^n$, the cumulative periodogram (also known as Kolmogorov-Smirnov hypothesis test) is considered. If this test does not reject the null hypothesis that $\{\hat{x}_t\}_{t=1}^n$ is a white noise process, the model is not discarded.

Further residuals analysis is performed by following the same approach as Haas et al. (2004). The procedure consists on employing a density transformation, as presented in Diebold et al. (1998), to test the assumption $\hat{x}_t|\mathcal{F}_{t-1} \sim F_t$, for some given target distribution F_t.

5.2 Forecasting Procedure

Once the parameters of the ARMA(p_1,q_1)-SFIEGARCH$(p_2,d,q_2)_s$ model are estimated, out-of-sample forecasting is performed. To obtain the predicted values $\hat{r}_{n+h}$, $\hat{\sigma}^2_{n+h}$ and $\hat{x}^2_{n+h}$, given $\{r_t\}_{t=1}^n$, with $n=4232$ and $h \in \{1,\cdots,270\}$, we proceed as described in steps **1** - **9** below. We shall denote by $\boldsymbol{\eta}$ the true parameters, namely,

$$\boldsymbol{\eta} = (d,\theta,\gamma,\omega,\mu,\phi_1,\cdots,\phi_{p_1},\varphi_1,\cdots,\varphi_{q_1},\alpha_1,\cdots,\alpha_{p_2},\beta_{q_1},\cdots,\beta_{q_2})',$$

and by $\hat{\boldsymbol{\eta}}$ the estimated values. With obvious identifications, the forecasting considering the other ARCH-type models is analogous.

1. The true parameters values $(d,\alpha_1,\cdots,\alpha_{p_2},\beta_1,\cdots,\beta_{q_2})'$ are replaced by the estimated ones, namely, $(\hat{d},\hat{\alpha}_1,\cdots,\hat{\alpha}_{p_2},\hat{\beta}_1,\cdots,\hat{\beta}_{q_2})'$, and the recurrence formula given in Proposition 1.1 is used to calculate the corresponding coefficients $\{\hat{\lambda}_{d,k}\}_{k=0}^{n+270}$. Notice that, with obvious identifications, this recurrence formula can be also used to calculate the coefficients $\{\hat{\psi}_{d,k}\}_{k=0}^{n+270}$ associated to $\hat{\phi}_1,\cdots,\hat{\phi}_{p_1}$, $\hat{\varphi}_1,\cdots,\hat{\varphi}_{q_1}$.

2. Upon setting $r_t = \hat{\mu}$ and $\hat{x}_t = g(\hat{z}_t) = 0$, whenever $t < 0$, the time series $\{\hat{x}_t\}_{t=1}^n$, $\{\hat{\sigma}_t\}_{t=1}^n$ and $\{\hat{z}_t\}_{t=1}^n$ are obtained recursively as follows[5]

$$\hat{x}_t = \sum_{k=0}^{p_1} \phi_k (r_{t-k} - \hat{\mu}) + \sum_{j=1}^{q_1} \varphi_j \hat{x}_{t-j},$$
$$\hat{\sigma}_t = \exp\left\{ \frac{\hat{\omega}}{2} + \frac{1}{2} \sum_{k=0}^{n-1} \hat{\lambda}_{d,k} \left[\hat{\theta} z_{t-1-k} + \hat{\gamma}\left(|\hat{z}_{t-1-k}| - \sqrt{2/\pi} \right) \right] \right\} \quad \text{and} \quad \hat{z}_t = \frac{\hat{x}_t}{\sigma_t},$$

for all $t \in \{1, \cdots, n\}$. Note that, in particular, $\hat{x}_1 = r_1 - \hat{\mu}$, $\hat{\sigma}_1 = e^{\hat{\omega} 0.5}$ and $\hat{z}_1 = \hat{x}_1 \hat{\sigma}_1^{-1}$.

3. Since the h-step ahead predictor for X_{n+h} given $\mathcal{F}_n$ is zero, it is set $\hat{x}_{n+h} = 0$, for all $h > 0$.

4. The h-step ahead forecast $\hat{r}_{n+h}$ for r_{n+h} is given by (see, for instance Brockwell and Davis, 1991)

$$\hat{r}_{n+h} = \hat{\mu} + \sum_{k=1}^{p_1} \phi_k (\hat{r}_{n+h-k} - \hat{\mu}) + \sum_{j=1}^{q_1} \varphi_j \hat{x}_{n+h-j}, \qquad \text{with } \hat{r}_t = r_t, \text{ if } t \leq n. \tag{5.3}$$

5. An estimate $\hat{\sigma}_g^2$ for σ_g^2 is obtained by replacing $\mathbb{E}(|Z_0|)$ and $\mathbb{E}(Z_0|Z_0|)$, in expression (1.14), by their respective sample estimates, that is,

$$\hat{\sigma}_g^2 = \hat{\theta}^2 + \hat{\gamma}^2 - (\hat{\gamma}\hat{\mu}_{|z|})^2 + 2\,\hat{\theta}\,\hat{\gamma} \left[\frac{1}{n} \sum_{t=1}^n \hat{z}_t |\hat{z}_t| \right], \quad \text{with} \quad \hat{\mu}_{|z|} := \frac{1}{n} \sum_{t=1}^n |\hat{z}_t|.$$

6. An estimative for $E_{\boldsymbol{\eta}}(\ell) := \mathbb{E}(\exp\{\lambda_{d,\ell} g(Z_0)\})$, given in (4.1), is obtained by considering the respective sample estimator

$$\hat{E}_{\hat{\boldsymbol{\eta}}}(\ell) = \frac{1}{n} \sum_{t=1}^n \exp\left\{ \hat{\lambda}_{d,\ell} \left[\hat{\theta} z_t + \hat{\gamma}(|\hat{z}_t| - \hat{\mu}_{|z|}) \right] \right\}, \quad \text{for any} \quad \ell \in \{0, \cdots, h-2\}.$$

7. Since σ_{n+1}^2 is $\mathcal{F}_n$-measurable, $\hat{\sigma}_{n+1}^2 = \sigma_{n+1}^2$ and it is computed as in step **2**.

8. The predictor $\breve{\sigma}_{n+h}^2$ is obtained upon replacing the true parameter values by the estimated ones in (4.4), with the additional assumption $g(\hat{z}_t) = 0$, if $t < 0$. Then, from (4.6), $\hat{\sigma}_{n+h}^2$ and $\tilde{\sigma}_{n+h}^2$ are obtained by setting, respectively,

$$\hat{\sigma}_{n+h}^2 = \breve{\sigma}_{n+h}^2 \prod_{\ell=0}^{h-2} \hat{E}_{\hat{\boldsymbol{\eta}}}(\ell) \quad \text{and} \quad \tilde{\sigma}_{n+h}^2 = \tilde{\sigma}_{n+h}^2 \left[1 + \frac{1}{2} \hat{\sigma}_g^2 \sum_{k=0}^{h-2} \hat{\lambda}_{d,k}^2 \right]^{-1} \prod_{\ell=0}^{h-2} \hat{E}_{\hat{\boldsymbol{\eta}}}(\ell), \quad \text{for any} \quad h > 2.$$

9. The predictor $\hat{r}_{n+h}^2$ is obtained through (4.8), with the additional assumption $\hat{x}_t = 0$, if $t < 0$.

Remark 5.7. In the literature, the time series $\{\hat{\mu}_t\}_{t=1}^n$, given by,

$$\hat{\mu}_t = \sum_{k=1}^{p_1} \phi_k (r_{t-k} - \hat{\mu}) + \sum_{j=1}^{q_1} \varphi_j \hat{x}_{t-j}, \quad \text{for } t \in \{1, \cdots, n\},$$

is called *fitted values* or *in-sample forecasts*[6]. Consequently, the residuals time series $\{\hat{x}_t\}_{t=1}^n$ is also denoted *in-sample errors of forecast*. Furthermore, since $\hat{z}_t = \hat{x}_t \hat{\sigma}_t^{-1}$, for all $t \in \{1, \cdots, n\}$, the time series $\{\hat{z}_t\}_{t=1}^n$ is often called *standardized residuals*.

[5] If the pseudo-likelihood is used instead of the quasi-likelihood, the value $\sqrt{2/\pi}$ must be replace by the value of $\mathbb{E}(|Z_0|)$ associated to the distribution considered in the estimation procedure.

[6] From (5.3) it is clear that $\hat{\mu}_t$ is the 1-step ahead forecast for r_t, given $\mathcal{F}_{t-1}$, for any $t > 0$.

5.3 Forecasting Performance and Models Comparison

Without loss of generality, let $\{\hat{y}_t\}_{t=t_0}^{t_0-1+n_p}$ denote either the in-sample ($t_0 = 1$ and $n_p = 4232$) or the out-of-sample ($t_0 = 4233$ and $n_p = 270$) forecast values corresponding to the time series $\{y_t\}_{t=t_0}^{t_0-1+n_p}$, where y_t is either r_t or r_t^2 and n_p is the number of predicted values. Denote by $\mathcal{M}$ any model used to obtain $\{\hat{y}_t\}_{t=t_0}^{t_0-1+n_p}$. The forecasting performance of model $\mathcal{M}$ is evaluated by computing the mean absolute error (mae), the mean percentage error (mpe) and the maximum absolute error (max_{ae}) measures, namely,

$$mae(\mathcal{M}) = \frac{1}{n_p}\sum_{t=1}^{n_p}|e_t|, \quad mpe(\mathcal{M}) := \frac{1}{n_p}\sum_{t=1}^{n_p}\frac{|e_t|}{|y_t|} \quad \text{and} \quad max_{ae}(\mathcal{M}) := \max_{t\in\{1,\cdots,n_p\}}\{|e_t|\}$$

where $e_t := y_t - \hat{y}_t$ denotes the forecasting error at step t. The statistical significance of the out-of-sample forecasting performance is evaluated by using the so called Diebold and Mariano hypothesis test (see Diebold and Mariano, 1995).

Remark 5.8. The mpe is an interesting measure since it considers not only the magnitude of the error (as does the mae) but also the proportion between the error and the true values so it is easier to decide whether the error is small or not. A drawback of the mpe is that this measure is highly affected when observations are too close to zero.

The predictive performance of model $\mathcal{M}$ is also evaluated by measuring the quality of the one-step ahead density forecasts (see, for instance, Paolella, 2013). The measure used for this analysis is the normalized sum of the realized predictive log-likelihood, given by

$$S(\mathcal{M}) = \frac{1}{n_p}\sum_{t=n+1}^{n+n_p}\ln\big(f^{\mathcal{M}}_{t|\mathcal{F}_{t-1}}(y_t;\hat{\boldsymbol{\eta}})\big),$$

where $\{y_t\}_{t=1}^{n+n_p}$ is a sample from $\{Y_t\}_{t\in\mathbb{Z}}$, n is the size of the sample used to estimate the parameters for model $\mathcal{M}$, n_p is the number of predicted values, $f^{\mathcal{M}}_{t|\mathcal{F}_{t-1}}(\cdot;\hat{\boldsymbol{\eta}})$ denotes the conditional probability density function of Y_t given $\mathcal{F}_{t-1}$ and $\boldsymbol{\eta}$ is the parameter vector for model $\mathcal{M}$.

Forecast efficiency regressions (see Mincer and Zarnowitz, 1969) are also used to compare the quality of the volatility forecasts among the different models fitted to the data. The standard Mincer-Zarnowitz regression for forecast efficiency is given by

$$y_{t+h} = \gamma_0 + \gamma_1\hat{y}_{t+h} + \varepsilon_t, \tag{5.4}$$

where y_{t+h} is the variable of interest and $\hat{y}_{t+h}$ is the h-step ahead forecast for y_{t+h} given $\mathcal{F}_t$. Under the null hypothesis of forecast efficiency $\gamma_0 = 0$ and $\gamma_1 = 1$. The coefficients γ_0 and γ_1 in (5.4) are obtained by ordinary least square (OLS) estimation. The standard errors of the estimates are corrected for heteroskedasticity and autocorrelation by using the HAC estimator (see Newey and West, 1987). Since the forecasts $\hat{y}_{n+h}$ are obtained from a model $\mathcal{M}$ for which the true parameter values are unknown, the uncertainty concerning parameter estimation is corrected by multiplying the Newey-West standard errors by $\lambda = \sqrt{1 + n_p/n}$ (see West and McCraken, 1998), where n is the sample size used to fit the model and n_p is the number of predicted values.

Since the true volatility cannot be directly measured, the forecast efficiency regression is performed by considering the realized volatility instead. The ideas for this analysis were adapted from Klaassen (2002), where a slightly different definition[7] for the "observed volatility" was considered. The definitions adopted here are given below.

[7] Klaassen (2002) considers daily log-returns and the models fitted to the data do not include the ARMA regression. So, r_t was written as $r_t = \mu + X_t$, where X_t follows an ARCH-type model. In this case, the author replaced $\bar{y}_t$ by $\hat{\mu}$ in the definition of v_t. In our case, μ is replaced by μ_t, which may vary according to each ARCH-type model associated to X_t. Therefore we shall use the traditional definition of realized volatility to let v_t be model free.

Definition 5.1. Let $r_{(t-1)M+k}$ be the log-return value corresponding to the k-th period of day t, for $k \in \{1, \cdots, M\}$ and $t = 1, \cdots, N$, where M is the number of intraday periods and N is the number of observed days.

a) The daily log-return, denoted by $r_t^{(d)}$, is defined through $r_t^{(d)} = \sum_{k=1}^{M} r_{(t-1)M+k}$, for all $t \in \{1, \cdots, N\}$.

b) The daily realized volatility, denoted by v_t, is given by

$$v_t = \sum_{k=1}^{M} (r_{(t-1)M+k} - \bar{r}_t)^2, \quad \text{where} \quad \bar{r}_t := \frac{1}{M} \sum_{j=1}^{M} r_{(t-1)M+j}, \quad \text{for all } t \in \{1, \cdots, N\}.$$

c) The log-return over the period of h days, denoted by $r_{t-1}^{(d)}[h]$, is given by

$$r_{t-1}^{(d)}[h] = \sum_{j=0}^{h-1} r_{t+j}^{(d)} = \sum_{j=0}^{h-1} \left[\sum_{k=1}^{M} r_{(t+j-1)M+k} \right] = \sum_{k=1}^{hM} r_{(t-1)M+k}, \quad \text{for any } h \geq 1 \text{ and } t \in \{1, \cdots, N\}.$$

In particular, $r_{t-1}^{(d)}[1] = r_t^{(d)}$ so the log-return of period 1-day is simply the daily log-return.

d) The realized volatility over the period of h days, denoted by $v_{t-1}[h]$, is given by

$$v_{t-1}[h] = \sum_{j=0}^{h-1} \left[\sum_{k=1}^{M} (r_{(t+j-1)M+k} - \bar{r}_{t+j})^2 \right] = \sum_{j=0}^{h-1} v_{t+j}, \quad \text{for any } h \geq 1 \text{ and } t \in \{1, \cdots, N\}.$$

In particular, $v_{t-1}[h] = v_t$ so the realized volatility over the period of 1-day is simply the daily volatility.

By following the same steps as in the proof of proposition 2.2 in Prass and Lopes (2013), one can show that, if $\{r_t\}_{t \in \mathbb{Z}}$ follows an ARMA(p_1, q_1)-SFIEGARCH$(p_2, d, q_2)_s$ model then the forecast for the log-return over the period of h days, given the information up to day[8] $t-1$, is given by

$$\hat{r}_{t-1}^{(d)}[h] = \sum_{j=0}^{h-1} \hat{r}_{t+j}^{(d)} = \sum_{k=1}^{hM} \hat{r}_{(t-1)M+k}, \quad \text{for any } h \geq 1 \text{ and } t \in \{1, \cdots, N\}. \tag{5.5}$$

where $\hat{r}_{(t-1)M+k} = \mathbb{E}(r_{(t-1)M+k} | \mathcal{F}_{(t-1)M})$ is the k-step ahead forecast for $r_{(t-1)M+k}$, obtained from the ARMA model. Moreover, the forecast for the conditional variance of log-return over the period of h, given the information up to day $t-1$, namely $\sigma_{t-1}^{2\,(d)}[h] = \mathrm{Var}(r_{t-1}^{(d)}[h] | \mathcal{F}_{(t-1)M})$, is given by

$$\hat{\sigma}_{t-1}^{2\,(d)}[h] = \sum_{j=1}^{hM} \Psi_j \hat{\sigma}_{(t-1)M+j}^2, \quad \text{with } \Psi_j := \left[\sum_{k=0}^{hM-j} \psi_k \right]^2, \tag{5.6}$$

where $\hat{\sigma}_{(t-1)M+k}^2 = \mathbb{E}(X_{(t-1)M+k}^2 | \mathcal{F}_{(t-1)M})$ is the k-step ahead forecast for $X_{(t-1)M+k}^2$, obtained from the SFIEGARCH model. Equivalent result is derived upon replacing the SFIEGARCH by any other ARCH-type model.

Remark 5.9. Notice that Definition 5.1 and expressions (5.5) and (5.6) assume M constant. When M varies over time, similar results are derived upon making the following adjustments: replace $(t-1)M$ by the number of intraday log-returns available up to day $(t-1)$ (included); replace hM by the number of intraday log-returns over the period from t to $t+h$; replace $\sum_{k=1}^{M}$ and $1/M$, respectively, by $\sum_{k=1}^{M_t}$ (in Definition 5.1 c) and d), by $\sum_{k=1}^{M_{t+j}}$) and $1/M_t$, where M_t is the number of intraday log-returns available for day t.

Remark 5.10. For any $h \geq 1$ fixed, the forecast efficiency regression is obtained upon replacing, in (5.4), y_{t+1} and $\hat{y}_{t+1}$, respectively, by $v_t[h]$ and $\hat{\sigma}_t^{2(d)}[h]$, for $t = n, \cdots, N-h$, where N is the sample size of the log-return time series and n is the size of the sample used to fit the model.

[8] Note that, since we are considering intraday log-returns, the information up to day $t-1$ corresponds to $\mathcal{F}_{(t-1)M}$, where M is number of the intraday periods.

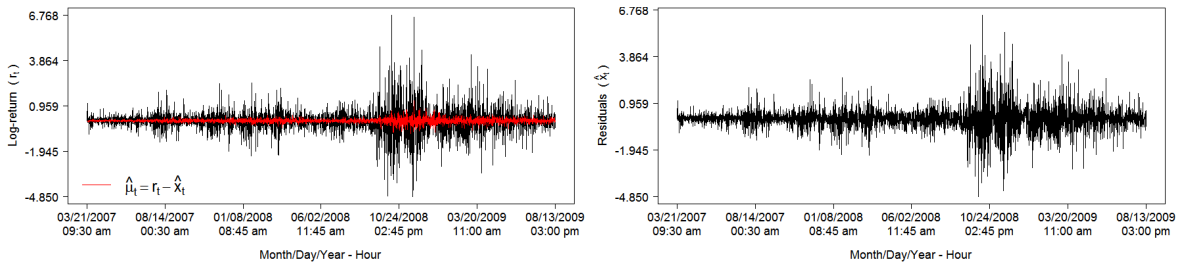
5.4 Results

Constrained ARMA$(263, 0)$ and ARMA$(0, 263)$ models, with $\phi_i = \theta_j = 0$, whenever $i, j \notin \{7, 8, 13, 18, 53, 60, 109, 119, 139, 190, 191, 203, 218, 263\}$, where analyzed. It turns out that several initial considered parameters where not significant and were removed from the models. The most parsimonious found model is given by (the number in parenthesis is the robust standard error)

$$\begin{aligned} r_t = X_t &- \underset{(0.0128)}{0.0391} X_{t-7} - \underset{(0.0169)}{0.0834} X_{t-13} + \underset{(0.0204)}{0.0803} X_{t-53} + \underset{(0.0215)}{0.0748} X_{t-109} - \underset{(0.0093)}{0.0433} X_{t-190} \\ &+ \underset{(0.0221)}{0.0742} X_{t-203} + \underset{(0.0167)}{0.0680} X_{t-218} - \underset{(0.0111)}{0.0421} X_{t-263} \\ &:= X_t + \mu_t, \end{aligned}$$

for all $t \in \{1, \cdots, 4232\}$, with $X_t = 0$, if $t \leq 0$.

The observed time series $\{r_t\}_{t=1}^{4232}$ and the corresponding fitted values $\{\hat{\mu}_t\}_{t=1}^{n}$, obtained from the ARMA model, are given in Figure 5.4 (a). Figure 5.4 (b) shows the residuals time series $\{\hat{x}_t\}_{t=1}^{4232}$. The p-values for the Box-Pierce and Ljung-Box test statistic for $\{\hat{x}_t\}_{t=1}^{n}$, that is, the residuals of the ARMA model, were smaller than 0.05 for any lag higher than 15. On the other hand, upon applying the cumulative periodogram test, the null hypothesis that $\{\hat{x}_t\}_{t=1}^{n}$ is white noise process, was not rejected (the cumulative periodogram figure was omitted to save space and may be obtained from the authors upon request). As expected, both the Box-Pierce (or Ljung-box) and the cumulative periodogram tests reject the null hypothesis that $\{\hat{x}_t^2\}_{t=1}^{n}$ is a white noise process.



(a) (b)

Figure 5.4: This figure shows the S&P500 US stock index log-return time series observed in the period from from March 21, 2007 at 9:30 am to August 13, 2009 at 03:00 pm and the corresponding fitted values and residuals obtained form the constrained ARMA model. Panel (a) gives the observed time series $\{r_t\}_{t=1}^{4232}$ (in black) and the fitted values $\{\hat{\mu}_t\}_{t=1}^{4232}$ (in red). Panel (b) shows the residuals time series $\{\hat{x}_t\}_{t=1}^{4232}$.

The estimated values and the corresponding standard errors for the parameters of considered ARCH-type models are given in Table 5.2. For any model in Table 5.2, the p-values for both the Box-Pierce and Ljung-Box test statistics corresponding to the time series $\{\hat{z}_t\}_{t=1}^{4232}$, that is, the residual of the ARCH-type model, were always higher than 0.05, for any lag $h > 0$. The same applies to the time series $\{\hat{z}_t^2\}_{t=1}^{4232}$. The histogram (or the kernel density) and the QQ-Plot (both omitted here to save space but available from the authors upon request) indicate that, although symmetric, the distribution of $\{\hat{z}_t\}_{t=1}^{4232}$ is not Gaussian.

Remark 5.11. The SFIEGARCH model given in Table 5.2 is the most parsimonious model such that $\beta(\cdot)$ has no roots in the closed disk $\{z : |z| \leq 1\}$. The p-value for α_3, in the FIEGARCH model, is 0.1141. On the other hand, the model fitted without this coefficient does not lead to uncorrelated residuals. The polynomial $\beta(\cdot)$ associated to the FIEGARCH model has two roots with absolute value 1.0023 and two other roots with absolute value 1.0024. Therefore, very close to the unit circle. Analogously, the polynomial $\beta(\cdot)$ associated to the EGARCH model has two roots with absolute value 1.0024 and another one with absolute value 1.0001. Despite this fact, the EGARCH model presents sligltly better performance in terms of log-likelihood, AIC, BIC and HQC criteria.

Table 5.2: Estimated values and the corresponding robust standard errors (in parenthesis) for the parameters of the SFIEGARCH, FIEGARCH and EGARCH models fitted to the S&P500 US stock index log-return time series observed in the period from March 21, 2007 at 9:30 am to August 13, 2009 at 03:00 pm. This table also presents the corresponding log-likelihood, AIC, BIC and HQC criteria values.

Parameter	SFIEGARCH	FIEGARCH	EGARCH
d	0.4532 (0.0104)	0.4529 (0.0057)	-
ω	-1.1810 (0.0129)	-1.6972 (0.0049)	-0.8601 (0.0058)
θ	-0.0820 (0.0115)	-0.1100 (0.0017)	-0.0954 (0.0016)
γ	0.2127 (0.0213)	0.2393 (0.0053)	0.2197 (0.0025)
α_1	0.1655 (0.0040)	0.0657 (0.0065)	-0.2718 (0.0015)
α_2	0.1963 (0.0263)	0.3021 (0.0088)	0.3761 (0.0019)
α_3	0.1821 (0.0045)	0.0151 (0.0095)	-0.2503 (0.0021)
α_4	-0.3095 (0.0048)	-0.4206 (0.0079)	-0.4603 (0.0039)
α_5	-0.4115 (0.0094)	-0.5962 (0.0136)	-0.5355 (0.0038)
α_6	-0.3139 (0.0112)	0.4580 (0.0023)	0.9040 (0.0019)
β_1	0.3151 (0.0095)	0.1378 (0.0039)	0.3271 (0.0013)
β_2	-0.0668 (0.0174)	0.1231 (0.0025)	0.0779 (0.0016)
β_3	0.2092 (0.0198)	-0.0081 (0.0030)	0.1545 (0.0025)
β_4	-0.3621 (0.0079)	-0.2854 (0.0042)	-0.4983 (0.0014)
β_5	0.0785 (0.0124)	-0.1842 (0.0020)	0.1895 (0.0023)
β_6	0.6534 (0.0115)	0.8896 (0.0029)	0.7389 (0.0026)
log-likelihood	-3053.9066	-2936.1820	-2878.2630
AIC	6139.8131	5904.3641	5786.5260
BIC	6241.4201	6005.9709	5881.7824
HQC	6175.7272	5940.2781	5820.1954

To apply the density transform procedure (for details, see Haas et al., 2004; Diebold et al., 1998) the GED distribution, with different values for the tail-thickness parameter ν. Under this scenario, the null hypothesis to be tested is

$$H_0 : \hat{x}_t|\mathcal{F}_{t-1} \sim \text{GED}(\nu, 0, \sigma_t) \quad \text{or, equivalently,} \quad H_0 : F_t^{-1}(\hat{x}_t) \sim \mathcal{U}(0,1),$$

where $\text{GED}(\nu, 0, \sigma_t)$ denotes the generalized error distribution with tail-thickness parameter ν, mean zero and standard deviation σ_t, and $F_t(\cdot)$ is the corresponding cumulative distribution function. In particular, when $\nu = 2$, we have the Gaussin distribution. The time series $\{x_t\}_{t=1}^{4232}$ corresponds to the residuals of the ARMA model fitted to the one-hour log-return time series $\{r_t\}_{t=1}^{4232}$ and $\{\hat{\sigma}_t^2\}_{t=1}^{4232}$ denotes the conditional variance of the log-returns, obtained from the SFIEGARCH, FIEGARCH or from the EGARCH model. Table 5.3 reports the results for the Kolmogorov-Smirnov (K-S) hypothesis test used to compare the sample $\{F^{-1}(\hat{x}_t)\}_{t=1}^{4232}$ with the uniform distribution.

Table 5.3 confirms the results obtained with the QQ-Plot, that is, $\{\hat{z}_t\}_{t=1}^{n}$ does have Gaussian distribution. This table also indicates that the assumption that $\{\hat{z}_t\}_{t=1}^{n}$ follows a $\text{GED}(\nu)$ distribution holds for more than one value of ν. The next step in this analysis would be to replace the QMLE by the log-likelihood estimation procedure, using the GED distribution, and estimate ν alongside with the other parameters. The information on the ν parameter could then be incorporated in the forecasting equation to see whether forecast efficiency improves or not. This analysis shall be performed in a future work.

The mean absolute error (mae), the mean percentage error (mpe) and the maximum absolute error (max_{ae}) measures for the selected models are reported in Tables 5.4 and 5.5. For the in-sample analysis, the mae, mpe and max_{ae} values were obtained by letting $e_t = r_t - \hat{\mu}_t$, for all $t \in \{1, \cdots, 4232\}$ (see Section 5.3). For the out-of-sample comparison, we consider not only the forecasts for r_{t+h} but also for r_{t+h}^2. The out-of-sample mae, mpe and max_{ae} values were obtained

Table 5.3: Results for the Kolmogorov-Smirnov (K-S) hypothesis test used to compare the sample $\{F_t^{-1}(\hat{x}_t)\}_{t=1}^{4232}$ with the uniform distribution. The values reported are the p-value for the K-S test statistic. The null hypothesis considered is $H_0 : \hat{x}_t|\mathcal{F}_{t-1} \sim \text{GED}(\nu, 0, \sigma_t)$, for different values of ν. The time series $\{x_t\}_{t=1}^{4232}$ corresponds to the residuals of the ARMA model fitted to the one-hour log-return time series $\{r_t\}_{t=1}^{4232}$ and $\{\hat{\sigma}_t^2\}_{t=1}^{4232}$ denotes the conditional variance of the log-returns, obtained from the SFIEGARCH, FIEGARCH or from the EGARCH model.

ν	SFIEGARCH	FIEGARCH	EGARCH
1.45	0.56	0.03	0.25
1.50	0.45	0.09	0.19
1.55	0.36	0.21	0.14
1.70	0.11	0.28	0.04
2.00	0.00	0.00	0.00

by letting $e_{t+h} = r_{t+h} - \hat{r}_{t+h}$, for $t = 2432$ (fixed) and $h = 1, \cdots, 270$; by setting $h = 1$ fixed and letting $e_{t+1} = r_{t+1} - \hat{r}_{t+1}$ for $t = 4232, \cdots, 4501$; by letting $e_{t+h} = r_{t+h}^2 - \hat{r}_{t+h}^2$, for $t = 2432$ (fixed) and $h = 1, \cdots, 270$; and also by letting $h = 1$ fixed and considering $e_{t+1} = r_{t+1}^2 - \hat{r}_{t+1}^2$, for $t = 4232, \cdots, 4501$. For any ARCH-type model, $\hat{r}_{t+h}^2$, for any $h \geq 1$ and $t \in \mathbb{Z}$, was obtained according to Theorem 4.3. While Table 5.4 reports the *mae*, *mpe* and max_{ae} associated to the in-sample and out-of-sample forecasts for r_{t+h}, Table 5.5 gives the values associated to r_{t+h}^2. Notice that, since the ARMA model was selected independently of the ARCH-type models, all values in Table 5.4 do not depend on the model for the conditional variance σ_t^2.

Table 5.4: The mean absolute error (*mae*), the mean percentage error (*mpe*) and the maximum absolute error (max_{ae}) measures for the out-of-sample forecasts for r_{t+h}. Case 1 are the values corresponding to the in-sample forecasts, with $e_t = r_t - \hat{\mu}_t$, for $t \in \{1, \cdots, 4232\}$. Case 2 and 3 correspond to the out-of-sample forecasts. For Case 2, $e_{t+h} = r_{t+h} - \hat{r}_{t+h}$, for $t = 4232$ and $h \in \{1, \cdots, 270\}$. Case 3 assumes $e_{t+1} = r_{t+1} - \hat{r}_{t+1}$, for $t \in \{4232, \cdots, 4501\}$.

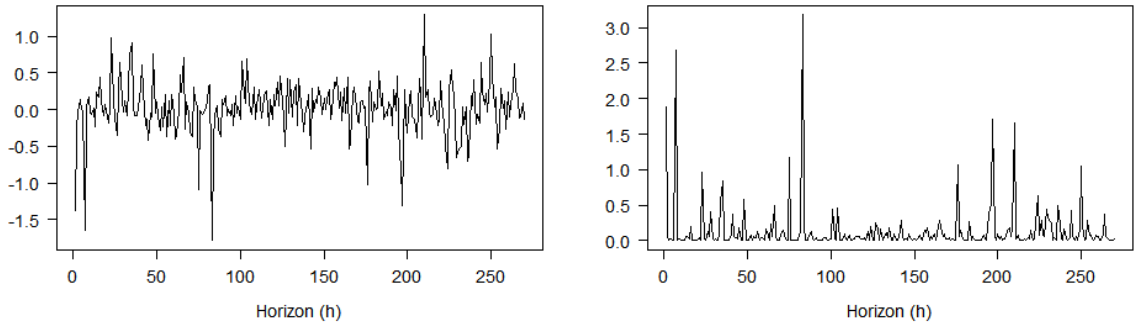
Measure	Case 1	Case 2	Case 3
mae	0.4309	0.2611	0.1525
mape	1.6291	1.3480	1.7712
max_{ae}	6.4438	1.7892	0.9703

Table 5.5: The mean absolute error (*mae*), the mean percentage error (*mpe*) and the maximum absolute error (max_{ae}) measures for the out-of-sample forecasts for r_{t+h}^2. For each model the left column considers $e_t = r_{t+h}^2 - \hat{r}_{t+h}^2$, for $t = 4232$ and $h \in \{1, \cdots, 270\}$, and the right column assumes $e_{t+1} = r_{t+1}^2 - \hat{r}_{t+1}^2$, for $t \in \{4232, \cdots, 4501\}$.

Measure	SFIEGARCH		FIEGARCH		EGARCH	
mae	0.2845	0.1865	0.2829	0.1724	0.2272	0.1687
mape	970.1824	531.1481	684.6831	409.1600	417.9443	372.8434
max_{ae}	2.9271	2.7457	2.9426	2.9844	2.9597	3.0164

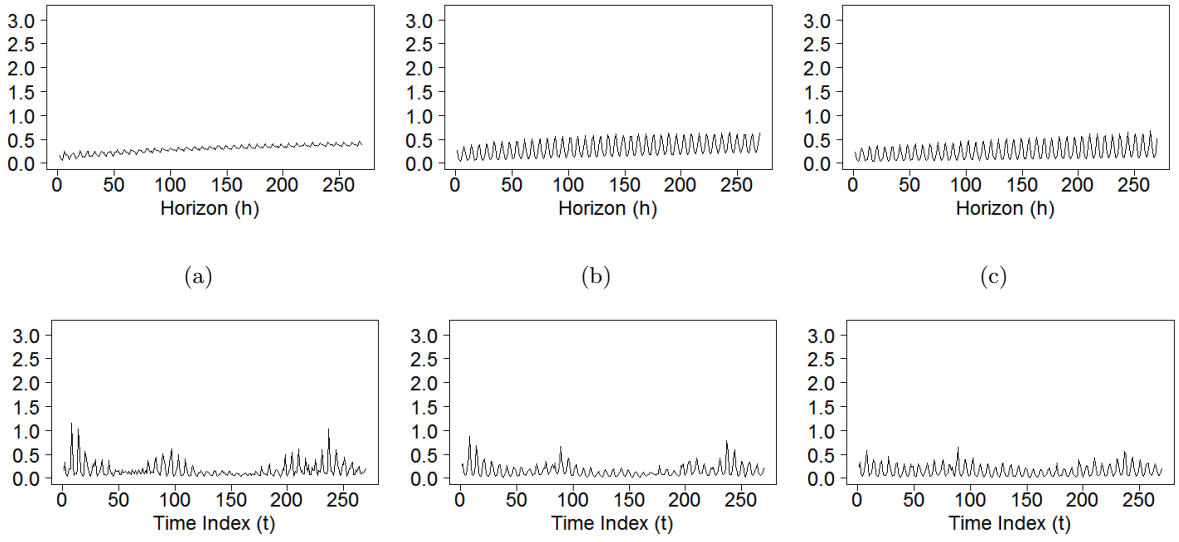
Table 5.5 indicates that the SFIEGARCH model presents the best performance, among all models, only in terms of max_{ae}. Figures 5.5 and 5.6 show, respectively, the observed values r_{t+h}, r_{t+h}^2 and the fitted ones, obtained from the ARCH-type models, for $h = 1, \cdots, 270$. These figures help to explain the reason why the SFIEGARCH model have higher *mae* and *mape* than the other ARCH-type models.

From Figure 5.5 it is clear that r_{t+h}^2 is very close to zero, for several values of $h \in \{1, \cdots, 270\}$. From Figure 5.6 (first row, from top to bottom) it is clear that the h-step ahead forecasts obtained from the SFIEGARCH model converge to a fixed value, which is expected since $\hat{\sigma}_{t+h}^2$ converges to the unconditional variance as h goes to infinity. From Figure 5.6 (first row, from top to bottom) is also evident that the forecasts for the FIEGARCH and EGARCH models do not converge at all.



(a) (b)

Figure 5.5: Sample from the one-hour log-return time series corresponding to the period from August 14, 2009 at 09:15 am to October 8, 2009 at 03:00 pm. Panel (a) gives the log-returns time series $\{r_{t+h}\}_{h=1}^{270}$. Panel (b) gives the squared log-returns $\{r_{t+h}^2\}_{h=1}^{270}$. The time index $t = 4232$ corresponds to August 13, 2009 at 03:00 pm.



(a) (b) (c)

(d) (e) (f)

Figure 5.6: Out-of-sample forecasts for the sample from the one-hour log-return time series corresponding to the period from August 14, 2009 at 09:15 am to October 8, 2009 at 03:00 pm. Panels (a), (b) and (c) give the h-step ahead forecasts $\{\hat{r}_{t+h}^2\}_{h=1}^{270}$, with forecasting origin $t = 2432$, for the SFIEGARCH, FIEGARCH and EGARCH models, respectively. Panels (d), (e) and (b) give the one-step ahead forecast $\{\hat{r}_{t+1}^2\}_{t=4232}^{4501}$, obtained by considering, respectively, the SFIEGARCH, FIEGARCH and EGARCH models fitted to the sample $\{r_t\}_{t=1}^{4232}$.

On the contrary, the forecast values for these two models seem to oscillate around a curve which converges to the same value as the forecasts from the SFIEGARCH model. Moreover, the amplitude of these oscillations increases over time. This should be expected since both the FIEGARCH and the EGARCH models are such that the polynomial $\beta(\cdot)$ has at least one unit root (or a root close enough to the unit root).

Figure 5.6 (second row, from top to bottom) also shows that the one-step ahead forecasts for the SFIEGARCH model were able to capture the peaks in the square log-returns much better than the other two models. Although this is a good feature of the SFIEGARCH model it also makes the *mae* and the *mape* values increase since the forecast values tend to increase in the region around the

peaks while the observed time series shows several values close to zero in the same region. Also, notice that, since the forecasts for r^2_{t+h} are always positive, it is evident that a model which oscilates as the FIEGARCH and the EGARCH do will provide h-step ahead forecasts close to zero more often than a model for which the forecast value converges to a non-zero constant.

Remark 5.12. We apply the Diebold and Mariano hypothesis test (see Diebold and Mariano, 1995) to verify the statistical significance of the out-of-sample forecasting performance. We consider the absolute error as loss function so the loss-differential series is given by $\{\mathrm{d}_t\}$, with $\mathrm{d}_t = |y_{t_0+t+1} - \hat{y}_{t_0+t+1}| - |y_{t_o+t+1}|$, for $t_0 = 2432$ and $t = 0, \cdots, 269$. The variable y_{t_0+t+1} denotes either the log-returns r_{t_0+t+1} or the squared log-returns $r^2_{t_0+t+1}$ and $\hat{y}_{t_0+t+1}$ are the corresponding one-step ahead forecasts obtained from model $\mathcal{M}$. For all models the p-value for the test statistic was smaller than 0.0002. Therefore, the null hypothesis that $\hat{y}_{t_0+t+1} = 0$, for all $t > 0$, was always rejected.

Table 5.6 reports the values of the normalized sum of the realized predictive likelihood for each ARCH-type models. The statistic was obtained by considering the GED(ν) probability density function, for different values of ν. In particular, for $\nu = 2$ we have the Gaussian case. The results in 5.6 show that $\nu = 1.45$ provides slightly better density forecasts for all three models, compared to the other values of $\nu < 2$. Moreover, density forecasts from the EGARCH model are slightly better than for the other two models.

Table 5.6: Normalized sum values of the realized predictive likelihood for each ARCH-type models considering the GED(ν) probability density function for different values of ν.

ν	SFIEGARCH	FIEGARCH	EGARCH
1.45	1.0050	1.0959	1.1331
1.50	0.9973	1.0893	1.1269
1.55	0.9899	1.0829	1.1209
1.70	0.9688	1.0649	1.1040
2.00	0.9318	1.0338	1.0749

Table 5.7 shows the estimated values of the coefficients γ_0 and γ_1 in (5.4). From Table 5.7 one concludes that, in all cases, the null hypothesis of forecast efficiency ($\gamma_0 = 0$ and $\gamma_1 = 1$) is rejected. Table 5.7 also indicates that the three models have a very similar performance. On the other hand, as we mentioned earlier, the log-return time series has several values very close to zero. Moreover, the volatility forecasts obtained from the SFIEGARCH model converge to a constant while the forecasts obtained from the FIEGARCH and EGARCH models show an oscillating behavior leading to forecasts close to zero more often than the SFIEGARCH model.

Table 5.7: Estimated values of the coefficients γ_0 and γ_1 defined in the forecast efficiency regression. The number in parenthesis are the standard errors corrected for heteroskedasticity and autocorrelation (by using the HAC estimator) and uncertainty concerning parameter estimation (upon multiplying by $\lambda = \sqrt{1 + n_p/n}$, where n is the sample size used to fit the model and n_p is the number of predicted values).

h	SFIEGARCH		FIEGARCH		EGARCH	
	γ_0	γ_1	γ_0	γ_1	γ_0	γ_1
1	0.8447 (0.2668)	-0.1087 (0.2043)	0.8018 (0.2453)	-0.0789 (0.1796)	0.7636 (0.4685)	-0.0458 (0.4522)
2	1.5941 (0.6158)	-0.0942 (0.1874)	1.5559 (0.5474)	-0.0840 (0.1799)	1.8090 (0.9887)	-0.2285 (0.4327)
3	2.2665 (1.1055)	-0.0657 (0.1933)	2.2181 (0.9726)	-0.0538 (0.1758)	2.6404 (1.8851)	-0.2181 (0.5425)
4	3.1568 (1.3799)	-0.0906 (0.2194)	3.1464 (1.1611)	-0.0918 (0.1865)	3.8928 (2.4312)	-0.3186 (0.5390)
5	4.0411 (2.3326)	-0.1074 (0.2713)	4.0263 (2.2802)	-0.1077 (0.2753)	5.0712 (4.0225)	-0.3666 (0.7009)
6	5.0353 (2.6067)	-0.1345 (0.2473)	5.0360 (2.3323)	-0.1374 (0.2297)	6.3321 (3.4384)	-0.4120 (0.4943)
7	6.0834 (3.4448)	-0.1598 (0.2711)	6.1792 (3.4706)	-0.1743 (0.2811)	7.7567 (4.0976)	-0.4684 (0.5108)
8	7.2811 (3.2498)	-0.1938 (0.2080)	7.3698 (3.2783)	-0.2050 (0.2247)	9.1453 (3.1329)	-0.5013 (0.3597)
9	8.7566 (4.5220)	-0.2508 (0.2554)	8.7339 (3.9764)	-0.2485 (0.2409)	10.4351 (3.0866)	-0.5120 (0.3238)
10	9.8906 (2.9644)	-0.2636 (0.2124)	9.7976 (2.6865)	-0.2548 (0.2050)	11.4789 (3.4867)	-0.4921 (0.4433)

6 Conclusions

In this work we presented several theoretical results regarding seasonal FIEGARCH (SFIEGARCH) processes. The similarities/differences between this model and the PLM-EGARCH model, introduced by Bordignon et al. (2009), were also discussed.

We proved here that $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is a SARFIMA$(0,d,0)\times(p,0,q)_s$ process. With this result we provided a complete description of the process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ given that necessary and sufficient conditions for the existence, stationarity and ergodicity, as well as the autocovariance structure and spectral representation of the SARFIMA processes are well known. These results were used to establish the conditions for the existence, stationarity and ergodicity of the process $\{X_t\}_{t\in\mathbb{Z}}$ itself. We also provided conditions for the existence of the r-th moment of the random variables $\{\sigma_t^2\}_{t\in\mathbb{Z}}$ and $\{X_t^2\}_{t\in\mathbb{Z}}$ when the underlying distribution is GED. Expressions for the asymmetry and kurtosis measures of any stationary SFIEGARCH process were also derived.

In this paper we also contributed to the theory of SARFIMA$(0,d,0)\times(p,0,q)_s$ processes by extending the range of the parameter d for the invertibility property and by providing an alternative asymptotic expression for the autocovariance function $\gamma_{\ln(\sigma_t^2)}(\cdot)$ of the process $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$. We also derived the exact and the asymptotic expressions for the autocovariance and spectral density functions of the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$.

As an illustration, we analyzed the behavior of the intraday volatility of the S&P500 US stock index log-returns in the period from December 13, 2004 to October 10, 2009. To account for serial correlation in the log-return time series we considered a constrained ARMA model. An SFIEGARCH model was used to account for both the long memory and seasonal behavior for the volatility. FIEGARCH and EGARCH models were also considered in order to analyze the influence of including or not the seasonal parameter in the volatility equation. We conclude that, for this particular time series, not including the seasonal parameter (FIEGARCH model) or ignoring the long-memory behavior (EGARCH model) lead to models which are close enough to the non-stationary region.

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, to the Editor and two anonymous referees whose comments lead to considerable improvement of this paper.

Appendix A: Proofs

In this section we provide the proofs of all propositions, lemmas, corollaries and theorems stated in Sections 1 - 3, in the same order as they appear in the text.

Proof of Proposition 1.2:

Straightforward. ∎

Proof of Theorem 1.1:

From (1.5) and (1.9), one has

$$\sum_{k=0}^{\infty}\lambda_{d,k}z^k=\lambda(z)=\frac{\alpha(z)}{\beta(z)}(1-z^s)^{-d}=\left(\sum_{k=0}^{\infty}f_kz^k\right)\left(\sum_{j=0}^{\infty}\pi_{d,j}z^j\right)=\sum_{k=0}^{\infty}\left(\sum_{j=0}^{k}f_j\pi_{d,k-j}\right)z^k.$$

It follows that, $\lambda_{d,k}=\sum_{j=0}^{k}f_j\pi_{d,k-j}$ or, equivalently,

$$\lambda_{d,sk+r}=\sum_{j=0}^{sk+r}f_j\pi_{d,sk+r-j}=\sum_{j=0}^{k}f_{sj+r}\pi_{d,sk-sj},\quad \text{for all } k\geq 0 \text{ and } r\in\{0,\cdots,s-1\}. \tag{A.1}$$

Let $\mathscr{K}(\cdot)$ be defined as

$$\mathscr{K}(sk+r)=\sum_{j=0}^{\lfloor\sqrt{k}\rfloor}f_{sj+r}\frac{\pi_{d,sk-sj}}{\pi_{sk}},\quad \text{for all } k\in\mathbb{N}\quad\text{and all } r\in\{0,\cdots,s-1\}.$$

From expression (1.10), $\pi_{d,sk-sj}\sim\pi_{d,sk}$, uniformly, for all $0\leq j\leq\lfloor\sqrt{k}\rfloor$. Also, since $\beta(\cdot)$ has no roots in the closed disk $\{z:|z|\leq 1\}$, there exist constants $a>0$ and $B>0$ such that $|f_k|<Be^{-ak}$, for all $k\in\mathbb{N}$ (see Kokoszka and Taqqu, 1994). Hence, $\sum_{j>k_0}|f_j|\leq\frac{Be^{-a(k_0+1)}}{1-e^{-a}}$, for all $k_0\in\mathbb{N}$. It follows that,

$$\lim_{k\to\infty}\sum_{r=0}^{s-1}\mathscr{K}(sk+r)=\sum_{r=0}^{s-1}\left[\lim_{k\to\infty}\sum_{j=0}^{\lfloor\sqrt{k}\rfloor}f_{sj+r}\frac{\pi_{d,sk-sj}}{\pi_{sk}}\right]=\sum_{j=0}^{\infty}f_j=\frac{\alpha(1)}{\beta(1)}.$$

Moreover,

$$\left|\sum_{j=\lfloor\sqrt{k}\rfloor+1}^{k}f_{sj+r}\pi_{d,sk-sj}\right|\leq\sup_{k\in\mathbb{N}}\{|\pi_{d,k}|\}\sum_{j>\lfloor\sqrt{k}\rfloor}|f_{sj+r}|=o(k^{-\nu}),\quad\text{for any }\nu>0.$$

Thus,

$$\lambda_{d,sk+r}=\pi_{d,sk}\sum_{j=0}^{\lfloor\sqrt{k}\rfloor}f_{sj+r}\frac{\pi_{d,sk-sj}}{\pi_{sk}}+\sum_{j=\lfloor\sqrt{k}\rfloor+1}^{k}f_{sj+r}\pi_{d,sk-sj}=\pi_{d,sk}\,\mathscr{K}(sk+r)+o(h^{-\nu}),$$

for any $\nu>0$, and

$$\lim_{k\to\infty}\left[\left(\sum_{r=0}^{s-1}\lambda_{d,sk+r}\right)\left(\pi_{sk}\frac{\alpha(1)}{\beta(1)}\right)^{-1}\right]=\frac{\beta(1)}{\alpha(1)}\lim_{k\to\infty}\left[\sum_{r=0}^{s-1}\mathscr{K}(sk+r)+\frac{o(h^{-\nu})}{\pi_{sk}}\right]=1.$$

Therefore, the result holds. ∎

Proof of Theorem 1.2:

From expression (A.1), for all $r\in\{0,\cdots,s-1\}$,

$$\begin{aligned}\lambda_{d,sk+r}-\frac{1}{\Gamma(d)k^{1-d}}\frac{\alpha(1)}{\beta(1)}&=\sum_{j=0}^{sk+r}f_j\left[\pi_{d,sk+r-j}-\frac{1}{\Gamma(d)k^{1-d}}\right]-\frac{1}{\Gamma(d)k^{1-d}}\sum_{j>sk+r}f_j\\&=\sum_{j=0}^{k}f_{sj+r}\left[\pi_{d,sk-sj}-\frac{1}{\Gamma(d)k^{1-d}}\right]-\frac{1}{\Gamma(d)k^{1-d}}\sum_{j=0}^{sk+r}f_j\mathbb{I}_{\mathbb{R}\setminus\mathbb{N}}\left(\tfrac{|j-r|}{s}\right)\\&\quad-\frac{1}{\Gamma(d)k^{1-d}}\sum_{j>sk+r}f_j.\end{aligned}$$

Since $\sum_{j=0}^{\infty}|f_j|\leq\infty$, there exist $a>0$ and $B>0$ such that $|f_j|<Be^{-aj}$, for all $j\in\mathbb{N}$. Consequently,

$$\frac{1}{\Gamma(d)k^{1-d}}\sum_{j>sk+r}f_j=o(k^{-\nu})\quad\text{and}\quad\frac{1}{\Gamma(d)k^{1-d}}\sum_{j=0}^{sk+r}f_j\,\mathbb{I}_{\mathbb{R}\setminus\mathbb{N}}\left(\tfrac{|j-r|}{s}\right)=O(k^{d-1})\mathbb{I}_{\mathbb{N}\setminus\{0,1\}}(s),$$

for any $\nu > 0$, as k goes to infinity. From Theorem 1.1, one concludes that

$$\sum_{j=0}^{k} f_{sj+r}\left[\pi_{d,sk-sj} - \frac{1}{\Gamma(d)k^{1-d}}\right] = \sum_{j=0}^{\lfloor\sqrt{k}\rfloor} f_{sj+r}\left[\pi_{d,sk-sj} - \frac{1}{\Gamma(d)k^{1-d}}\right] + o(k^{-\nu}) - \frac{1}{\Gamma(d)k^{1-d}}\sum_{j=\lfloor\sqrt{k}\rfloor+1}^{k} f_{sj+r}$$
$$= \sum_{j=0}^{\lfloor\sqrt{k}\rfloor} f_{sj+r}\left[\pi_{d,sk-sj} - \frac{1}{\Gamma(d)k^{1-d}}\right] + O(k^{2-d}),$$

for any $r \in \{0, \cdots, s-1\}$. Since

$$\int_1^{\sqrt{k}} e^{-a(sx+r)}\left[\left(\frac{k-x}{k}\right)^{d-1} - 1\right]\mathrm{d}x = e^{-ar}\int_1^{\sqrt{k}} e^{-asx}\left[\left(1-\frac{x}{k}\right)^{d-1} - 1\right]\mathrm{d}x \quad \text{(setting } x = ky\text{)}$$
$$= e^{-ar}k\int_{1/k}^{\sqrt{k}/k} e^{-asky}\left[(1-y)^{d-1} - 1\right]\mathrm{d}y$$
$$\leq (1-d)2^{1-d}e^{-ar}k\int_0^{\sqrt{k}/k} e^{-asky}y\,\mathrm{d}y \quad \text{(setting } u = asky\text{)}$$
$$= (1-d)2^{1-d}\frac{e^{-ar}}{ask}\int_0^{as\sqrt{k}} ue^{-u}\mathrm{d}u = O(k^{-1}), \quad \text{as } k \to \infty,$$

by using equality (1.5), one concludes that

$$\sum_{j=0}^{\lfloor\sqrt{k}\rfloor} f_{sj+r}\left[\pi_{d,sk-sj} - \frac{1}{\Gamma(d)k^{1-d}}\right] = O(k^{d-2})\sum_{j=0}^{\lfloor\sqrt{k}\rfloor} f_{sj+r} + \frac{1}{\Gamma(d)k^{1-d}}\sum_{j=0}^{\lfloor\sqrt{k}\rfloor} f_{sj+r}\left[\left(\frac{k-j}{k}\right)^{d-1} - 1\right]$$
$$= O(k^{d-2}), \qquad \text{as } k \to \infty.$$

Therefore, equation (1.12) holds. ∎

Proof of Proposition 1.1:

By definition,

$$\frac{\alpha(z)}{\beta(z)}(1-z^s)^{-d} = \sum_{k=0}^{\infty}\lambda_{d,k}z^k \quad \Longrightarrow \quad \alpha(z) = \beta(z)(1-z^s)^d\Big(\sum_{k=0}^{\infty}\lambda_{d,k}z^k\Big). \tag{A.2}$$

Set

$$\alpha_m^* = \begin{cases} \alpha_m, & \text{if} \quad 0 \leq m \leq p, \\ 0, & \text{if} \quad m > p, \end{cases} \quad \text{and} \quad \beta_m^* = \begin{cases} \beta_m, & \text{if} \quad 0 \leq m \leq q, \\ 0, & \text{if} \quad m > q. \end{cases} \tag{A.3}$$

Then,

$$\beta(z)(1-z^s)^d = \Big(\sum_{k=0}^{\infty} -\beta_k^* z^k\Big)\Big(\sum_{k=0}^{\infty}\delta_{d,k}\mathcal{B}^{sk}\Big) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} -\beta_i^*\delta_{d,j}\mathcal{B}^{sj+i} = \sum_{k=0}^{\infty}\tau_k\mathcal{B}^k, \tag{A.4}$$

where $\tau_k = \sum_{i=0}^{k} -\beta_i^*\delta^*_{d,\frac{k-i}{s}}$, for all $k \in \mathbb{N}$, and δ_m^* is defined in (1.13).

Thus, (A.2) can be rewritten as

$$\beta(\mathcal{B})(1-\mathcal{B}^s)^d\Big(\sum_{k=0}^{\infty}\lambda_{d,k}\mathcal{B}^k\Big) = \Big(\sum_{k=0}^{\infty}\tau_k\mathcal{B}^k\Big)\Big(\sum_{k=0}^{\infty}\lambda_{d,k}\mathcal{B}^k\Big) = \sum_{k=0}^{\infty}\Big(\sum_{i=0}^{k}\lambda_{d,i}\tau_{k-i}\Big)\mathcal{B}^k$$
$$= \sum_{k=0}^{\infty}\left[\sum_{i=0}^{k}\lambda_{d,i}\Big(-\sum_{j=0}^{k-i}\beta_j^*\delta^*_{d,\frac{k-i-j}{s}}\Big)\right]\mathcal{B}^k$$
$$= \lambda_{d,0} + \sum_{k=1}^{\infty}\left[\lambda_{d,k} - \sum_{i=0}^{k-1}\lambda_{d,i}\Big(\sum_{j=0}^{k-i}\beta_j^*\delta^*_{d,\frac{k-i-j}{s}}\Big)\right]\mathcal{B}^k. \tag{A.5}$$

Therefore, from (1.13) and (A.5), expression (A.2) holds if and only if

$$-\alpha_0^* = \lambda_{d,0} \quad \text{and} \quad -\alpha_k^* = \lambda_{d,k} - \sum_{i=0}^{k-1}\lambda_{d,i}\Big(\sum_{j=0}^{k-i}\beta_j^*\delta^*_{d,\frac{k-i-j}{s}}\Big), \quad \text{for all} \quad k \geq 1,$$

and the result holds. ∎

Proof of Lemma 1.1:

From Theorem 1.1, $\sum_{k=0}^{\infty}\lambda_{d,k}^2<\infty$ if and only if $d<0.5$. Therefore, if (1.15) is a.s. convergent, by applying the three series theorem (see Billingsley, 1995), one concludes that, necessarily, $d<0.5$. On the other hand, if $d<0.5$, by applying Kolmogorov's convergence criteria (Billingsley, 1995, theorem 22.6), one concludes that (1.15) is a.s. convergent. Finally, from Theorem 1.1, if $d<0$, $\sum_{k=0}^{\infty}|\lambda_{d,k}|<\infty$ and, from proposition 3.1.1 in Brocwel and Davis (1991), the series (1.15) is absolutely a.s. convergent. ∎

Proof of Corollary 1.1:

The result follows immediately from Proposition 1.2, Lemma 1.1 and the definition of SARFIMA processes in Bisognin and Lopes (2009). In particular, if $s=1$, it is an ARFIMA(q,d,p) process (see Brockwell and Davis, 1991). ∎

Proof of Corollary 1.2:

From Lemma 1.1, if $d<0.5$, the random variable σ_t^2 is finite with probability one, for all $t\in\mathbb{Z}$. Since, by assumption $\mathbb{E}(Z_t^2)=\mathbb{E}(Z_0^2)=1$, the random variable Z_t is finite with probability one, for all $t\in\mathbb{Z}$. Therefore, from expression (1.1), the result follows. ∎

Proof of Theorem 1.3:

By Hölder's inequality, it suffices to prove the result for $\mathfrak{p}=2$. From Bisognin and Lopes (2009), the spectral density function of $\{Y_t\}_{t\in\mathbb{Z}}$ is given by

$$f_{\ln(\sigma_t^2)}(\lambda)=\frac{\sigma_g^2}{2\pi}\frac{|\alpha(e^{-i\lambda})|^2}{|\beta(e^{-i\lambda})|^2}|1-e^{-is\lambda}|^{-2d}=\frac{\sigma_g^2}{2\pi}\frac{|\alpha(e^{-i\lambda})|^2}{|\beta(e^{-i\lambda})|^2}\left[2\sin\left(\frac{s\lambda}{2}\right)\right]^{-2d}, \tag{A.6}$$

for all $\lambda\in[0,\pi]$. The authors also show that, for all $k=0,\cdots,\left\lfloor\frac{s}{2}\right\rfloor$,

$$f_Y(\lambda)\sim\frac{\sigma_\varepsilon^2}{2\pi}\left|\frac{\alpha(e^{-i\lambda_k})}{\beta(e^{-i\lambda_k})}\right|^2 s^{-2d}\,|\lambda-\lambda_k|^{-2d},\quad\text{as}\quad\lambda\to\lambda_k:=\frac{2\pi k}{s}. \tag{A.7}$$

Suppose first that the L^2 convergence holds. Notice that, there exists a real constant $c>0$ such that $f_Y(\lambda)\sim c\lambda^{-2d}$, as $\lambda\to0^+$. Consequently, from proposition 1.5.8 in Bingham et al. (1987), if $d\geq0.5$, the function $f_Y(\cdot)\notin L^1$ and hence cannot be a spectral density function. From Theorem 1.1, if $d\leq-1$, then $\tilde{\lambda}_{d,k}\not\to0$, when $k\to\infty$. Consequently, $\tilde{\lambda}_{d,k}Y_{t-k}\not\to0$ in $L^{\mathfrak{p}}$-norm, when $k\to\infty$, and the series representation $\sum_{k=0}^{\infty}\tilde{\lambda}_{d,k}Y_{t-k}$ cannot converge in $L^{\mathfrak{p}}$-norm, for any $0<\mathfrak{p}\leq2$. Therefore, necessarily, $d\in(-1,0.5)$.

Suppose now that $d\in(-1,0.5)$ and assume $s>1$ (the case $s=1$ can be found in Bondon and Palma, 2007). Notice that $d=d_1+d_2$, where $d_1\in(-0.5,0)$ and $d_2\in(-0.5,0.5)$. Define the functions $f_1(\cdot)$ and $f_2(\cdot)$ as follows,

$$f_1(\lambda):=\frac{f_Y(\lambda)}{f_2(\lambda)}\quad\text{and}\quad f_2(\lambda):=|1-e^{is\lambda}|^{-2d_2},\quad\text{for all }\lambda\in[0,\pi].$$

Since $d_1\in(-0.5,0)$, from expressions (A.6) and (A.7), it is obvious that, $f_1(\cdot)\in L^\infty$ and

$$f_1(\lambda)\sim l\Big(\frac{1}{\lambda-2\pi k/s}\Big)\Big|\lambda-\frac{2\pi k}{s}\Big|^{-2d_1},\quad\text{as }\lambda\to\frac{2\pi k}{s},\quad\text{for all }0\leq k\leq\left\lfloor\frac{s}{2}\right\rfloor,$$

where $l(\cdot)$ is a slowly varying function at infinity (for details, see Bingham et al., 1987). Since $d_2\in(-0.5,0.5)$, from expression (A.6), one easily concludes that, for any $\epsilon\in(0,2\pi/s)$, the function

$f_1^{-1}(\cdot)$ is bounded in the interval $[2\pi k/s+\epsilon, 2\pi(k+1)/s-\epsilon]$, for all $0 \leq k \leq \lfloor s/2 \rfloor$. Moreover, from expression (A.7), there exists an $\epsilon \in (0, 2\pi/s)$, such that

$$f_1^{-1}(\lambda) \leq 2l_1\Big(\frac{1}{\lambda - 2\pi k/s}\Big)\Big|\lambda - \frac{2\pi k}{s}\Big|^{2d_1}, \quad \text{for all} \quad \lambda \in \Big(\frac{2\pi k}{s} - \epsilon, \frac{2\pi k}{s}\Big) \cup \Big(\frac{2\pi k}{s}, \frac{2\pi k}{s} + \epsilon\Big)$$

and $k \in \{0, \cdots, \lfloor s/2 \rfloor\}$, where $l_1(\cdot) = l^{-1}(\cdot)$, is also a slowly varying function. It follows that $\int_0^\pi f_1^{-1}(\lambda)d\lambda$ can be written as

$$\int_0^\pi f_1^{-1}(\lambda)d\lambda = \begin{cases} S_1, & \text{if } s \text{ is even;} \\ S_1 + S_2, & \text{if } s \text{ is odd,} \end{cases}$$

where

$$S_1 := \sum_{k=0}^{\lfloor s/2 \rfloor - 1} \left[\int_{2\pi k/s}^{2\pi k/s+\epsilon} f_1^{-1}(\lambda)d\lambda + \int_{2\pi k/s+\epsilon}^{2\pi(k+1)/s-\epsilon} f_1^{-1}(\lambda)d\lambda + \int_{2\pi(k+1)/s-\epsilon}^{2\pi(k+1)/s} f_1^{-1}(\lambda)d\lambda \right],$$
$$S_2 := \int_{2\pi\lfloor s/2 \rfloor/s}^{2\pi\lfloor s/2 \rfloor/s+\epsilon} f_1^{-1}(\lambda)d\lambda + \int_{2\pi\lfloor s/2 \rfloor/s+\epsilon}^{\pi} f_1^{-1}(\lambda)d\lambda,$$

and it satisfies

$$\int_0^\pi f_1^{-1}(\lambda)d\lambda \leq \begin{cases} S_1^*, & \text{if } s \text{ is even;} \\ S_1^* + S_2^*, & \text{if } s \text{ is odd,} \end{cases}$$

where

$$S_1^* := \sum_{k=0}^{\lfloor s/2 \rfloor - 1} \left[2\int_{2\pi k/s}^{2\pi k/s+\epsilon} l_1\Big(\frac{1}{\lambda - 2\pi k/s}\Big)\Big|\lambda - \frac{2\pi k}{s}\Big|^{2d_1} d\lambda + \int_{2\pi k/s+\epsilon}^{2\pi(k+1)/s-\epsilon} f_1^{-1}(\lambda)d\lambda \right.$$
$$\left. + 2\int_{2\pi(k+1)/s-\epsilon}^{2\pi(k+1)/s} l_1\Big(\frac{1}{\lambda - 2\pi k/s}\Big)\Big|\lambda - \frac{2\pi k}{s}\Big|^{2d_1} d\lambda \right],$$
$$S_2^* := 2\int_{2\pi\lfloor s/2 \rfloor/s}^{\pi\lfloor s/2 \rfloor 2/s+\epsilon} l_1\Big(\frac{1}{\lambda - 2\pi k/s}\Big)\Big|\lambda - \frac{2\pi k}{s}\Big|^{2d_1} d\lambda + \int_{2\pi\lfloor s/2 \rfloor/s+\epsilon}^{\pi} f_1^{-1}(\lambda)d\lambda.$$

Thus, since $d_1 > -0.5$, by proposition 1.5.10 in Bingham et al. (1987) we conclude that $f_1^{-1}(\cdot) \in L^1$. By comparing the function $f_2(\cdot)$ with the corresponding one in Bondon and Palma (2007), we conclude that $f_2(\lambda) = |1 - e^{is\lambda}|$ satisfies condition A_p, with $p = 2$, from theorem 3 in Bloomfield (1985). Therefore, taking $p = 1$ in theorem 4 from Bloomfield (1985) we conclude that $\tilde{\lambda}(\cdot)$ has a Fourier series that converges in $L^2(f)$, where $F(\cdot)$ is the spectral distribution function of Y_t, for all $t \in \mathbb{Z}$, and the result follows. ∎

Proof of Lemma 2.1:

See Bisognin and Lopes (2009). ∎

Proof of Corollary 2.1:

It follows immediately from Lemma 2.1. ∎

Proof of Theorem 2.1

Suppose $d < 0.5$. (i) By hypothesis, Z_t is finite with probability one, for all $t \in \mathbb{Z}$. From Corollary 1.1, the random variables $\ln(\sigma_t^2) - \omega$ is finite with probability one, for all $t \in \mathbb{Z}$, so it is σ_t. Thus, from theorem 3.5.8 in Stout (1974), $\{X_t\}_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic process.

(ii) Assume that $\mathbb{E}(|\ln(Z_0^2)|) < \infty$. It follows that the random variable $|\ln(Z_t^2)|$ is finite with probability one, for all $t \in \mathbb{Z}$. From Corollary 1.1, $\mathbb{E}(|\ln(\sigma_t^2)|) < \infty$, for all $t \in \mathbb{Z}$. By expression (1.2), $\ln(X_t^2) = \ln(\sigma_t^2) + \ln(Z_t^2)$, for all $t \in \mathbb{Z}$. It follows that the random variable $\ln(X_t^2)$ 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. From theorem 3.5.8 in Stout (1974), the stochastic process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is strictly stationary and ergodic. Moreover, if $\mathbb{E}([\ln(Z_0^2)]^2) < \infty$ then $\text{Var}\big(\ln(X_t^2)\big) = \text{Var}\big(\ln(X_0^2)\big) = \text{Var}(\ln(\sigma_0^2)) + \text{Var}(\ln(Z_0^2)) < \infty$, for all $t \in \mathbb{Z}$. Therefore, $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is weakly stationary. ∎

Proof of Theorem 2.2:

From Corollary 2.1 and Theorem 2.1 both processes $\{X_t\}_{t\in\mathbb{Z}}$ and $\{\sigma_t\}_{t\in\mathbb{Z}}$ are strictly stationary and hence, any existing moments are time invariant. Let $r > 2$ be any real number such that $\mathbb{E}(|Z_t|^r) < \infty$. By the independence hypothesis $\mathbb{E}(|X_t|^r) = \mathbb{E}(|\sigma_t|^r)\mathbb{E}(|Z_t|^r)$, for all $t \in \mathbb{Z}$. Since $\mathbb{E}(|Z_t|^r) < \infty$, one has to show that $\mathbb{E}(|\sigma_t|^r) < \infty$. From expression (1.2), and from the i.i.d. hypothesis on the random variables Z_t, for all $t \in \mathbb{Z}$, one has

$$\mathbb{E}(|\sigma_t|^r) = \mathbb{E}\Big(\Big|\exp\Big\{\frac{\omega}{2} + \frac{1}{2}\sum_{k=0}^{\infty}\lambda_{d,k}g(Z_{t-1-k})\Big\}\Big|^r\Big) = e^{\frac{r\omega}{2}}\prod_{k=0}^{\infty}\mathbb{E}\Big(\exp\Big\{\frac{r}{2}\lambda_{d,k}g(Z_0)\Big\}\Big). \tag{A.8}$$

From (2.1), expression (A.8) converges to a non-zero constant (see, section 0.25 in Gradshteyn and Ryzhik, 2000). By Hölder's inequality the result follows for all $0 < m < r$. ∎

Proof of Corollary 2.2:

By hypothesis, $d < 0.5$ and $\beta(z)$ has no roots in the closed disk $\{z : |z| \leq 1\}$. From Theorem 1.1, it follows that $\sum_{k=0}^{\infty}\lambda_{d,k}^2 < \infty$. Moreover, from expression A2.4 in Nelson (1991), for all $r > 0$, $\mathbb{E}(\exp\{r\lambda_{d,k}g(Z_0)\}) = 1 + O(\lambda_{d,k}^2)$, as $k \to \infty$. Thus, the result follows immediately from Theorem 2.2. ∎

Proof of Proposition 2.1:

Let $\{X_t\}_{t\in\mathbb{Z}}$ be any stationary SFIEGARCH process. Let $\lambda(\cdot)$ be the polynomial defined by (1.9). Since $\mathbb{E}(X_t^r) = \mathbb{E}(\sigma_t^r Z_t^r)$, for all $r > 0$, $\mathbb{E}(X_t) = 0$ and $\mathbb{E}(Z_t^2) = 1$, for all $t \in \mathbb{Z}$, the asymmetry and kurtosis measures of $\{X_t\}_{t\in\mathbb{Z}}$ are given, respectively, by

$$A_X = \frac{\mathbb{E}(X_t^3)}{\big(\mathbb{E}(X_t^2)\big)^{3/2}} = \frac{\mathbb{E}(\sigma_t^3)\mathbb{E}(Z_t^3)}{\big(\mathbb{E}(\sigma_t^2)\big)^{3/2}} \quad \text{and} \quad K_X = \frac{\mathbb{E}(X_t^4)}{\big(\mathbb{E}(X_t^2)\big)^{2}} = \frac{\mathbb{E}(\sigma_t^4)\mathbb{E}(Z_t^4)}{\big(\mathbb{E}(\sigma_t^2)\big)^{2}}. \tag{A.9}$$

Upon replacing expression (A.8) in (A.9) the result follows. ∎

Proof of Lemma 2.2:

See Bisognin and Lopes (2009). ∎

Proof of Corollary 2.3:

Notice that, for any $\eta \in \mathbb{N}$ one can write

$$\gamma_{\ln(\sigma_t^2)}(sh + r) = \sum_{|k|\leq\eta}\gamma_A(sk + r)\gamma_V(sh - sk) + \sum_{|k|>\eta}\gamma_A(sk + r)\gamma_V(sh - sk).$$

Since $\sum_{h\in\mathbb{Z}}|\gamma_A(h)| < \infty$, there exist $a > 0$ and $B > 0$ such that $|\gamma_A(h)| < Be^{-ah}$, for all $h \in \mathbb{N}$ (see Kokoszka and Taqqu, 1994). Also, $\gamma_A(sk + r) = \gamma_A(-s|k| + r) = \gamma_A(s|k| - r)$, for all $k < -\eta$. Thus,

$$\Big|\sum_{|k|>\eta}\gamma_A(sk + r)\gamma_V(sh - sk)\Big| \leq \gamma_V(0)\sum_{k>\eta}\Big(|\gamma_A(sk + r)| + |\gamma_A(sk - r)|\Big)$$

$$\leq \frac{B\gamma_V(0)(e^{-ar}+e^{ar})}{1-e^{-a}} e^{-as(\sqrt{h}+1)} = o(h^{-\nu}), \quad \text{for all } \nu > 0.$$

On the other hand,

$$\sum_{|k|\leq\eta} \gamma_A(sk+r)\gamma_V(sh-sk) = \gamma_V(sh) \sum_{|k|\leq\eta} \gamma_A(sk+r)\frac{\gamma_V(sh-sk)}{\gamma_V(sh)} := \gamma_V(sh)\mathscr{G}(sh+r).$$

Let $\eta = \lfloor\sqrt{h}\rfloor$. Then, for any $|k| \leq \eta = \lfloor\sqrt{h}\rfloor$, one has $(h-k)^{2d-1} \sim h^{2d-1}$, uniformly, as $h \to \infty$. From expression (2.6),

$$\gamma_V(sh) = \sigma_g^2 \frac{(-1)^h\Gamma(1-2d)}{\Gamma(1-d+h)\Gamma(1-d-h)} = \sigma_g^2 \frac{\Gamma(1-2d)}{\Gamma(1-d)\Gamma(d)} \frac{\Gamma(h+d)}{\Gamma(h+1-d)} \sim \sigma_g^2 \frac{\Gamma(1-2d)}{\Gamma(1-d)\Gamma(d)} h^{2d-1}, \quad \text{as} \quad h \to \infty, \tag{A.10}$$

where σ_g^2 is given by (1.14). Hence, $\gamma_V(sh-sk) = \gamma_V(s(h-k)) \sim \gamma_V(sh)$, uniformly, for all $|k| \leq \eta = \lfloor\sqrt{h}\rfloor$. Since $\sum_{h\in\mathbb{Z}} |\gamma_A(h)| < \infty$, it follows that

$$\lim_{h\to\infty} \sum_{r=0}^{s-1} \mathscr{G}(sh+r) = \lim_{h\to\infty} \sum_{r=0}^{s-1} \sum_{|k|\leq\eta} \gamma_A(sk+r)\frac{\gamma_V(sh-sk)}{\gamma_V(sh)} = \sum_{k\in\mathbb{Z}} \gamma_A(k),$$

and expression (2.7) holds. ∎

Proof of Theorem 2.3:

From Theorem 2.1, $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is a stationary (weakly and strictly) process. Thus,

$$\gamma_{\ln(X_t^2)}(h) = \text{Cov}(\ln(X_t^2), \ln(X_{t+h}^2)) = \text{Cov}(\ln(X_0^2), \ln(X_h^2)), \quad \text{for all } t, h \in \mathbb{Z}$$

From expression (1.1), one concludes that

$$\gamma_{\ln(X_t^2)}(h) = \text{Cov}(\ln(\sigma_0^2), \ln(\sigma_h^2)) + \text{Cov}(\ln(\sigma_0^2), \ln(Z_h^2)) + \text{Cov}(\ln(Z_0^2), \ln(\sigma_h^2)) + \text{Cov}(\ln(Z_0^2), \ln(Z_h^2)),$$

for all $h \in \mathbb{Z}$. Notice that $\text{Cov}(\ln(\sigma_0^2), \ln(\sigma_h^2)) = \gamma_{\ln(\sigma_t^2)}(h)$ is the autocovariance function of $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$, given in Lemma 2.2, and $\text{Cov}(\ln(Z_0^2), \ln(Z_h^2)) = \sigma^2_{\ln(Z_t^2)}\mathbb{I}_{\{0\}}(h)$. Moreover, from expression (1.2),

$$\text{Cov}(\ln(\sigma_h^2), \ln(Z_0^2)) = \text{Cov}\Big(\sum_{k=0}^{\infty} \lambda_{d,k} g(Z_{h-1-k}), \ln(Z_0^2)\Big) = \begin{cases} 0, & \text{if} \quad h \leq 0; \\ C_1\lambda_{d,h-1}, & \text{if} \quad h > 0, \end{cases} \tag{A.11}$$

where $C_1 = \text{Cov}(\ln(Z_0^2), g(Z_0)) = \theta\,\mathbb{E}(Z_0\ln(Z_0^2)) + \gamma\mathbb{E}(|Z_0|\ln(Z_0^2)) - \gamma\mathbb{E}(|Z_0|)\mathbb{E}(\ln(Z_0^2))$. Thus, from expression (A.11), $\text{Cov}(\ln(\sigma_0^2), \ln(Z_h^2)) + \text{Cov}(\ln(Z_0^2), \ln(\sigma_h^2)) = C_1\lambda_{d,|h|-1}\mathbb{I}_{\mathbb{Z}^*}(h)$, for all $h \in \mathbb{Z}$, and expression (2.8) holds. Expression (2.9) follows directly from Corollary 2.3 and Theorem 1.1. ∎

Proof of Corollary 2.4:

From Corollary 2.3 and equation (A.10),

$$\sum_{r=0}^{s-1} \gamma_{\ln(\sigma_t^2)}(sh+r) = \gamma_V(sh) \sum_{r=0}^{s-1} \mathscr{G}(sh+r) + o(h^{-\nu}) \sim \mathscr{C}_1 h^{2d-1}, \quad \text{as } h \to \infty.$$

From Theorem 1.1 and equation (1.10) one concludes that $\mathscr{K}(sh-1) = \mathscr{K}(s(h-1)+(s-1))$, for all $h \geq 1$, and

$$\sum_{r=0}^{s-1} \lambda_{d,sh+r-1} = \pi_{d,s(h-1)}\mathscr{K}(sh-1) + \pi_{d,sh} \sum_{r=1}^{s-1} \mathscr{K}(sk+r-1) + o(h^{-\nu}) \sim \mathscr{C}_2 k^{d-1}, \text{ as } h \to \infty,$$

for any $\nu > 0$, where $\mathscr{C}_1(\cdot)$ and $\mathscr{C}_2(\cdot)$ are given in equation (2.11). Since

$$\lim_{h\to\infty} \frac{h^{2d-1}}{h^{d-1}} = \lim_{h\to\infty} h^d = 0, \quad \text{if } d < 0, \qquad \lim_{h\to\infty} \frac{h^{d-1}}{h^{2d-1}} = \lim_{h\to\infty} h^{-d} = 0, \quad \text{if } d > 0,$$

$$\lim_{h\to\infty} \mathscr{C}_1(h) = \sigma_g^2 \frac{\Gamma(1-2d)}{\Gamma(1-d)\Gamma(d)} \sum_{k\in\mathbb{Z}} \gamma_A(k) \qquad \text{and} \qquad \lim_{h\to\infty} \mathscr{C}_2(h) = \frac{1}{\Gamma(d)} \frac{\alpha(1)}{\beta(1)},$$

expression (2.10) holds. ∎

Proof of Theorem 3.1:

From Theorem 2.1, if $d < 0.5$, the stochastic process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is strictly stationary and ergodic. Moreover, if $\mathrm{Var}(\ln(Z_0^2)) < \infty$, then it is weakly stationary and hence, it has a spectral distribution function. Thus, from Herglotz's theorem (see Brockwell and Davis, 1991), it suffices to show that $f_{\ln(X_t^2)}(\cdot)$, given by (3.1), is a continuous, non-negative function and it satisfies

$$\gamma_{\ln(X_t^2)}(h) = \int_{(-\pi,\pi]} e^{\mathrm{i}h\lambda} f_{\ln(X_t^2)}(\lambda)d\lambda, \quad \text{for all } h \in \mathbb{Z},$$

with $\gamma_{\ln(X_t^2)}(\cdot)$ given in Theorem 2.3.

The continuity of $f_{\ln(X_t^2)}(\cdot)$ follows immediately from its definition. To prove non-negativity notice that, from expression (A.6) and from the i.i.d. hypothesis on the $\{Z_t\}_{t\in\mathbb{Z}}$ process, one has

$$f_{\ln(\sigma_t^2)}(\lambda) = \frac{\sigma_g^2}{2\pi}|\Lambda(\lambda)|^2 \quad \text{and} \quad f_{\ln(Z_t^2)}(\lambda) = \frac{\sigma^2_{\ln(Z_t^2)}}{2\pi}, \quad \text{for all } \lambda \in [0,\pi],$$

where $\Lambda(z) := \lambda(e^{-\mathrm{i}z})$ and $\lambda(\cdot)$ is defined in (1.9). Moreover, $|C_1| \le \sigma_g\sigma_{\ln(Z_t^2)}$ and $|\Re(z)| \le |z|$, for all $z \in \mathbb{C}$. Thus,

$$\begin{aligned} f_{\ln(\sigma_t^2)}(\lambda) + \frac{C_1}{\pi}\Re\big(e^{-\mathrm{i}\lambda}\Lambda(\lambda)\big) + f_{\ln(Z_t^2)}(\lambda) &\ge f_{\ln(\sigma_t^2)}(\lambda) - \frac{\sigma_g\sigma_{\ln(Z_t^2)}}{\pi}\big|e^{-\mathrm{i}\lambda}\Lambda(\lambda)\big| + f_{\ln(Z_t^2)}(\lambda) \\ &= \left(\frac{\sigma_g}{\sqrt{2\pi}}|\Lambda(\lambda)| - \frac{\sigma_{\ln(Z_t^2)}}{\sqrt{2\pi}}\right)^2 \ge 0, \quad \text{for all } \lambda \in [0,\pi]. \end{aligned}$$

To complete the proof, observe that

$$\gamma_{\ln(\sigma_t^2)}(h) = \int_{(-\pi,\pi]} e^{\mathrm{i}h\lambda} f_{\ln(\sigma_t^2)}(\lambda)d\lambda, \qquad \gamma_{\ln(Z_t^2)}(h) = \int_{(-\pi,\pi]} e^{\mathrm{i}h\lambda} f_{\ln(Z_t^2)}(\lambda)d\lambda$$

and

$$\begin{aligned} \int_{(-\pi,\pi]} e^{\mathrm{i}h\lambda}\frac{C_1}{\pi}\Re\big(e^{-\mathrm{i}\lambda}\Lambda(\lambda)\big)d\lambda &= \frac{C_1}{\pi}\sum_{k=0}^{\infty}\lambda_{d,k}\int_{(-\pi,\pi]} e^{\mathrm{i}h\lambda}\cos((k+1)\lambda)d\lambda \\ &= \frac{C_1}{\pi}\sum_{k=0}^{\infty}\lambda_{d,k}\pi\mathbb{I}_{\{0\}}(k+1-|h|) = C_1\lambda_{d,|h|-1}\mathbb{I}_{\mathbb{N}^*}(|h|), \end{aligned}$$

for all $h \in \mathbb{Z}$. Therefore, the result holds. ∎

Proof of Theorem 4.1

Let $\mathscr{S}_1 := \prod_{k=0}^{h-2}\exp\{\lambda_{d,k}g(Z_{n+h-1-k})\}$ and $\mathscr{S}_2 := \prod_{k=h-1}^{\infty}\exp\{\lambda_{d,k}g(Z_{n+h-1-k})\}$, for any $n \in \mathbb{N}$ and $h > 1$ fixed. Notice that, from expression (1.2), one can write

$$\sigma_{n+h}^2 = e^{\omega}\prod_{k=0}^{\infty}\exp\{\lambda_{d,k}g(Z_{n+h-1-k})\} := e^{\omega}\mathscr{S}_1\mathscr{S}_2, \quad \text{for all } n \in \mathbb{N} \text{ and } h > 1. \tag{A.12}$$

Also, observe that the hypothesis $\mathbb{E}(\sigma_t^2) < \infty$ implies $0 < \mathscr{S}_1, \mathscr{S}_2 < \infty$ with probability one.

The $\mathcal{F}_n$-measurability of $Z_{n+h-1-k}$, when $k \ge h-1$, and the the i.i.d. property of $\{Z_t\}_{t\in\mathbb{Z}}$ imply that

$$\mathbb{E}(\sigma_{n+h}^2|\mathcal{F}_n) = e^{\omega}\mathscr{S}_2\mathbb{E}(\mathscr{S}_1) = e^{\omega}\prod_{k=h-1}^{\infty}\exp\{\lambda_{d,k}g(Z_{n+h-1-k})\}\prod_{k=0}^{h-2}\mathbb{E}\big(\exp\{\lambda_{d,k}g(Z_0)\}\big),$$

and expression (4.1) holds.

Now, the independence of $\mathscr{S}_1$ and $\mathscr{S}_2$ implies that

$$\mathbb{E}\big(\big[\sigma_{n+h}^2 - \hat{\sigma}_{n+h}^2\big]^2\big) = \mathbb{E}\big(\big[e^{\omega}\mathscr{S}_1\mathscr{S}_2 - e^{\omega}\mathscr{S}_2\mathbb{E}(\mathscr{S}_1)\big]^2\big) = e^{2\omega}\mathbb{E}(\mathscr{S}_2^2)\mathbb{E}\big(\big[\mathscr{S}_1 - \mathbb{E}(\mathscr{S}_1)\big]^2\big), \quad \text{for any } h > 1.$$

Since $\mathbb{E}\big(\big[\mathscr{S}_1 - \mathbb{E}(\mathscr{S}_1)\big]^2\big) = \mathbb{E}(\mathscr{S}_1^2) - \big[\mathbb{E}(\mathscr{S}_1)\big]^2$, expression (4.2) holds.

To conclude the proof observe that

$$\mathbb{E}\big(\big[X_{n+h}^2 - \hat{X}_{n+h}^2\big]^2\big) = \mathbb{E}\big(\big[X_{n+h}^2 - \hat{\sigma}_{n+h}^2\big]^2\big) = \mathbb{E}\big(\big[e^{\omega}\mathscr{S}_1\mathscr{S}_2 Z_{n+h}^2 - e^{\omega}\mathscr{S}_2\mathbb{E}(\mathscr{S}_1)\big]^2\big), \quad \text{for any } h > 1.$$

Then, by adding and subtracting $\mathbb{E}\big(\big[e^{\omega}\mathscr{S}_1\mathscr{S}_2\big]^2\big)$ to the right hand side of the above equation, one can rewrite

$$\begin{aligned}\mathbb{E}\big(\big[X_{n+h}^2 - \hat{X}_{n+h}^2\big]^2\big) &= \mathbb{E}\big(\big[e^{\omega}\mathscr{S}_1\mathscr{S}_2\big]^2\big)\big[\mathbb{E}(Z_{n+h}^4) - 1\big] + e^{2\omega}\mathbb{E}(\mathscr{S}_2^2)\big(\mathbb{E}(\mathscr{S}_1^2) - \big[\mathbb{E}(\mathscr{S}_1)\big]^2\big)\\ &= \mathbb{E}(\sigma_0^4)\big[\mathbb{E}(Z_0^4) - 1\big] + mse(\sigma_{n+h}^2), \quad \text{for any } h > 1,\end{aligned}$$

and the result follows. ∎

Proof of Theorem 4.2

Expression (4.4) and the first equation in (4.5) follow immediately by mimicking the proof of proposition 4 in Lopes and Prass (2013). The second equation in (4.5) is obtained by replacing $\check{\ln}(X_{n+h}^2)$ by $\hat{\ln}(X_{n+h}^2)$ and mimicking the proof of proposition 5 in Lopes and Prass (2013). ∎

Proof of Theorem 4.3

From (4.7), for any fixed $n \in \mathbb{Z}$,

$$r_{n+h}^2 = \mu^2 + 2\mu\sum_{i=0}^{\infty}\psi_i X_{n+h-i} + \sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\psi_k\psi_j X_{n+h-k}X_{n+h-j}, \quad \text{for all } h > 0.$$

Thus the result follows by observing that $\mathbb{E}(X_{n+h-i}|\mathcal{F}_n) = X_{n+h-i}$, if $i \geq h$, and 0 otherwise, and that

$$\mathbb{E}(X_{n+h-k}X_{n+h-j}|\mathcal{F}_n) = \begin{cases} \mathbb{E}(X_{n+h-k}^2|\mathcal{F}_n) = \hat{\sigma}_{n+h-k}^2, & \text{for any } k,j < h, \text{ with } k = j;\\ 0, & \text{for any } k,j < h, \text{ with } k \neq j;\\ X_{n+h-k}X_{n+h-j}, & \text{for any } k,j \geq h, \text{ with } k \neq j.\end{cases}$$

for any $k, j \in \mathbb{N}$ and $h > 0$. ∎

References

T. Andersen, T. Bollerslev, F. Diebold and P. Labys (2000). "Great realizations". *Risk*, vol. 13, 105-108,

R. Baillie, T. Bollerslev and H. Mikkelsen (1996). "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity". *Journal of Econometrics*, vol. 74, 3-30.

P. Billingsley (1995). *Probability and Measure*. New York: Wiley.

N.H. Bingham, C.M. Goldie and J.L. Teugels (1987). *Regular Variation*. New York: Cambridge University Press.

C. Bisognin and S.R.C. Lopes (2009). "Properties of seasonal long memory processes." *Mathematical and Computer Modelling*, vol. 49, 1837-1851.

P. Bloomfield (1985). "On Series Representations for Linear Predictors." *The Annals of Probability*, vol. 13(1), 226-233.

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

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

T. Bollerslev and J.M. Wooldridge (1992). "Quasi-maximum Likelihood Estimation and Inference in Dynamic Models with Time-varying Covariances". *Econometric Reviews*, vol. 11, 143-172.

P. Bondon and W. Palma (2007). "A class of antipersistent processes". *Journal of Time Series Analysis*, vol. 28, 261-273.

S. Bordignon, M. Caporin and F. Lisi (2007). "Generalized long memory GARCH for intradaily volatility modelling". *Computational Statistics & Data Analysis*, vol. 51, 5900-5912.

S. Bordignon, M. Caporin and F. Lisi (2009). "Periodic Long-Memory GARCH models." *Econometric Reviews*, vol. 28(1), 60-82.

P.J. Brockwell and R.A. Davis (1991). *Time Series: Theory and Methods*. Second Edition. New York: Springer-Verlag.

J. Davidson (2004). "Moment and Memory Properties of Linear Conditional Heteroscedasticity Model, and a New Model." *Journal of Business & Economic Statistics*, vol. 22, 16-29.

F.X. Diebold and R.S. Mariano (1995). "Comparing Predictive Accuracy". *Journal of Business & Economic Statistics*, vol. 13(3), 134-144.

F.X. Diebold, T.A. Gunther and A.S. Tay (1998). "Evaluating Density Forecasts, with Applications to Financial Risk Management". *International Economic Review*, vol. 39, 863-883.

I.S. Gradshteyn and I.M. Ryzhik (2000). *Table of Integrals, Series, and Products*. New York: Academic Press.

M. Haas, S. Mittnik and M. S. Paolella (2004). "A New Approach to Markov-Switching GARCH Models". *Journal of Financial Econometrics*, vol. 2(4), 493-530.

J. Hosking (1981). "Fractional differencing". *Biometrika*, vol. 68, 165-176.

C.M. Hurvich, E. Moulines and P. Soulier (2005). "Estimating Long Memory in Volatility". *Econometrica*, vol. 73(4), 1283-1328.

F. Klaassen (2002). "Improving GARCH Volatility Forecasts with Regime-Switching GARCH". *Empirical Economics*, vol. 27, 363-394.

P.S. Kokoszka and M.S. Taqqu (1994). "Infinite variance stable ARMA processes". *Journal of Time Series Analysis*, vol. 15, 203-220.

P.S. Kokoszka and M.S. Taqqu (1995). "Fractional ARIMA with stable Innovations." *Stochastic Processes and their Applications*, vol. 60, 19-47.

S.R.C. Lopes (2008). "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)*, vol. 60, 497-525, Birkhäuser.

S.R.C. Lopes and T.S. Prass (2013). "Theoretical Results on Fractionally Integrated Exponential Generalized Autoregressive Conditional Heteroskedastic Processes." In Revision.

(1969) J.A. Mincer and V. Zarnowitz. "The evaluation of economic forecasts", In: J. Mincer (Eds.) *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, New York: Columbia University Press, 3-46.

D.B. Nelson (1991). "Conditional Heteroskedasticity in Asset Returns: A New Approach". *Econometrica*, vol. 59, 347-370.

W.K. Newey and K.D. West (1987). "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". *Econometrica*, vol. 55(3), 703-708.

W. Palma (2007). *Long-Memory Time Series: Theory and Methods*. New Jersey: John Wiley.

M. Paolella (2013). "Multivariate Asset Return Prediction with Mixture Models". Forthcoming in: *European Journal of Finance*.

T.S. Prass and S.R.C. Lopes (2012). "Risk Measures Estimation on FIEGARCH Processes." Working Paper.

T.S. Prass and S.R.C. Lopes (2013). "VaR, Teste de Estresse e MaxLoss na Presença de Heteroscedasticidade e Longa Dependência na Volatilidade". Forthcoming in: *Revista Brasilieira de Estatística*.

W.F. Stout (1974). *Almost Sure Convergence*. New York: Academic Press.

K.D. West and M.W. McCracken (1998). "Regression-based Tests of Predictive Ability". *International Economic Review*, vol. 39(4), 817-840.

Apêndice E

Tabelas Complementares

Neste apêndice apresentamos as tabelas correspondentes às simulações envolvendo o cálculo dos coeficientes do polinômio $\lambda(\cdot) = \sum_{k=0}^{\infty} \lambda_{d,k} z^k$, definido por (4.9). Também apresentamos os da comparação de amostras de processos SFIEGARCH gerados a partir das mesmas inovações para diferentes pontos de truncamento.

E.1 Tabelas Complementares para a Seção 4.2.1

Tabela E.1: Valores do ponto de truncamento m do polinômio $\lambda(\cdot)$ e da diferença $diff = \lambda_{d,m} - \hat{\lambda}_{d,m}$, para diferentes valores de d, $s = 1$ e diferentes critérios de convergência ε. Valores de $\lambda_{d,m}$ e da diferença $diff = \lambda_{d,m} - \hat{\lambda}_{d,m}$, para $m = 50000$ e $s = 1, 2$.

d	m estimado para $s=1$				$m = 50000$ (fixo)			
	$\varepsilon = 10^{-3}$		$\varepsilon = 10^{-4}$		$s = 1$		$s = 2$	
	m	$diff$	m	$diff$	$\lambda_{d,m}$	$diff$	$\lambda_{d,m}$	$diff$
				$p = q = 0$, nos casos C1 e C2				
-0.45	49	-6.63e-06	238	-1.37e-07	-4.28e-08	-2.79e-13	-1.17e-07	-1.53e-12
-0.40	55	-5.05e-06	282	-9.92e-08	-7.09e-08	-3.97e-13	-1.87e-07	-2.10e-12
-0.30	66	-2.96e-06	387	-5.04e-08 †	-1.80e-07	-7.02e-13	-4.43e-07	-3.46e-12
-0.20	73	-1.65e-06	497	-2.41e-08	-3.95e-07	-9.47e-13	-9.07e-07	-4.35e-12
-0.10	62	-8.89e-07	503	-1.09e-08	-6.34e-07	-6.98e-13	-1.36e-06	-2.99e-12
0.10	177	-2.54e-07	2278	-1.97e-09	6.20e-06	-5.58e-12	1.16e-05	-2.08e-11
0.20	837	-9.56e-08	14882	-5.38e-10	3.79e-05	-6.07e-11	6.60e-05	-2.11e-10
0.30	4036	-2.60e-08	108251	-9.70e-11	1.72e-04	-3.61e-10 ⋆	2.79e-04	-1.17e-09 ⋆
0.40	26507	-4.53e-09	1230304	-9.75e-12	6.83e-04	-1.64e-09 ⋆	1.04e-03	-4.97e-09 †
0.45	83158	-1.49e-09	5471184	-2.26e-12	1.32e-03	-3.27e-09 †	1.94e-03	-9.59e-09 †

† indica que $\lambda_{d,m} < -\varepsilon \leq \hat{\lambda}_{d,m} < 0$ ou $0 < \hat{\lambda}_{d,m} \leq \varepsilon < \lambda_{d,m}$, para o correspondente ε.
No caso $m = 50000$, significa que $|\lambda_{d,m}| > \varepsilon$, para $\varepsilon = 10^{-3}$.

⋆ indica que $|10^{-4}| < |\lambda_{d,m}| \leq |10^{-3}|$.

* indica que o valor não foi calculado devido a limitações computacionais.

Continua

Tabela E.2: (**Continuação da Tabela E.1**) Valores do ponto de truncamento m do polinômio $\lambda(\cdot)$ e da diferença $diff = \lambda_{d,m} - \hat{\lambda}_{d,m}$, para diferentes valores de d, $s = 1$ e diferentes critérios de convergência ε. Valores de $\lambda_{d,m}$ e da diferença $diff = \lambda_{d,m} - \hat{\lambda}_{d,m}$, para $m = 50000$ e $s = 1, 2$.

d	m estimado para $s=1$				$m = 50000$ (fixo)			
	$\varepsilon = 10^{-3}$		$\varepsilon = 10^{-4}$		$s = 1$		$s = 2$	
	m	$diff$	m	$diff$	$\lambda_{d,m}$	$diff$	$\lambda_{d,m}$	$diff$
d				$p = 0$ e $q = 1$, no caso C1				
−0.45	148	−4.45e-05 †	721	−8.62e-07 †	−2.14e-07	−2.62e-11	−3.25e-07	2.60e-07
−0.40	172	−3.64e-05 †	889	−6.69e-07 †	−3.54e-07	−4.17e-11	−5.20e-07	4.16e-07
−0.30	227	−2.49e-05 †	1335	−4.07e-07 †	−9.00e-07	−9.71e-11	−1.23e-06	9.85e-07
−0.20	279	−1.83e-05 †	1899	−2.60e-07 †	−1.97e-06	−1.94e-10	−2.52e-06	2.01e-06
−0.10	268	−1.72e-05 †	2171	−2.06e-07 †	−3.17e-06	−2.83e-10	−3.78e-06	3.02e-06
0.10	1055	3.40e-06 †	13616	2.61e-08 †	3.10e-05	2.21e-09	3.22e-05	−2.57e-05
0.20	6257	4.99e-07 †	111263	2.80e-09 †	1.90e-04	1.18e-08 ⋆	1.83e-04	−1.47e-04 ⋆
0.30	40215	6.70e-08 †	1078845	*	8.59e-04	4.63e-08 ⋆	7.75e-04	−6.20e-04 ⋆
0.40	387522	5.88e-09 †	17987143	*	3.42e-03	1.56e-07 †	2.88e-03	−2.30e-03 †
0.45	1551509	*	102078879	*	6.61e-03	2.75e-07 †	5.38e-03	−4.30e-03 †
d				$p = 0$ e $q = 1$, no caso C2				
−0.45	70	−1.88e-05	338	−3.85e-07 †	−7.13e-08	−1.84e-12	−1.39e-07	5.57e-08
−0.40	79	−1.57e-05 †	406	−3.00e-07 †	−1.18e-07	−2.87e-12	−2.23e-07	8.91e-08
−0.30	98	−1.11e-05 †	574	−1.85e-07	−3.00e-07	−6.37e-12	−5.28e-07	2.11e-07
−0.20	112	−8.36e-06 †	760	−1.21e-07 †	−6.58e-07	−1.21e-11	−1.08e-06	4.32e-07
−0.10	99	−8.12e-06 †	800	−9.88e-08 †	−1.06e-06	−1.67e-11	−1.62e-06	6.47e-07
0.10	312	1.79e-06	4017	1.38e-08 †	1.03e-05	1.15e-10	1.38e-05	−5.51e-06
0.20	1585	2.86e-07 †	28181	1.61e-09	6.32e-05	5.73e-10	7.86e-05	−3.14e-05
0.30	8372	4.32e-08	224573	1.61e-10 †	2.86e-04	2.07e-09 ⋆	3.32e-04	−1.33e-04 ⋆
0.40	62101	4.51e-09	2882439	*	1.14e-03	6.38e-09 †	1.23e-03	−4.93e-04 †
0.45	210505	1.15e-09	13849774	*	2.20e-03	1.07e-08 †	2.31e-03	−9.22e-04 †
				$p = 0$ e $q = 2$, no caso C1				
−0.45	41	1.38e-05	199	2.86e-07	−3.29e-08	3.73e-13	−7.93e-08	1.06e-08
−0.40	46	1.27e-05	234	2.50e-07	−5.45e-08	6.34e-13	−1.27e-07	1.69e-08
−0.30	54	1.15e-05	317	1.91e-07	−1.38e-07	1.68e-12	−3.01e-07	4.01e-08
−0.20	59	1.07e-05	399	1.56e-07	−3.04e-07	3.76e-12	−6.15e-07	8.21e-08
−0.10	49	1.30e-05	396	1.57e-07	−4.88e-07	6.07e-12	−9.23e-07	1.23e-07
0.10	132	−4.56e-06	1702	−3.52e-08	4.77e-06	−5.71e-11	7.86e-06	−1.05e-06
0.20	603	−9.50e-07	10721	−5.34e-09	2.92e-05	−3.34e-10	4.48e-05	−5.98e-06
0.30	2774	−1.93e-07	74414	−7.20e-10	1.32e-04	−1.42e-09 ⋆	1.89e-04	−2.52e-05 ⋆
0.40	17118	−2.86e-08	794524	−6.16e-11	5.26e-04	−5.14e-09 ⋆	7.03e-04	−9.37e-05 ⋆
0.45	51610	−8.96e-09	3395562	*	1.02e-03	−9.41e-09 †	1.31e-03	−1.75e-04 †

† indica que $\lambda_{d,m} < -\varepsilon \leq \hat{\lambda}_{d,m} < 0$ ou $0 < \hat{\lambda}_{d,m} \leq \varepsilon < \lambda_{d,m}$, para o correspondente ε.
No caso $m = 50000$, significa que $|\lambda_{d,m}| > \varepsilon$, para $\varepsilon = 10^{-3}$.

⋆ indica que $|10^{-4}| < |\lambda_{d,m}| \leq |10^{-3}|$.

* indica que o valor não foi calculado devido a limitações computacionais.

Continua

Tabela E.3: (**Continuação da Tabela E.1**) Valores do ponto de truncamento m do polinômio $\lambda(\cdot)$ e da diferença $diff = \lambda_{d,m} - \hat{\lambda}_{d,m}$, para diferentes valores de d, $s = 1$ e diferentes critérios de convergência ε. Valores de $\lambda_{d,m}$ e da diferença $diff = \lambda_{d,m} - \hat{\lambda}_{d,m}$, para $m = 50000$ e $s = 1, 2$.

d	m estimado para $s=1$				$m = 50000$ (fixo)			
	$\varepsilon = 10^{-3}$		$\varepsilon = 10^{-4}$		$s = 1$		$s = 2$	
	m	$diff$	m	$diff$	$\lambda_{d,m}$	$diff$	$\lambda_{d,m}$	$diff$
d					$p = 0$ e $q = 2$, no caso C2			
−0.45	46	4.90e-06	223	9.24e-08	−3.89e-08	1.56e-13	−9.21e-08	1.42e-08
−0.40	51	4.93e-06	263	8.88e-08	−6.44e-08	2.95e-13	−1.47e-07	2.27e-08
−0.30	62	4.67e-06	360	7.78e-08	−1.64e-07	9.09e-13	−3.49e-07	5.37e-08
−0.20	68	4.79e-06	459	6.93e-08	−3.59e-07	2.27e-12	−7.14e-07	1.10e-07
−0.10	57	6.27e-06	461	7.52e-08	−5.77e-07	3.98e-12	−1.07e-06	1.65e-07
0.10	159	−2.36e-06	2049	−1.82e-08	5.64e-06	−4.20e-11	9.12e-06	−1.40e-06
0.20	743	−5.00e-07	13210	−2.81e-09	3.45e-05	−2.56e-10	5.20e-05	−8.00e-06
0.30	3522	−1.02e-07	94471	−3.81e-10	1.56e-04	−1.12e-09 ⋆	2.20e-04	−3.38e-05 ⋆
0.40	22613	−1.50e-08	1049602	*	6.21e-04	−4.20e-09 ⋆	8.16e-04	−1.26e-04 ⋆
0.45	69927	−4.63e-09	4600679	*	1.20e-03	−7.79e-09 †	1.53e-03	−2.35e-04 †
d					$p = 1$ e $q = 0$, no caso C1			
−0.45	35	1.89e-05	167	3.89e-07	−2.57e-08	3.29e-13	−1.17e-07	−4.68e-08
−0.40	38	1.81e-05	196	3.36e-07	−4.25e-08	5.56e-13	−1.87e-07	−7.48e-08
−0.30	45	1.53e-05	262	2.57e-07	−1.08e-07	1.45e-12	−4.43e-07	−1.77e-07
−0.20	48	1.45e-05	325	2.10e-07	−2.37e-07	3.22e-12	−9.07e-07	−3.63e-07
−0.10	39	1.79e-05	316	2.15e-07	−3.81e-07	5.16e-12	−1.36e-06	−5.44e-07
0.10	100	−6.50e-06	1291	−5.00e-08	3.72e-06	−4.80e-11	1.16e-05	4.63e-06
0.20	442	−1.39e-06	7859	−7.80e-09	2.28e-05	−2.79e-10	6.60e-05	2.64e-05
0.30	1946	−2.94e-07	52180	−1.10e-09	1.03e-04	−1.18e-09 ⋆	2.79e-04	1.12e-04 ⋆
0.40	11314	−4.60e-08	525127	−9.90e-11	4.10e-04	−4.26e-09 ⋆	1.04e-03	4.14e-04 †
0.45	32851	−1.49e-08	2161325	*	7.94e-04	−7.79e-09 ⋆	1.94e-03	7.75e-04 †
d					$p = 1$ e $q = 0$, no caso C2			
−0.45	16	3.86e-04	79	7.00e-06	−8.55e-09	9.37e-13	−1.17e-07	−9.35e-08
−0.40	18	3.07e-04	90	5.95e-06	−1.42e-08	1.51e-12	−1.87e-07	−1.50e-07
−0.30	20	2.55e-04	113	4.45e-06	−3.60e-08	3.60e-12	−4.43e-07	−3.54e-07
−0.20	20	2.37e-04	130	3.63e-06	−7.89e-08	7.39e-12	−9.07e-07	−7.25e-07
−0.10	15	2.99e-04	117	3.73e-06	−1.27e-07	1.10e-11	−1.36e-06	−1.09e-06
0.10	30	−1.23e-04	381	−9.59e-07	1.24e-06	−9.04e-11	1.16e-05	9.26e-06
0.20	112	−2.95e-05	1991	−1.65e-07	7.58e-06	−4.98e-10	6.60e-05	5.28e-05
0.30	405	−7.18e-06	10862	−2.67e-08	3.43e-05	−2.00e-09	2.79e-04	2.23e-04 ⋆
0.40	1813	−1.39e-06	84152	−2.99e-09	1.37e-04	−6.89e-09 ⋆	1.04e-03	8.29e-04 †
0.45	4458	−5.21e-07	293243	−7.92e-10	2.65e-04	−1.23e-08 ⋆	1.94e-03	1.55e-03 †

† indica que $\lambda_{d,m} < -\varepsilon \leq \hat{\lambda}_{d,m} < 0$ ou $0 < \hat{\lambda}_{d,m} \leq \varepsilon < \lambda_{d,m}$, para o correspondente ε.
No caso $m = 50000$, significa que $|\lambda_{d,m}| > \varepsilon$, para $\varepsilon = 10^{-3}$.

⋆ indica que $|10^{-4}| < |\lambda_{d,m}| \leq |10^{-3}|$.

* indica que o valor não foi calculado devido a limitações computacionais.

Continua

Tabela E.4: (**Continuação da Tabela E.1**) Valores do ponto de truncamento m do polinômio $\lambda(\cdot)$ e da diferença $diff = \lambda_{d,m} - \hat{\lambda}_{d,m}$, para diferentes valores de d, $s = 1$ e diferentes critérios de convergência ε. Valores de $\lambda_{d,m}$ e da diferença $diff = \lambda_{d,m} - \hat{\lambda}_{d,m}$, para $m = 50000$ e $s = 1, 2$.

d	m estimado para $s=1$				$m = 50000$ (fixo)			
	$\varepsilon = 10^{-3}$		$\varepsilon = 10^{-4}$		$s = 1$		$s = 2$	
	m	$diff$	m	$diff$	$\lambda_{d,m}$	$diff$	$\lambda_{d,m}$	$diff$
d				$p = 1$ e $q = 1$, no caso C1				
−0.45	104	−5.54e-05 †	507	−1.04e-06 †	−1.28e-07	−1.32e-11	−2.21e-07	1.30e-07
−0.40	120	−4.50e-05 †	617	−8.16e-07 †	−2.13e-07	−2.10e-11	−3.53e-07	2.08e-07
−0.30	154	−3.14e-05 †	901	−5.08e-07 †	−5.40e-07	−4.89e-11	−8.37e-07	4.92e-07
−0.20	183	−2.37e-05 †	1241	−3.34e-07 †	−1.18e-06	−9.76e-11	−1.71e-06	1.01e-06
−0.10	169	−2.32e-05 †	1364	−2.75e-07 †	−1.90e-06	−1.42e-10	−2.57e-06	1.51e-06
0.10	598	5.01e-06 †	7719	3.83e-08 †	1.86e-05	1.10e-09	2.19e-05	−1.29e-05
0.20	3304	7.85e-07 †	58754	4.40e-09 †	1.14e-04	5.89e-09 ⋆	1.25e-04	−7.34e-05 ⋆
0.30	19385	1.15e-07 †	520035	4.29e-10 †	5.15e-04	2.30e-08 ⋆	5.27e-04	−3.10e-04 ⋆
0.40	165405	1.14e-08 †	7677402	*	2.05e-03	7.71e-08 †	1.96e-03	−1.15e-03 †
0.45	612905	2.79e-09 †	40325015	*	3.97e-03	1.36e-07 †	3.66e-03	−2.15e-03 †
d				$p = 1$ e $q = 1$, no caso C2				
−0.45	23	2.30e-04	112	4.12e-06	−1.43e-08	1.29e-12	−9.46e-08	−5.57e-08
−0.40	25	2.02e-04	129	3.47e-06	−2.36e-08	2.07e-12	−1.51e-07	−8.91e-08
−0.30	29	1.55e-04	167	2.51e-06	−6.00e-08	4.97e-12	−3.59e-07	−2.11e-07
−0.20	30	1.39e-04	199	1.97e-06	−1.32e-07	1.02e-11	−7.34e-07	−4.32e-07
−0.10	23	1.76e-04	186	1.96e-06	−2.11e-07	1.53e-11	−1.10e-06	−6.47e-07
0.10	53	−5.89e-05	672	−4.54e-07	2.07e-06	−1.26e-10	9.37e-06	5.51e-06
0.20	212	−1.31e-05	3770	−7.29e-08	1.26e-05	−6.94e-10	5.35e-05	3.14e-05
0.30	840	−2.91e-06	22534	−1.08e-08	5.72e-05	−2.79e-09	2.26e-04	1.33e-04 ⋆
0.40	4248	−4.99e-07	197156	−1.08e-09	2.28e-04	−9.66e-09 ⋆	8.38e-04	4.93e-04 ⋆
0.45	11283	−1.73e-07	742315	−2.64e-10	4.41e-04	−1.73e-08 ⋆	1.57e-03	9.22e-04 †
d				$p = 1$ e $q = 2$, no caso C1				
−0.45	29	4.77e-05	140	1.10e-06	−1.97e-08	6.05e-13	−7.51e-08	−2.12e-08
−0.40	32	3.03e-05	162	9.41e-07	−3.27e-08	9.91e-13	−1.20e-07	−3.39e-08
−0.30	37	3.63e-05	214	6.88e-07	−8.31e-08	2.44e-12	−2.85e-07	−8.02e-08
−0.20	39	3.77e-05	261	5.44e-07	−1.82e-07	5.17e-12	−5.83e-07	−1.64e-07
−0.10	31	6.05e-05	249	5.45e-07	−2.93e-07	7.94e-12	−8.74e-07	−2.46e-07
0.10	75	−1.60e-05	965	−1.24e-07	2.86e-06	−6.86e-11	7.44e-06	2.09e-06
0.20	319	−3.46e-06	5662	−1.95e-08	1.75e-05	−3.87e-10	4.24e-05	1.20e-05
0.30	1338	−7.49e-07	35870	−2.79e-09	7.93e-05	−1.59e-09	1.79e-04	5.05e-05 ⋆
0.40	7307	−1.22e-07	339125	−2.62e-10	3.15e-04	−5.61e-09 ⋆	6.65e-04	1.87e-04 ⋆
0.45	20388	−4.07e-08	1341376	*	6.11e-04	−1.01e-08 ⋆	1.24e-03	3.51e-04 †

† indica que $\lambda_{d,m} < -\varepsilon \leq \hat{\lambda}_{d,m} < 0$ ou $0 < \hat{\lambda}_{d,m} \leq \varepsilon < \lambda_{d,m}$, para o correspondente ε.
No caso $m = 50000$, significa que $|\lambda_{d,m}| > \varepsilon$, para $\varepsilon = 10^{-3}$.

⋆ indica que $|10^{-4}| < |\lambda_{d,m}| \leq |10^{-3}|$.

* indica que o valor não foi calculado devido a limitações computacionais.

Continua

Tabela E.5: (**Continuação da Tabela E.1**) Valores do ponto de truncamento m do polinômio $\lambda(\cdot)$ e da diferença $diff = \lambda_{d,m} - \hat{\lambda}_{d,m}$, para diferentes valores de d, $s = 1$ e diferentes critérios de convergência ε. Valores de $\lambda_{d,m}$ e da diferença $diff = \lambda_{d,m} - \hat{\lambda}_{d,m}$, para $m = 50000$ e $s = 1, 2$.

d	m estimado para $s=1$				$m = 50000$ (fixo)			
	$\varepsilon = 10^{-3}$		$\varepsilon = 10^{-4}$		$s = 1$		$s = 2$	
	m	$diff$	m	$diff$	$\lambda_{d,m}$	$diff$	$\lambda_{d,m}$	$diff$
d					$p = 1$ e $q = 2$, no caso C2			
−0.45	15	1.70e-04	74	8.11e-06	−7.78e-09	9.34e-13	−8.08e-08	−5.95e-08
−0.40	17	4.08e-04	84	6.93e-06	−1.29e-08	1.50e-12	−1.29e-07	−9.52e-08
−0.30	18	3.27e-04	105	5.21e-06	−3.27e-08	3.58e-12	−3.06e-07	−2.26e-07
−0.20	18	3.05e-04	120	4.28e-06	−7.17e-08	7.34e-12	−6.26e-07	−4.62e-07
−0.10	14	3.10e-04	107	4.44e-06	−1.15e-07	1.09e-11	−9.39e-07	−6.92e-07
0.10	27	−1.46e-04	343	−1.16e-06	1.13e-06	−8.96e-11	8.00e-06	5.89e-06
0.20	100	−3.56e-05	1767	−2.02e-07	6.90e-06	−4.92e-10	4.56e-05	3.36e-05
0.30	354	−8.92e-06	9480	−3.33e-08	3.12e-05	−1.97e-09	1.93e-04	1.42e-04 ⋆
0.40	1547	−1.77e-06	71792	−3.81e-09	1.24e-04	−6.80e-09 ⋆	7.16e-04	5.27e-04 ⋆
0.45	3748	−6.73e-07	246586	−1.02e-09	2.41e-04	−1.21e-08 ⋆	1.34e-03	9.86e-04 †
d					$p = 2$ e $q = 0$, no caso C1			
−0.45	52	−1.71e-05 †	254	−3.39e-07 †	−4.71e-08	−8.03e-13	−1.52e-07	−2.34e-08
−0.40	59	−1.37e-05	302	−2.62e-07	−7.80e-08	−1.23e-12	−2.43e-07	−3.74e-08
−0.30	71	−9.70e-06 †	417	−1.61e-07	−1.98e-07	−2.64e-12	−5.76e-07	−8.86e-08
−0.20	79	−7.25e-06 †	538	−1.04e-07	−4.34e-07	−4.83e-12	−1.18e-06	−1.81e-07
−0.10	68	−6.89e-06	548	−8.34e-08 †	−6.98e-07	−6.35e-12	−1.77e-06	−2.72e-07
0.10	197	1.45e-06	2532	1.12e-08	6.82e-06	3.85e-11	1.50e-05	2.32e-06
0.20	943	2.24e-07	16764	1.26e-09 †	4.17e-05	1.76e-10	8.58e-05	1.32e-05
0.30	4624	3.24e-08	124041	1.21e-10	1.89e-04	5.65e-10 ⋆	3.63e-04	5.58e-05 ⋆
0.40	31070	3.16e-09	1442116	*	7.52e-04	1.48e-09 ⋆	1.35e-03	2.07e-04 †
0.45	98892	7.71e-10	6506400	*	1.46e-03	2.22e-09 †	2.52e-03	3.87e-04 †
d					$p = 2$ e $q = 0$, no caso C2			
−0.45	59	−2.11e-05 †	285	−4.31e-07 †	−5.56e-08	−1.36e-12	−1.75e-07	−2.34e-08
−0.40	66	−1.78e-05 †	340	−3.37e-07 †	−9.22e-08	−2.10e-12	−2.81e-07	−3.74e-08
−0.30	81	−1.26e-05 †	474	−2.11e-07 †	−2.34e-07	−4.66e-12	−6.65e-07	−8.86e-08
−0.20	91	−9.64e-06 †	618	−1.39e-07 †	−5.13e-07	−8.81e-12	−1.36e-06	−1.81e-07
−0.10	79	−9.49e-06 †	638	−1.15e-07 †	−8.25e-07	−1.21e-11	−2.04e-06	−2.72e-07
0.10	236	2.18e-06 †	3048	1.67e-08 †	8.06e-06	8.21e-11	1.74e-05	2.32e-06
0.20	1162	3.55e-07 †	20657	2.00e-09 †	4.93e-05	4.07e-10	9.90e-05	1.32e-05
0.30	5870	5.55e-08 †	157474	2.07e-10	2.23e-04	1.45e-09 ⋆	4.18e-04	5.58e-05 ⋆
0.40	41045	6.07e-09	1905100	*	8.88e-04	4.43e-09 ⋆	1.55e-03	2.07e-04 †
0.45	133990	1.60e-09	8815584	*	1.72e-03	7.39e-09 †	2.91e-03	3.87e-04 †

† indica que $\lambda_{d,m} < -\varepsilon \leq \hat{\lambda}_{d,m} < 0$ ou $0 < \hat{\lambda}_{d,m} \leq \varepsilon < \lambda_{d,m}$, para o correspondente ε.
No caso $m = 50000$, significa que $|\lambda_{d,m}| > \varepsilon$, para $\varepsilon = 10^{-3}$.

⋆ indica que $|10^{-4}| < |\lambda_{d,m}| \leq |10^{-3}|$.

* indica que o valor não foi calculado devido a limitações computacionais.

Continua

Tabela E.6: (**Continuação da Tabela E.1**) Valores do ponto de truncamento m do polinômio $\lambda(\cdot)$ e da diferença $diff = \lambda_{d,m} - \hat{\lambda}_{d,m}$, para diferentes valores de d, $s = 1$ e diferentes critérios de convergência ε. Valores de $\lambda_{d,m}$ e da diferença $diff = \lambda_{d,m} - \hat{\lambda}_{d,m}$, para $m = 50000$ e $s = 1, 2$.

d	m estimado para $s=1$				$m = 50000$ (fixo)			
	$\varepsilon = 10^{-3}$		$\varepsilon = 10^{-4}$		$s = 1$		$s = 2$	
	m	$diff$	m	$diff$	$\lambda_{d,m}$	$diff$	$\lambda_{d,m}$	$diff$
d				$p = 2$ e $q = 1$, no caso C1				
−0.45	158	−4.51e-05 †	770	−8.76e-07 †	−2.35e-07	−3.13e-11	−3.70e-07	2.73e-07
−0.40	184	−3.69e-05 †	952	−6.78e-07 †	−3.90e-07	−4.98e-11	−5.92e-07	4.36e-07
−0.30	245	−2.49e-05 †	1436	−4.12e-07 †	−9.90e-07	−1.16e-10	−1.40e-06	1.03e-06
−0.20	302	−1.83e-05 †	2055	−2.62e-07 †	−2.17e-06	−2.33e-10	−2.87e-06	2.12e-06
−0.10	292	−1.72e-05 †	2367	−2.06e-07 †	−3.49e-06	−3.39e-10	−4.31e-06	3.17e-06
0.10	1172	3.34e-06 †	15137	2.57e-08 †	3.41e-05	2.65e-09	3.67e-05	−2.70e-05
0.20	7049	4.84e-07 †	125340	2.72e-09 †	2.09e-04	1.42e-08 ⋆	2.09e-04	−1.54e-04 ⋆
0.30	46081	6.40e-08 †	1236207	*	9.45e-04	5.57e-08 ⋆	8.83e-04	−6.51e-04 ⋆
0.40	454238	5.50e-09 †	21083854	*	3.76e-03	1.88e-07 †	3.28e-03	−2.42e-03 †
0.45	1845073	*	121393470	*	7.28e-03	3.31e-07 †	6.13e-03	−4.52e-03 †
d				$p = 2$ e $q = 1$, no caso C2				
−0.45	83	−2.74e-05 †	405	−5.44e-07 †	−9.27e-08	−4.05e-12	−1.98e-07	5.56e-08
−0.40	95	−2.25e-05 †	489	−4.27e-07 †	−1.54e-07	−6.37e-12	−3.16e-07	8.91e-08
−0.30	120	−1.58e-05 †	702	−2.66e-07 †	−3.90e-07	−1.45e-11	−7.49e-07	2.11e-07
−0.20	139	−1.22e-05 †	946	−1.76e-07 †	−8.55e-07	−2.84e-11	−1.53e-06	4.32e-07
−0.10	126	−1.18e-05 †	1015	−1.45e-07 †	−1.37e-06	−4.03e-11	−2.30e-06	6.47e-07
0.10	417	2.67e-06 †	5377	2.06e-08 †	1.34e-05	2.98e-10	1.96e-05	−5.51e-06
0.20	2200	4.30e-07 †	39118	2.42e-09 †	8.22e-05	1.55e-09	1.12e-04	−3.14e-05 ⋆
0.30	12178	6.51e-08 †	326688	2.43e-10 †	3.72e-04	5.90e-09 ⋆	4.72e-04	−1.33e-04 ⋆
0.40	96161	6.75e-09 †	4463399	*	1.48e-03	1.92e-08 †	1.75e-03	−4.93e-04 †
0.45	339181	1.71e-09 †	22315798	*	2.87e-03	3.33e-08 †	3.27e-03	−9.22e-04 †
d				$p = 2$ e $q = 2$, no caso C1				
−0.45	44	1.12e-06	212	1.90e-08	−3.62e-08	2.81e-14	−1.01e-07	−2.12e-09
−0.40	49	1.57e-06	250	2.95e-08	−6.00e-08	8.69e-14	−1.62e-07	−3.39e-09
−0.30	58	2.36e-06	341	3.89e-08	−1.52e-07	4.03e-13	−3.83e-07	−8.02e-09
−0.20	64	2.87e-06	432	4.23e-08	−3.34e-07	1.22e-12	−7.84e-07	−1.64e-08
−0.10	54	4.11e-06	432	5.14e-08	−5.37e-07	2.38e-12	−1.18e-06	−2.46e-08
0.10	147	−1.85e-06	1892	−1.44e-08	5.25e-06	−2.85e-11	1.00e-05	2.09e-07
0.20	680	−4.14e-07	12077	−2.33e-09	3.21e-05	−1.81e-10	5.71e-05	1.19e-06
0.30	3179	−8.85e-08	85269	−3.30e-10	1.45e-04	−8.17e-10 ⋆	2.41e-04	5.05e-06 ⋆
0.40	20065	−1.35e-08	931311	−2.91e-11	5.78e-04	−3.13e-09 ⋆	8.95e-04	1.87e-05 ⋆
0.45	61375	−4.27e-09	4038044	*	1.12e-03	−5.87e-09 †	1.67e-03	3.50e-05 †

† indica que $\lambda_{d,m} < -\varepsilon \leq \hat{\lambda}_{d,m} < 0$ ou $0 < \hat{\lambda}_{d,m} \leq \varepsilon < \lambda_{d,m}$, para o correspondente ε.
No caso $m = 50000$, significa que $|\lambda_{d,m}| > \varepsilon$, para $\varepsilon = 10^{-3}$.

⋆ indica que $|10^{-4}| < |\lambda_{d,m}| \leq |10^{-3}|$.

* indica que o valor não foi calculado devido a limitações computacionais.

Continua

Tabela E.7: (**Continuação da Tabela E.1**) Valores do ponto de truncamento m do polinômio $\lambda(\cdot)$ e da diferença $diff = \lambda_{d,m} - \hat{\lambda}_{d,m}$, para diferentes valores de d, $s = 1$ e diferentes critérios de convergência ε. Valores de $\lambda_{d,m}$ e da diferença $diff = \lambda_{d,m} - \hat{\lambda}_{d,m}$, para $m = 50000$ e $s = 1, 2$.

d	m estimado para $s=1$				$m = 50000$ (fixo)			
	$\varepsilon = 10^{-3}$		$\varepsilon = 10^{-4}$		$s = 1$		$s = 2$	
	m	$diff$	m	$diff$	$\lambda_{d,m}$	$diff$	$\lambda_{d,m}$	$diff$
d	$p = 2$ e $q = 2$, no caso C2							
-0.45	55	-1.27e-05	267	-2.59e-07	-5.06e-08	-6.99e-13	-1.35e-07	2.83e-09
-0.40	62	-1.02e-05	318	-1.99e-07	-8.38e-08	-1.06e-12	-2.17e-07	4.53e-09
-0.30	75	-7.05e-06 †	440	-1.19e-07 †	-2.13e-07	-2.22e-12	-5.13e-07	1.07e-08
-0.20	84	-5.08e-06 †	571	-7.40e-08	-4.66e-07	-3.94e-12	-1.05e-06	2.20e-08
-0.10	73	-4.56e-06	585	-5.68e-08 †	-7.50e-07	-4.98e-12	-1.57e-06	3.30e-08
0.10	213	8.55e-07	2742	6.62e-09	7.33e-06	2.66e-11	1.34e-05	-2.81e-07
0.20	1032	1.18e-07	18337	6.62e-10	4.48e-05	1.09e-10	7.64e-05	-1.60e-06
0.30	5123	1.39e-08	137428	5.18e-11	2.03e-04	2.89e-10 ⋆	3.23e-04	-6.76e-06 ⋆
0.40	35016	8.87e-10	1625287	*	8.08e-04	5.02e-10 ⋆	1.20e-03	-2.51e-05 †
0.45	112671	1.31e-10	7412960	*	1.56e-03	4.60e-10 †	2.24e-03	-4.70e-05 †

† indica que $\lambda_{d,m} < -\varepsilon \leq \hat{\lambda}_{d,m} < 0$ ou $0 < \hat{\lambda}_{d,m} \leq \varepsilon < \lambda_{d,m}$, para o correspondente ε.
No caso $m = 50000$, significa que $|\lambda_{d,m}| > \varepsilon$, para $\varepsilon = 10^{-3}$.

⋆ indica que $|10^{-4}| < |\lambda_{d,m}| \leq |10^{-3}|$.

* indica que o valor não foi calculado devido a limitações computacionais.

E.2 Tabelas Complementares para a Seção 4.2.2

Tabela E.8: Valores mínimo e máximo (RANGE(X_t)) das variáveis aleatórias $\{X_t(m_1)\}_{t=1}^{n}$ e das estatísticas MIN, MAX, MAE, SD e $\bar{e}$, obtidas considerando-se m_1 o valor de m obtidos pela Tabela E.1, para $\varepsilon = 10^{-3}$ e $s = 1$, $m_2 = 50000$ e $m_3 = 100000$.

d	RANGE(X_t)	Medidas Obtidas com m_1 e m_2					Medidas Obtidas com m_1 e m_3				
		MIN	MAX	MAE	SD	$\bar{e}$	MIN	MAX	MAE	SD	$\bar{e}$
Caso $p = 0$ e $q = 0$											
0.10	[-3.16289, 3.56670]	0.00000	0.00583	0.00079	0.00086	-0.00003	0.00000	0.00591	0.00079	0.00087	-0.00003
0.20	[-3.26868, 3.55790]	0.00000	0.00954	0.00122	0.00122	-0.00007	0.00000	0.00879	0.00124	0.00120	-0.00008
0.30	[-3.44414, 3.62923]	0.00000	0.02848	0.00522	0.00423	0.00020	0.00000	0.02498	0.00468	0.00380	0.00018
0.40	[-3.66573, 3.74290]	0.00000	0.00893	0.00113	0.00138	0.00000	0.00000	0.01504	0.00265	0.00244	-0.00009
Caso $p = 0$ e $q = 1$											
0.10	[-4.04594, 4.45593]	0.00000	0.01443	0.00176	0.00190	-0.00001	0.00000	0.01386	0.00174	0.00185	-0.00001
0.20	[-4.79219, 4.49270]	0.00001	0.03585	0.00464	0.00459	0.00006	0.00000	0.03407	0.00408	0.00417	0.00005
0.30	[-6.01537, 4.48784]	0.00000	0.02772	0.00209	0.00257	-0.00002	0.00001	0.03883	0.00473	0.00458	-0.00009
0.40	[-10.10964, 5.93662]	0.00009	1.12158	0.10332	0.10439	0.00155	0.00007	1.05336	0.09245	0.09472	0.00141
Caso $p = 0$ e $q = 2$											
0.10	[-3.28433, 3.52165]	0.00000	0.00661	0.00068	0.00077	-0.00004	0.00000	0.00670	0.00068	0.00078	-0.00004
0.20	[-3.36130, 3.49368]	0.00000	0.00954	0.00118	0.00120	-0.00006	0.00000	0.00973	0.00125	0.00125	-0.00006
0.30	[-3.51800, 3.57499]	0.00000	0.02009	0.00459	0.00368	0.00024	0.00000	0.01853	0.00417	0.00335	0.00022
0.40	[-3.70364, 3.64370]	0.00000	0.01303	0.00242	0.00230	0.00017	0.00000	0.00758	0.00111	0.00114	0.00009
Caso $p = 1$ e $q = 0$											
0.10	[-3.06300, 3.61334]	0.00000	0.00575	0.00059	0.00067	-0.00003	0.00000	0.00578	0.00059	0.00067	-0.00004
0.20	[-3.12281, 3.58640]	0.00000	0.01032	0.00120	0.00128	-0.00005	0.00000	0.01077	0.00125	0.00134	-0.00005
0.30	[-3.24045, 3.58813]	0.00000	0.02043	0.00460	0.00367	0.00027	0.00000	0.01925	0.00427	0.00343	0.00026
0.40	[-3.38364, 3.54303]	0.00000	0.01810	0.00387	0.00305	0.00021	0.00000	0.01370	0.00253	0.00215	0.00014
Caso $p = 1$ e $q = 1$											
0.10	[-3.56267, 3.86997]	0.00000	0.01088	0.00136	0.00137	-0.00008	0.00000	0.01100	0.00141	0.00140	-0.00008
0.20	[-3.96734, 3.97186]	0.00000	0.02663	0.00458	0.00377	0.00018	0.00000	0.02575	0.00424	0.00352	0.00017
0.30	[-4.59435, 4.07443]	0.00001	0.04684	0.00860	0.00696	0.00037	0.00000	0.04114	0.00701	0.00576	0.00031
0.40	[-5.73612, 4.45224]	0.00002	0.15968	0.02443	0.02027	0.00092	0.00001	0.13328	0.01796	0.01519	0.00073
Caso $p = 1$ e $q = 2$											
0.10	[-3.19792, 3.70802]	0.00000	0.00368	0.00051	0.00056	-0.00002	0.00000	0.00372	0.00051	0.00056	-0.00002
0.20	[-3.20121, 3.65351]	0.00000	0.00766	0.00108	0.00111	-0.00007	0.00000	0.00802	0.00111	0.00115	-0.00007
0.30	[-3.31188, 3.62900]	0.00000	0.01919	0.00336	0.00316	0.00021	0.00000	0.01780	0.00312	0.00297	0.00020
0.40	[-3.46021, 3.59496]	0.00001	0.04618	0.01013	0.00804	0.00047	0.00001	0.04235	0.00910	0.00724	0.00042
Caso $p = 2$ e $q = 0$											
0.10	[-3.10770, 3.57372]	0.00000	0.00774	0.00085	0.00093	-0.00002	0.00000	0.00783	0.00085	0.00094	-0.00002
0.20	[-3.22821, 3.64784]	0.00000	0.01274	0.00155	0.00161	-0.00003	0.00000	0.01191	0.00154	0.00155	-0.00003
0.30	[-3.40957, 3.74688]	0.00000	0.02104	0.00460	0.00410	0.00031	0.00000	0.01907	0.00401	0.00376	0.00029
0.40	[-3.69718, 3.86746]	0.00001	0.03630	0.00753	0.00596	0.00032	0.00000	0.02282	0.00510	0.00408	0.00022
Caso $p = 2$ e $q = 1$											
0.10	[-4.14802, 4.36060]	0.00000	0.01881	0.00202	0.00234	0.00001	0.00000	0.01827	0.00197	0.00228	0.00001
0.20	[-5.02108, 4.35892]	0.00001	0.06117	0.00827	0.00727	0.00029	0.00001	0.05913	0.00766	0.00679	0.00027
0.30	[-6.43041, 4.64494]	0.00000	0.03756	0.00285	0.00355	0.00002	0.00001	0.05064	0.00576	0.00584	-0.00005
0.40	[-10.65987, 6.58220]	0.00010	1.27291	0.11210	0.11926	0.00117	0.00008	1.19126	0.10007	0.10828	0.00107
Caso $p = 2$ e $q = 2$											
0.10	[-3.18242, 3.58590]	0.00000	0.00766	0.00070	0.00078	-0.00003	0.00000	0.00771	0.00070	0.00079	-0.00003
0.20	[-3.26884, 3.56115]	0.00000	0.00812	0.00115	0.00116	-0.00009	0.00000	0.00840	0.00120	0.00120	-0.00009
0.30	[-3.43400, 3.54272]	0.00000	0.02636	0.00501	0.00403	0.00027	0.00000	0.02340	0.00456	0.00367	0.00025
0.40	[-3.65490, 3.60785]	0.00001	0.04213	0.00867	0.00681	0.00043	0.00000	0.03007	0.00680	0.00534	0.00035

Tabela E.9: Valores mínimo e máximo (RANGE($\ln(h_t^2)$)) das variáveis aleatórias $\{\ln(h_t^2(m_1))\}_{t=1}^n$ e das estatísticas MIN, MAX, MAE, SD e $\bar{e}$, obtidas considerando-se m_1 o valor de m obtidos pela Tabela E.1, para $\varepsilon = 10^{-3}$ e $s = 1$, $m_2 = 50000$ e $m_3 = 100000$.

d	RANGE($\ln(h_t^2)$)	Medidas Obtidas com m_1 e m_2					Medidas Obtidas com m_1 e m_3				
		MIN	MAX	MAE	SD	$\bar{e}$	MIN	MAX	MAE	SD	$\bar{e}$
Caso $p = 0$ e $q = 0$											
0.10	[-0.24759, 0.92402]	0.00001	0.00595	0.00200	0.00119	-0.00084	0.00000	0.00600	0.00201	0.00121	-0.00089
0.20	[-0.30769, 1.01494]	0.00001	0.00790	0.00316	0.00173	-0.00026	0.00001	0.00741	0.00320	0.00162	-0.00055
0.30	[-0.38850, 1.14466]	0.00939	0.01933	0.01331	0.00232	0.01331	0.00814	0.01700	0.01193	0.00210	0.01193
0.40	[-0.52906, 1.29566]	0.00000	0.01031	0.00307	0.00241	-0.00104	0.00143	0.01305	0.00678	0.00275	-0.00678
Caso $p = 0$ e $q = 1$											
0.10	[-0.88971, 1.63020]	0.00000	0.00999	0.00451	0.00249	0.00182	0.00000	0.00961	0.00446	0.00236	0.00159
0.20	[-1.16994, 2.09467]	0.00343	0.01850	0.01163	0.00373	0.01163	0.00184	0.01625	0.01016	0.00353	0.01016
0.30	[-1.55155, 2.69472]	0.00104	0.01004	0.00513	0.00208	-0.00513	0.00665	0.01880	0.01206	0.00273	-0.01206
0.40	[-1.85733, 3.78635]	0.23465	0.27483	0.25203	0.00809	0.25203	0.21175	0.23452	0.22332	0.00572	0.22332
Caso $p = 0$ e $q = 2$											
0.10	[-0.51561, 0.90810]	0.00000	0.00558	0.00174	0.00108	-0.00051	0.00000	0.00561	0.00174	0.00110	-0.00055
0.20	[-0.48752, 0.97925]	0.00003	0.00688	0.00301	0.00151	-0.00268	0.00003	0.00702	0.00317	0.00152	-0.00290
0.30	[-0.45744, 1.10453]	0.00902	0.01552	0.01165	0.00159	0.01165	0.00769	0.01369	0.01059	0.00146	0.01059
0.40	[-0.48623, 1.24258]	0.00006	0.01208	0.00626	0.00275	0.00600	0.00001	0.00691	0.00290	0.00152	0.00159
Caso $p = 1$ e $q = 0$											
0.10	[-0.42960, 0.90387]	0.00001	0.00523	0.00152	0.00098	-0.00028	0.00000	0.00525	0.00152	0.00100	-0.00030
0.20	[-0.38197, 0.85888]	0.00002	0.00668	0.00301	0.00160	-0.00277	0.00002	0.00688	0.00314	0.00168	-0.00294
0.30	[-0.33021, 0.93590]	0.00848	0.01555	0.01167	0.00178	0.01167	0.00739	0.01472	0.01084	0.00181	0.01084
0.40	[-0.33151, 1.04490]	0.00659	0.01495	0.00993	0.00154	0.00993	0.00147	0.01152	0.00648	0.00203	0.00648
Caso $p = 1$ e $q = 1$											
0.10	[-0.56848, 1.25656]	0.00009	0.00727	0.00345	0.00164	-0.00330	0.00002	0.00736	0.00355	0.00166	-0.00344
0.20	[-0.73154, 1.49547]	0.00831	0.01494	0.01168	0.00169	0.01168	0.00755	0.01409	0.01080	0.00167	0.01080
0.30	[-0.94980, 1.81813]	0.01812	0.02752	0.02215	0.00188	0.02215	0.01394	0.02306	0.01800	0.00210	0.01800
0.40	[-1.23194, 2.37968]	0.05502	0.07399	0.06381	0.00480	0.06381	0.03699	0.05566	0.04658	0.00490	0.04658
Caso $p = 1$ e $q = 2$											
0.10	[-0.58330, 0.97244]	0.00000	0.00482	0.00132	0.00087	-0.00008	0.00000	0.00484	0.00132	0.00087	-0.00010
0.20	[-0.57587, 0.93348]	0.00002	0.00648	0.00278	0.00152	-0.00198	0.00003	0.00662	0.00285	0.00159	-0.00212
0.30	[-0.55970, 0.89457]	0.00001	0.01390	0.00855	0.00364	0.00855	0.00000	0.01287	0.00795	0.00351	0.00791
0.40	[-0.55245, 0.99575]	0.01944	0.03034	0.02562	0.00195	0.02562	0.01684	0.02785	0.02297	0.00188	0.02297
Caso $p = 2$ e $q = 0$											
0.10	[-0.34051, 0.93080]	0.00002	0.00610	0.00214	0.00124	-0.00101	0.00000	0.00614	0.00216	0.00127	-0.00106
0.20	[-0.33070, 1.03329]	0.00000	0.00953	0.00399	0.00223	0.00103	0.00000	0.00907	0.00396	0.00206	0.00071
0.30	[-0.40798, 1.18039]	0.00384	0.01997	0.01165	0.00419	0.01165	0.00102	0.01812	0.01013	0.00439	0.01013
0.40	[-0.55266, 1.39219]	0.01378	0.02352	0.01939	0.00217	0.01939	0.00915	0.01691	0.01307	0.00178	0.01307
Caso $p = 2$ e $q = 1$											
0.10	[-0.94095, 1.86889]	0.00002	0.01031	0.00512	0.00268	0.00338	0.00000	0.01003	0.00499	0.00259	0.00312
0.20	[-1.23134, 2.35642]	0.01411	0.02636	0.02107	0.00313	0.02107	0.01278	0.02522	0.01946	0.00298	0.01946
0.30	[-1.66294, 2.97673]	0.00108	0.01218	0.00691	0.00297	-0.00691	0.00827	0.02393	0.01453	0.00394	-0.01453
0.40	[-2.00193, 4.13966]	0.25085	0.28791	0.26950	0.00835	0.26950	0.22030	0.25194	0.23791	0.00772	0.23791
Caso $p = 2$ e $q = 2$											
0.10	[-0.30897, 0.88003]	0.00000	0.00572	0.00183	0.00112	-0.00062	0.00000	0.00575	0.00184	0.00114	-0.00066
0.20	[-0.29451, 0.95945]	0.00001	0.00677	0.00294	0.00145	-0.00205	0.00001	0.00701	0.00306	0.00149	-0.00229
0.30	[-0.34228, 1.08515]	0.00951	0.01616	0.01279	0.00175	0.01279	0.00868	0.01502	0.01162	0.00163	0.01162
0.40	[-0.43864, 1.23829]	0.01782	0.02805	0.02228	0.00239	0.02228	0.01474	0.02375	0.01742	0.00190	0.01742

Tabela E.10: Valores mínimo e máximo (RANGE(X_t)) das variáveis aleatórias $\{X_t(m_1)\}_{t=1}^{n}$ e das estatísticas MIN, MAX, MAE, SD e $\bar{e}$, obtidas considerando-se m_1 o valor de m obtidos pela Tabela E.1, para $\varepsilon = 10^{-3}$ e $s = 12$, $m_2 = 50000$ e $m_3 = 100000$.

d	RANGE(X_t)	Medidas Obtidas com m_1 e m_2					Medidas Obtidas com m_1 e m_3				
		MIN	MAX	MAE	SD	$\bar{e}$	MIN	MAX	MAE	SD	$\bar{e}$
Caso $p = 0$ e $q = 0$											
0.10	[-3.43322, 3.35749]	0.00000	0.01129	0.00131	0.00140	-0.00003	0.00000	0.01140	0.00130	0.00146	-0.00002
0.20	[-3.44544, 3.35933]	0.00000	0.01591	0.00229	0.00245	-0.00005	0.00000	0.01641	0.00251	0.00283	-0.00002
0.30	[-3.48449, 3.31319]	0.00000	0.00539	0.00061	0.00073	-0.00001	0.00000	0.01691	0.00209	0.00232	0.00009
0.40	[-3.39907, 3.16456]	0.00002	0.09853	0.01243	0.01275	-0.00020	0.00002	0.11152	0.00745	0.00968	0.00019
Caso $p = 0$ e $q = 1$											
0.10	[-3.59777, 2.98743]	0.00000	0.00708	0.00082	0.00092	-0.00005	0.00000	0.00694	0.00092	0.00097	-0.00003
0.20	[-3.52060, 2.96125]	0.00000	0.00735	0.00105	0.00117	-0.00005	0.00000	0.00683	0.00067	0.00077	0.00002
0.30	[-3.53327, 2.95310]	0.00000	0.04710	0.00616	0.00653	-0.00038	0.00000	0.03313	0.00274	0.00302	-0.00009
0.40	[-4.16978, 3.14954]	0.00002	0.16229	0.02316	0.02502	0.00046	0.00000	0.16064	0.01869	0.02560	0.00147
Caso $p = 0$ e $q = 2$											
0.10	[-3.10758, 3.51519]	0.00000	0.01286	0.00169	0.00179	-0.00012	0.00000	0.01243	0.00164	0.00179	-0.00010
0.20	[-3.06396, 3.59559]	0.00000	0.02093	0.00270	0.00300	-0.00013	0.00000	0.02321	0.00268	0.00328	-0.00005
0.30	[-3.15775, 3.64247]	0.00000	0.02033	0.00319	0.00317	0.00004	0.00000	0.03763	0.00413	0.00508	0.00033
0.40	[-3.38739, 3.53428]	0.00001	0.13016	0.01379	0.01571	-0.00156	0.00001	0.06248	0.00838	0.00883	-0.00053
Caso $p = 1$ e $q = 0$											
0.10	[-3.55301, 3.47361]	0.00000	0.01237	0.00111	0.00139	-0.00010	0.00000	0.01245	0.00114	0.00140	-0.00011
0.20	[-3.57198, 3.50163]	0.00000	0.02599	0.00316	0.00395	-0.00006	0.00000	0.02957	0.00365	0.00409	-0.00007
0.30	[-3.60664, 3.45499]	0.00000	0.03230	0.00261	0.00400	-0.00012	0.00000	0.04483	0.00404	0.00592	-0.00013
0.40	[-3.53235, 3.31422]	0.00003	0.07536	0.01329	0.01300	-0.00008	0.00000	0.08175	0.00876	0.01098	-0.00006
Caso $p = 1$ e $q = 1$											
0.10	[-3.49605, 2.97878]	0.00000	0.00739	0.00095	0.00109	-0.00011	0.00000	0.00743	0.00090	0.00105	-0.00010
0.20	[-3.48309, 2.95401]	0.00000	0.01006	0.00128	0.00133	0.00006	0.00000	0.01417	0.00194	0.00193	0.00010
0.30	[-3.41781, 2.93486]	0.00000	0.02430	0.00359	0.00384	-0.00013	0.00000	0.01993	0.00223	0.00269	0.00005
0.40	[-3.81396, 3.04558]	0.00001	0.14716	0.01938	0.02013	-0.00005	0.00003	0.13896	0.01544	0.01650	0.00061
Caso $p = 1$ e $q = 2$											
0.10	[-3.13128, 3.60420]	0.00000	0.01454	0.00178	0.00212	-0.00016	0.00000	0.01453	0.00173	0.00206	-0.00016
0.20	[-3.08379, 3.70456]	0.00001	0.02172	0.00337	0.00346	-0.00008	0.00000	0.03009	0.00354	0.00379	-0.00004
0.30	[-3.06664, 3.77922]	0.00000	0.04398	0.00422	0.00573	-0.00007	0.00001	0.03951	0.00606	0.00653	0.00010
0.40	[-3.43677, 3.79740]	0.00000	0.09571	0.00926	0.01110	-0.00001	0.00000	0.05685	0.00505	0.00648	0.00061
Caso $p = 2$ e $q = 0$											
0.10	[-3.71493, 3.32825]	0.00000	0.00851	0.00121	0.00119	-0.00001	0.00000	0.00864	0.00126	0.00121	-0.00001
0.20	[-3.71347, 3.30737]	0.00001	0.01736	0.00230	0.00228	-0.00003	0.00001	0.01852	0.00262	0.00271	-0.00003
0.30	[-3.71984, 3.24492]	0.00000	0.00925	0.00123	0.00159	0.00009	0.00000	0.01540	0.00226	0.00256	0.00010
0.40	[-3.56069, 3.09231]	0.00005	0.10170	0.01432	0.01609	0.00018	0.00000	0.09180	0.00907	0.01154	0.00024
Caso $p = 2$ e $q = 1$											
0.10	[-3.70211, 2.83302]	0.00000	0.00670	0.00065	0.00078	-0.00004	0.00000	0.00658	0.00073	0.00080	-0.00003
0.20	[-3.61358, 2.85050]	0.00000	0.00728	0.00127	0.00122	-0.00003	0.00000	0.00393	0.00048	0.00052	0.00002
0.30	[-3.71377, 2.90129]	0.00001	0.07087	0.00822	0.00804	-0.00018	0.00000	0.05219	0.00440	0.00487	0.00001
0.40	[-4.30880, 3.12092]	0.00001	0.16402	0.02187	0.02271	0.00116	0.00001	0.17232	0.02140	0.02634	0.00186
Caso $p = 2$ e $q = 2$											
0.10	[-3.25544, 3.43717]	0.00000	0.01045	0.00134	0.00148	-0.00007	0.00000	0.01046	0.00130	0.00152	-0.00007
0.20	[-3.27820, 3.47277]	0.00000	0.02037	0.00223	0.00261	-0.00011	0.00000	0.02130	0.00243	0.00287	-0.00007
0.30	[-3.31765, 3.47302]	0.00000	0.02505	0.00197	0.00296	0.00016	0.00000	0.02790	0.00369	0.00464	0.00031
0.40	[-3.45516, 3.35295]	0.00000	0.10624	0.01427	0.01410	-0.00089	0.00002	0.07525	0.01005	0.00995	-0.00032

Tabela E.11: Valores mínimo e máximo (RANGE($\ln(h_t^2)$)) das variáveis aleatórias $\{\ln(h_t^2(m_1))\}_{t=1}^n$ e das estatísticas MIN, MAX, MAE, SD e $\bar{e}$, obtidas considerando-se m_1 o valor de m obtidos pela Tabela E.1, para $\varepsilon = 10^{-3}$ e $s = 12$, $m_2 = 50000$ e $m_3 = 100000$.

d	RANGE($\ln(h_t^2)$)	Medidas Obtidas com m_1 e m_2					Medidas Obtidas com m_1 e m_3				
		MIN	MAX	MAE	SD	$\bar{e}$	MIN	MAX	MAE	SD	$\bar{e}$
Caso $p = 0$ e $q = 0$											
0.10	[-0.24747, 1.12819]	0.00000	0.00866	0.00332	0.00196	0.00007	0.00000	0.00878	0.00330	0.00213	0.00017
0.20	[-0.30436, 1.15315]	0.00004	0.01397	0.00582	0.00351	-0.00136	0.00001	0.01472	0.00637	0.00425	-0.00086
0.30	[-0.38795, 1.12876]	0.00001	0.00500	0.00164	0.00128	-0.00006	0.00002	0.01257	0.00539	0.00349	0.00174
0.40	[-0.59580, 1.01850]	0.00296	0.06442	0.03224	0.01713	-0.00797	0.00005	0.06975	0.01922	0.01606	-0.00225
Caso $p = 0$ e $q = 1$											
0.10	[-0.77307, 1.37594]	0.00001	0.00494	0.00210	0.00128	-0.00030	0.00000	0.00484	0.00237	0.00125	0.00021
0.20	[-0.86645, 1.33502]	0.00000	0.00680	0.00274	0.00171	-0.00271	0.00000	0.00495	0.00172	0.00108	-0.00022
0.30	[-1.01789, 1.31865]	0.00000	0.03215	0.01636	0.00862	-0.01574	0.00246	0.02193	0.00739	0.00463	-0.00676
0.40	[-1.33264, 1.63430]	0.00494	0.11287	0.05987	0.03037	0.00216	0.00001	0.13133	0.04776	0.03971	0.03067
Caso $p = 0$ e $q = 2$											
0.10	[-0.61079, 1.13376]	0.00000	0.01118	0.00429	0.00257	-0.00005	0.00001	0.01109	0.00415	0.00262	0.00003
0.20	[-0.64291, 1.18129]	0.00001	0.01755	0.00697	0.00435	-0.00204	0.00002	0.01912	0.00688	0.00497	-0.00166
0.30	[-0.72983, 1.19132]	0.00001	0.01841	0.00815	0.00424	0.00429	0.00001	0.02841	0.01035	0.00760	0.00568
0.40	[-0.99135, 1.14161]	0.00014	0.08068	0.03495	0.02185	-0.00568	0.00045	0.04138	0.02076	0.01108	-0.00127
Caso $p = 1$ e $q = 0$											
0.10	[-0.54824, 1.11127]	0.00001	0.00978	0.00287	0.00222	-0.00016	0.00000	0.00988	0.00294	0.00219	-0.00009
0.20	[-0.59041, 1.16200]	0.00000	0.02228	0.00788	0.00600	-0.00222	0.00001	0.02227	0.00913	0.00564	-0.00192
0.30	[-0.68085, 1.15582]	0.00001	0.02609	0.00673	0.00701	0.00133	0.00000	0.04001	0.01040	0.01056	0.00242
0.40	[-0.83652, 1.08953]	0.00427	0.05899	0.03367	0.01391	-0.00402	0.00002	0.05605	0.02195	0.01673	-0.00057
Caso $p = 1$ e $q = 1$											
0.10	[-0.49855, 1.22036]	0.00000	0.00558	0.00247	0.00170	-0.00206	0.00000	0.00570	0.00234	0.00165	-0.00176
0.20	[-0.55309, 1.22988]	0.00000	0.00773	0.00337	0.00188	0.00329	0.00114	0.01056	0.00504	0.00242	0.00478
0.30	[-0.65192, 1.16055]	0.00001	0.01829	0.00929	0.00521	-0.00318	0.00001	0.01380	0.00581	0.00427	0.00221
0.40	[-0.89830, 1.17148]	0.00007	0.09926	0.05082	0.02709	-0.00512	0.00325	0.08614	0.04046	0.02269	0.01200
Caso $p = 1$ e $q = 2$											
0.10	[-0.74639, 1.15353]	0.00000	0.01347	0.00464	0.00329	-0.00015	0.00001	0.01325	0.00446	0.00318	-0.00010
0.20	[-0.81275, 1.23285]	0.00004	0.01998	0.00851	0.00440	-0.00096	0.00001	0.01979	0.00887	0.00500	-0.00073
0.30	[-0.85539, 1.27573]	0.00002	0.03153	0.01052	0.00915	-0.00007	0.00003	0.03497	0.01542	0.00892	0.00077
0.40	[-0.97039, 1.34281]	0.00000	0.05344	0.02377	0.01709	-0.00398	0.00002	0.05255	0.01350	0.01269	-0.00131
Caso $p = 2$ e $q = 0$											
0.10	[-0.41177, 1.13921]	0.00001	0.00669	0.00307	0.00155	0.00008	0.00000	0.00689	0.00320	0.00153	0.00019
0.20	[-0.47126, 1.16627]	0.00048	0.01365	0.00591	0.00298	-0.00123	0.00003	0.01445	0.00671	0.00378	-0.00068
0.30	[-0.57101, 1.14793]	0.00000	0.00966	0.00321	0.00271	-0.00098	0.00003	0.01378	0.00594	0.00409	0.00100
0.40	[-0.73706, 1.04991]	0.00392	0.09501	0.03701	0.02519	-0.00610	0.00012	0.05847	0.02351	0.01837	0.00020
Caso $p = 2$ e $q = 1$											
0.10	[-0.85372, 1.34330]	0.00000	0.00468	0.00171	0.00121	-0.00068	0.00000	0.00460	0.00190	0.00116	-0.00012
0.20	[-0.94457, 1.33676]	0.00016	0.00678	0.00333	0.00147	-0.00333	0.00000	0.00313	0.00125	0.00078	-0.00060
0.30	[-1.10332, 1.30147]	0.00451	0.04146	0.02227	0.01010	-0.02088	0.00002	0.03411	0.01214	0.00768	-0.01100
0.40	[-1.34264, 1.57039]	0.00203	0.11936	0.05787	0.02878	0.01447	0.00005	0.12517	0.05630	0.04055	0.04584
Caso $p = 2$ e $q = 2$											
0.10	[-0.34895, 1.13512]	0.00000	0.00934	0.00343	0.00218	-0.00002	0.00004	0.00934	0.00333	0.00231	0.00007
0.20	[-0.37788, 1.17435]	0.00002	0.01783	0.00565	0.00385	-0.00205	0.00003	0.01846	0.00612	0.00425	-0.00163
0.30	[-0.50619, 1.17573]	0.00000	0.02056	0.00507	0.00519	0.00385	0.00000	0.02480	0.00936	0.00734	0.00538
0.40	[-0.74588, 1.08122]	0.00007	0.07204	0.03623	0.01645	-0.00292	0.00511	0.04879	0.02526	0.01123	0.00193

Apêndice F

Função Densidade Espectral

Neste apêndice apresentamos exemplos gráficos do comportamento das funções densidade espectral dos processos $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$ e $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$. Ressaltamos que, pela expressão (4.11), espera-se que ambas as funções apresentem um comportamento assintótico semelhante, quando $\lambda \to 2\pi k/s$, para $k = 0, \cdots, \lfloor s/2 \rfloor$.

No que segue,

- para os gráficos de $f_{\ln(h_t^2)}(\cdot)$, sem perda de generalidade, fixamos $\sigma_g^2 = 1$;
- para os gráficos de $f_{\ln(X_t^2)}(\cdot)$, consideramos os casos $Z_0 \sim \mathcal{N}(0,1)$ e $Z_t = Z_t^*\sqrt{(\nu-2)/\nu}$, onde $Z_t^* \sim t_\nu$, com $\nu \in \{3, 6\}$. Tal comparação mostra-se necessária pois, no caso de $f_{\ln(X_t^2)}(\cdot)$ temos três constantes que dependem da distribuição de Z_0: $\sigma_g^2 := \mathrm{Var}(g(Z_0))$, $C_1 := \mathrm{Cov}\left(\ln\left(Z_0^2\right), g(Z_0)\right))$ e $\sigma^2_{\ln(Z_t^2)} := \mathrm{Var}(\ln(Z_t^2))$;
- consideramos $\theta = 0.15$ e $\gamma \in \{-0.25, 0.25\}$, sendo que esses valores são utilizados somente para a função densidade espectral do processo $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$. Note que, como as funções de distribuição consideradas são simétricas, o sinal de θ não tem influência no comportamento da função densidade espectral pois (4.10) e (4.12) são reescritas como

$$\sigma_g^2 = \theta^2 + \gamma^2\big(1 - \big[\mathbb{E}(|Z_0|)\big]^2\big) \quad \text{e} \quad C_1 = \gamma\big[\mathbb{E}\big(|Z_0|\ln(Z_0^2)\big) - \mathbb{E}\big(|Z_0|\big)\mathbb{E}\big(\ln(Z_0^2)\big)\big];$$

- os valores de $\mathbb{E}(|Z_0|), \mathbb{E}\big(|Z_0|\ln(Z_0^2)\big)$, $\mathbb{E}\big(\ln(Z_0^2)\big)$ e $\sigma^2_{\ln(Z_t^2)}$, para cada função de distribuição considerada neste estudo, e os respectivos valores de σ_g^2 e C_1, para $\gamma = 0.25$, são dados na Tabela F.1. Note que σ_g^2 não depende do sinal de γ e o valor de C_1, para $\gamma = -0.25$, pode ser obtido multiplicando-se por -1 o valor dado na Tabela F.1;
- para o caso $p = q = 0$, consideramos $s \in \{1, 2, 4\}$ e $d \in \{0.05,\ 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.499\}$;
- quando $p \neq 0$ ou $q \neq 0$, fixamos $s = 4$, $d = 0.25$. Consideramos ainda todas as possibilidades para $a_1, b_1 \in \{-0.9,\ -0.5,\ -0.1,\ 0.1,\ 0.5,\ 0.9\}$, com $a_1 \neq b_1$;
- para melhor visualização das características da função densidade espectral, todos os gráficos estão na mesma escala e foram truncados no eixo y;
- todos os gráficos consideram $\lambda \in [0, 2\pi]$ para auxiliar na identificação do número de picos que a função apresenta nesse intervalo. Entretanto, levando-se em conta a simetria e periodicidade da função densidade espectral, voltamos nossa atenção para o comportamento da mesma no intervalo $[0, \pi]$.

Tabela F.1: Valores de $\mathbb{E}(|Z_0|), \mathbb{E}\big(|Z_0|\ln(Z_0^2)\big)$, $\mathbb{E}\big(\ln(Z_0^2)\big)$ e $\sigma^2_{\ln(Z_t^2)}$ e das constantes σ_g^2 e C_1 para as funções de distribuição Gaussiana e t_ν padronizada, com $\nu \in \{3, 6\}$, para $\theta = 0.15$ e $\gamma = 0.25$.

Distribuição	$\mathbb{E}(\lvert Z_0\rvert)$	$\mathbb{E}\big(\lvert Z_0\rvert\ln(Z_0^2)\big)$	$\mathbb{E}\big(\ln(Z_0^2)\big)$	$\sigma^2_{\ln(Z_t^2)}$	σ_g^2	C_1
t_3	0.6366	0.0000	−2.0000	5.8696	0.0597	0.3183
t_6	0.7500	0.0794	−1.5000	5.3297	0.0498	0.3011
$\mathcal{N}(0,1)$	0.7979	0.0925	−1.2704	4.9348	0.0452	0.2765

Caso 1: $p = q = 0$

Na Figura F.1 apresentamos a função densidade espectral do processo $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$, fixando-se $\sigma_g^2 = 1$, isto é,

$$f_{\ln(h_t^2)}(\lambda) = \frac{1}{2\pi}\left[2\sin\left(\frac{s\lambda}{2}\right)\right]^{-2d}, \quad \text{para todo } \lambda \in [0, \pi].$$

Nas Figuras F.2 e F.3 apresentamos a função densidade espectral do processo estocástico $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, considerando-se diferentes distribuições para as inovações, ou seja,

$$f_{\ln(X_t^2)}(\lambda) = \frac{\sigma_g^2}{2\pi}\left[2\sin\left(\frac{s\lambda}{2}\right)\right]^{-2d} + \frac{C_1}{\pi}\mathfrak{Re}\Big(e^{-\mathrm{i}\lambda}(1-e^{-\mathrm{i}s\lambda})^{-d}\Big) + \frac{\sigma^2_{\ln(Z_t^2)}}{2\pi}, \quad \text{para todo } \lambda \in [0, \pi],$$

com $\sigma^2_{\ln(Z_t^2)}$, σ_g^2 e C_1 dados na Tabela F.1.

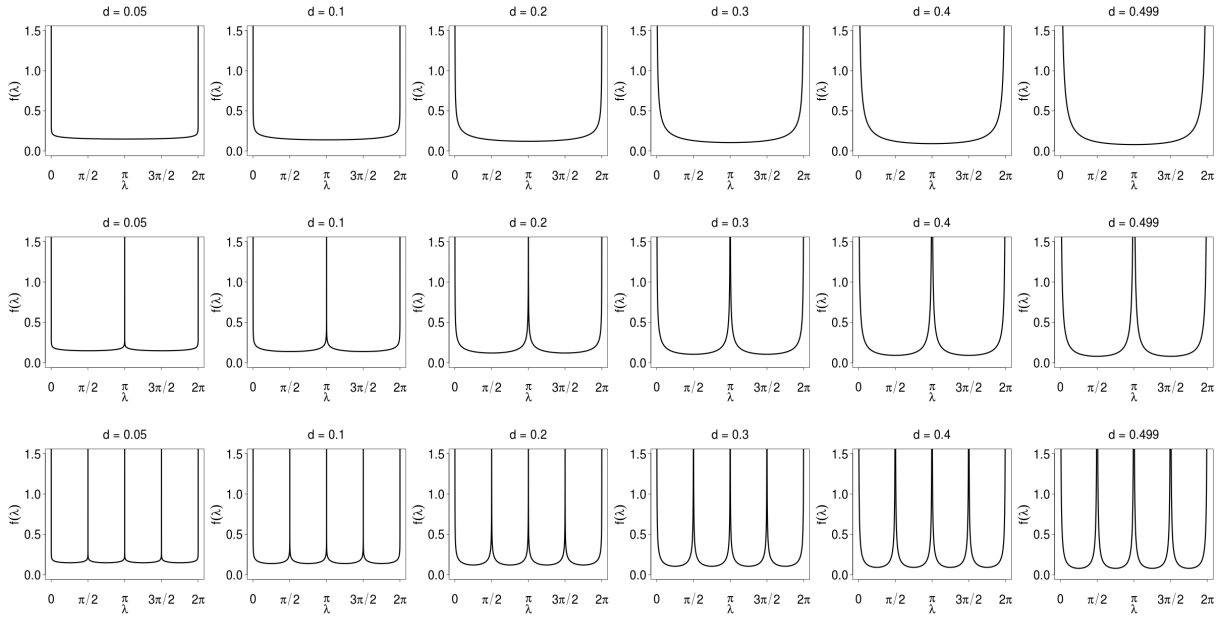


Figura F.1: Função densidade espectral do processo $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$, para $p = q = 0$, $d \in \{0.05,\ 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.499\}$ e $s \in \{1, 2, 4\}$.

Pela expressão (4.13) conclui-se que $f_{\ln(h_t^2)}(\frac{2k\pi}{s}) = \infty$, para todo $0 < d < 0.5$ e $k \in \mathbb{N}$. Consequentemente, pela expressão (4.11), $f_{\ln(X_t^2)}(\frac{2k\pi}{s}) = \infty$, para todo $0 < d < 0.5$ e $k \in \mathbb{N}$. Pela Figura F.1 observa-se que, para s fixo e $\lambda \to \frac{2k\pi}{s}$, para qualquer $k \in \{0, \cdots, s\}$, a função $f_{\ln(h_t^2)}(\cdot)$ tende ao infinito mais rapidamente, com o aumento de d.

Nas Figuras F.2 e F.3, a relação entre a velocidade de convergência ao infinito da função $f_{\ln(X_t^2)}(\cdot)$ e o valor do parâmetro d não é visível como no caso da função $f_{\ln(h_t^2)}(\cdot)$. Entretanto, observa-se que a função $f_{\ln(X_t^2)}(\cdot)$ apresenta comportamentos bem distintos para $\gamma < 0$ e $\gamma > 0$. Enquanto que, para

$\gamma < 0$ o mínimo da função é atingido próximo da frequência $\lambda = 0$, para $\gamma > 0$, a função atinge seu mínimo próximo de $\lambda = \pi$. Tal comportamento é o mesmo para qualquer valor de $s \in \{1, 2, 4\}$.

Como era esperado, para valores maiores de ν, a função densidade espectral aproxima-se da correspondente Gaussiana. Observa-se ainda que, quando consideramos $Z_t^* \sim t_\nu$, com $\nu \in \{3, 6\}$, os valores da função densidade espectral são mais elevados do que os correspondentes valores para o caso Gaussiano (veja a linha pontilhada em vermelho). Tal comportamento deixa clara a influência de $f_{\ln(Z_t^2)}(\lambda) = \sigma^2_{\ln(Z_t^2)} \times 1/(2\pi)$ pois, como mostra a Tabela F.1, as constantes σ_g^2 e C_1 assumem valores pequenos e não variam muito de uma distribuição para outra.

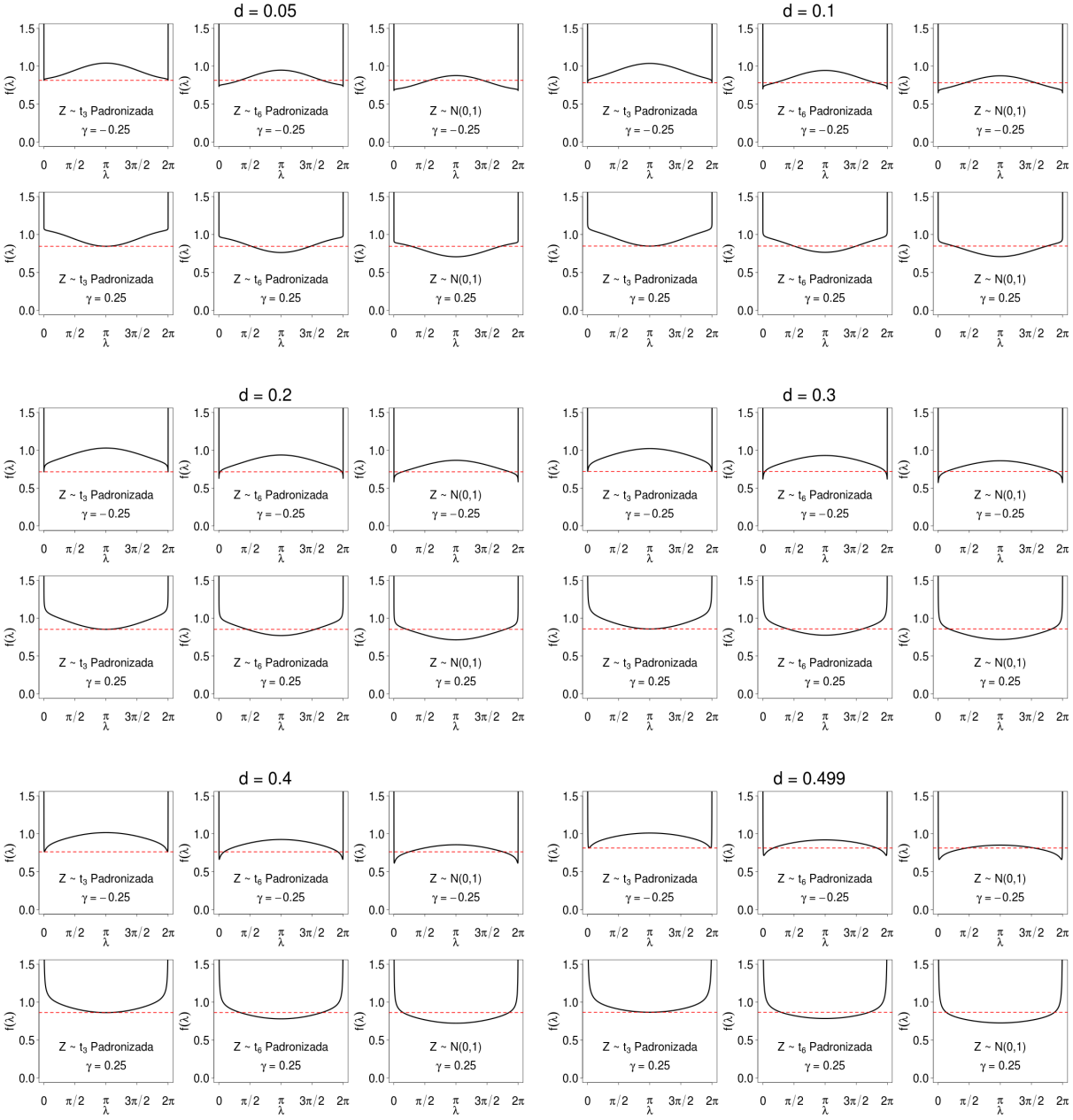
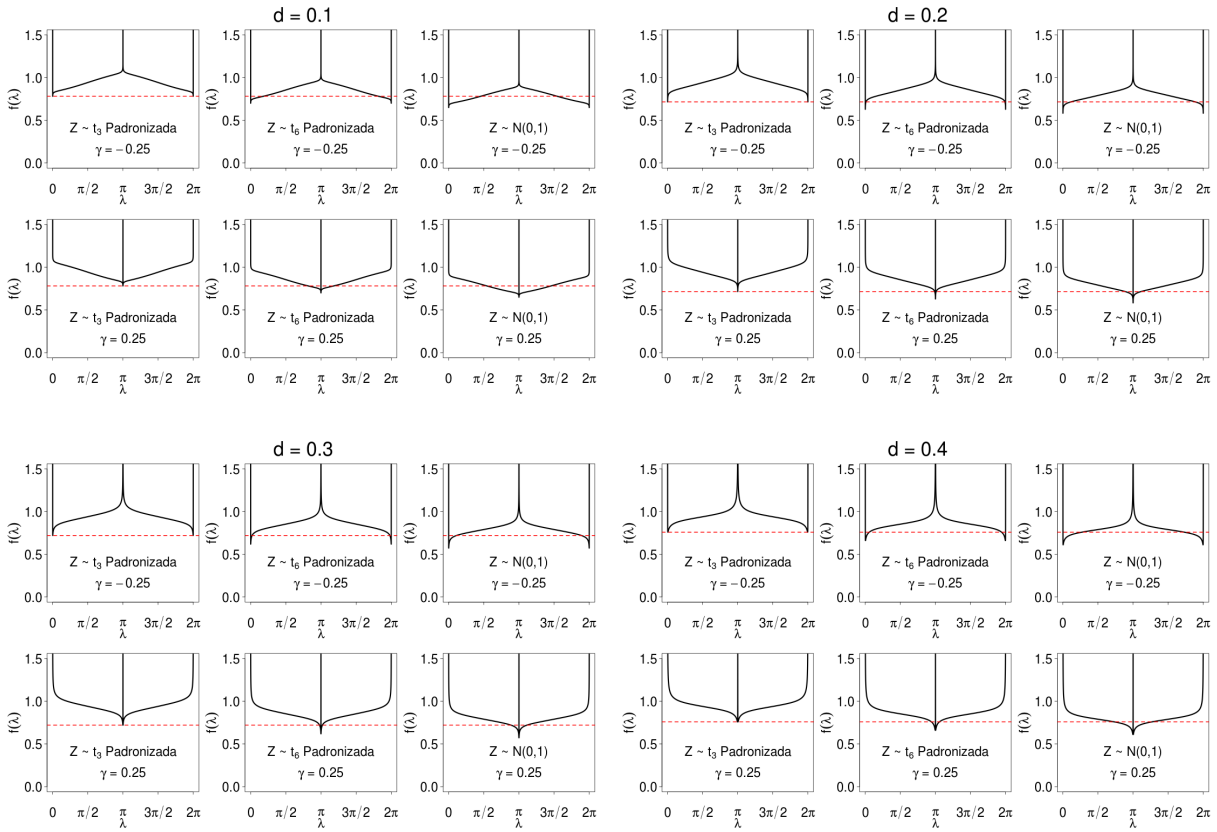
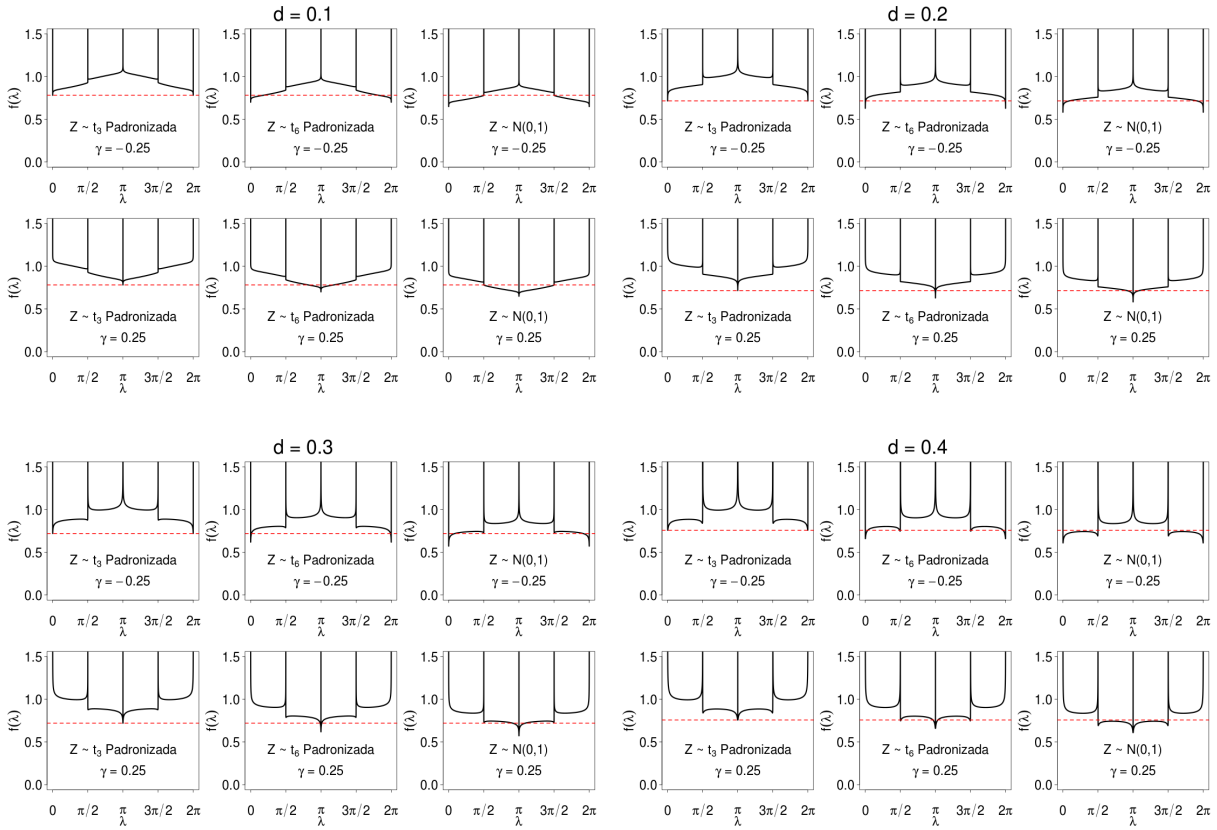


Figura F.2: Função densidade espectral do processo $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, para $s = 1$. Para cada valor de $d \in \{0.05, 0.1, 0.2, 0.3, 0.4, 0.499\}$, $Z_t = \sqrt{\frac{\nu-2}{\nu}} Z_t^*$, onde $Z_t^* \sim t_\nu$, para $\nu \in \{3, 6\}$ e $Z_t \sim \mathcal{N}(0, 1)$, $\theta = 0.15$ e $\gamma \in \{0.25, 0.25\}$.



(a) $s = 2$



(b) $s = 4$

Figura F.3: Função densidade espectral do processo $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, para $s \in \{2, 4\}$. Para cada valor de $d \in \{0.1, 0.2, 0.3, 0.4\}$: $Z_t = \sqrt{\frac{\nu-2}{\nu}} Z_t^*$, onde $Z_t^* \sim t_\nu$, para $\nu \in \{3, 6\}$ e $Z_t \sim \mathcal{N}(0, 1)$, $\theta = 0.15$ e $\gamma \in \{-0.25, 0.25\}$: (a) $s = 2$; (b) $s = 4$.

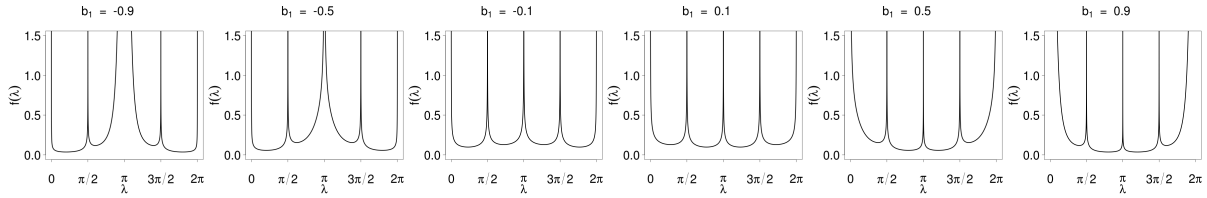


Figura F.4: Função densidade espectral do processo $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$, para $s = 4$, $d = 0.25$, $p = 0$, $q = 1$, com $b_1 \in \{-0.9,\ -0.5,\ -0.1,\ 0.1,\ 0.5,\ 0.9\}$ e $\sigma_g^2 = 1$.

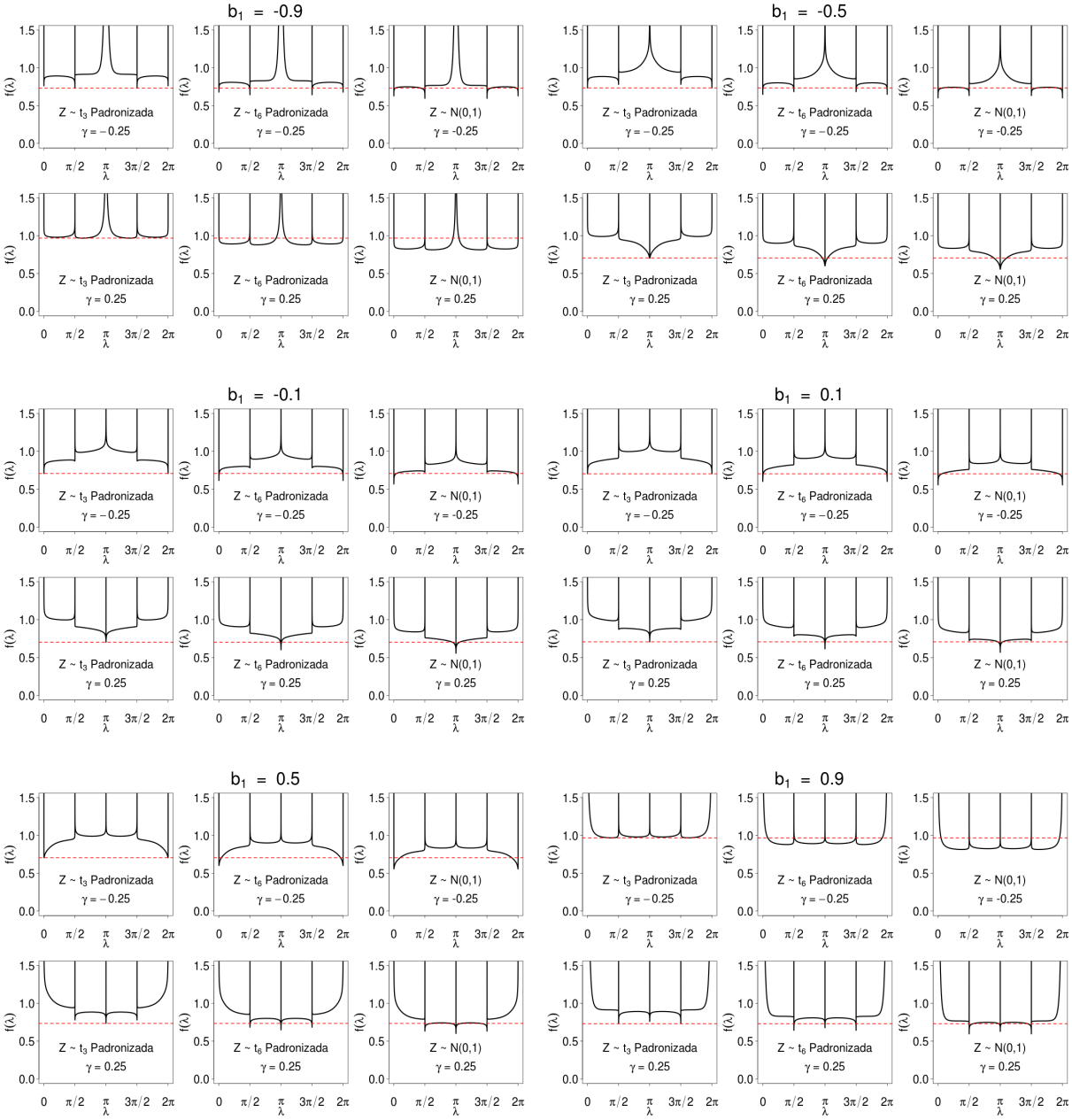


Figura F.5: Função densidade espectral do processo $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, para $s = 4$, $d = 0.25$, $p = 0$, $q = 1$. Para cada valor de $b_1 \in \{-0.9,\ -0.5,\ -0.1,\ 0.1,\ 0.5,\ 0.9\}$: $Z_t = \sqrt{\frac{\nu-2}{\nu}} Z_t^*$, onde $Z_t^* \sim t_\nu$, para $\nu \in \{3, 6\}$ e $Z_t \sim \mathcal{N}(0,1)$, $\theta = 0.15$ e $\gamma \in \{-0.25, 0.25\}$.

Caso 2: $p = 0$ e $q = 1$

Nas Figuras F.4 e F.5 apresentamos, respectivamente, a função densidade espectral do processo $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$, fixando-se $\sigma_g^2 = 1$,

$$f_{\ln(h_t^2)}(\lambda) = \frac{1}{2\pi}\frac{1}{|1 - b_1 e^{-\mathrm{i}\lambda}|^2}\left[2\sin\left(\frac{s\lambda}{2}\right)\right]^{-2d}, \quad \text{para todo } \lambda \in [0, \pi],$$

e a função densidade espectral do processo estocástico $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, considerando-se diferentes distribuições para as inovações,

$$f_{\ln(X_t^2)}(\lambda) = \frac{\sigma_g^2}{2\pi}\frac{1}{|1 - b_1 e^{-\mathrm{i}\lambda}|^2}\left[2\sin\left(\frac{s\lambda}{2}\right)\right]^{-2d} + \frac{C_1}{\pi}\mathfrak{Re}\left(\frac{e^{-\mathrm{i}\lambda}}{1 - b_1 e^{-\mathrm{i}\lambda}}(1 - e^{-\mathrm{i}s\lambda})^{-d}\right) + \frac{\sigma^2_{\ln(Z_t^2)}}{2\pi},$$

para todo $\lambda \in [0, \pi]$, com $\sigma^2_{\ln(Z_t^2)}$, σ_g^2 e C_1 dados na Tabela F.1.

Pela Figura F.4 observamos que, para $|b_1|$ próximo de zero, $f_{\ln(h_t^2)}(\cdot)$ tem comportamento semelhante ao caso $p = q = 0$ (Caso 1). Para valores negativos de b_1, a função tende ao infinito, quando $\lambda \to \pi$, mais rapidamente do que quando $b_1 > 0$. Por outro lado, quando $\lambda \to 0$, a divergência para o infinito é mais rápida para $b_1 > 0$. Além disso, em ambos os casos, a divergência é mais rápida para $|b_1|$ próximo de 1 do que próximo de 0.

Pela Figura F.5, observamos que a região onde a $f_{\ln(X_t^2)}(\cdot)$ atinge o seu mínimo depende, não somente do sinal de γ, mas também, do sinal de b_1. Assim como $f_{\ln(h_t^2)}(\cdot)$, a função $f_{\ln(X_t^2)}(\cdot)$ apresenta comportamento semelhante ao caso $p = q = 0$ (Caso 1), quando $|b_1| = 0.1$. Novamente, para a distribuição t-Student, os valores de $f_{\ln(X_t^2)}(\cdot)$ são mais distantes da origem do que para a distribuição Gaussiana. Observa-se também que, para $\gamma < 0$, as mudanças no contorno do gráfico $f_{\ln(X_t^2)}(\cdot)$ são mais pronunciadas quando b_1 passa de 0.5 para 0.9 do que quando esse parâmetro muda de ± 0.1 para ± 0.5 ou de -0.5 para -0.9. Para $\gamma > 0$, a mudança no traçado da função é mais visível quando b_1 passa de -0.9 para -0.5 ou de 0.1 para 0.5, do que quando este valor muda de -0.5 para -0.1 ou de 0.5 para 0.9.

Caso 3: $p = 1$ e $q = 0$

Nas Figuras F.6 e F.7 apresentamos, respectivamente, a função densidade espectral do processo $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$, fixando-se $\sigma_g^2 = 1$,

$$f_{\ln(h_t^2)}(\lambda) = \frac{1}{2\pi}|1 - a_1 e^{-\mathrm{i}\lambda}|^2\left[2\sin\left(\frac{s\lambda}{2}\right)\right]^{-2d}, \quad \text{para todo } \lambda \in [0, \pi],$$

e a função densidade espectral do processo estocástico $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, considerando-se diferentes distribuições para as inovações,

$$f_{\ln(X_t^2)}(\lambda) = \frac{\sigma_g^2}{2\pi}|1 - a_1 e^{-\mathrm{i}\lambda}|^2\left[2\sin\left(\frac{s\lambda}{2}\right)\right]^{-2d} + \frac{C_1}{\pi}\mathfrak{Re}\left(e^{-\mathrm{i}\lambda}(1 - a_1 e^{-\mathrm{i}\lambda})(1 - e^{-\mathrm{i}s\lambda})^{-d}\right) + \frac{\sigma^2_{\ln(Z_t^2)}}{2\pi},$$

para todo $\lambda \in [0, \pi]$, com σ_g^2 e C_1 dados na Tabela F.1.

Pela Figura F.6 observamos que, assim como no caso $p = 0$ e $q = 1$ (Caso 2), $f_{\ln(h_t^2)}(\cdot)$ tem um comportamento semelhante ao caso $p = q = 0$ (Caso 1) quando $|b_1| = 0.1$. A divergência para o infinito neste caso também é mais rápida para $|b_1|$ próximo de 1 do que próximo de 0. Entretanto, quando $\lambda \to 0$ ou $\lambda \to \pi$, a convergência dá-se de forma oposta ao que ocorre no caso $p = 0$ e $q = 1$ (Caso 2). Isto é, quando $\lambda \to 0$, a divergência para o infinito é mais rápida para $b_1 < 0$ e, quando $\lambda \to \pi$, é mais rápida para $b_1 > 0$.

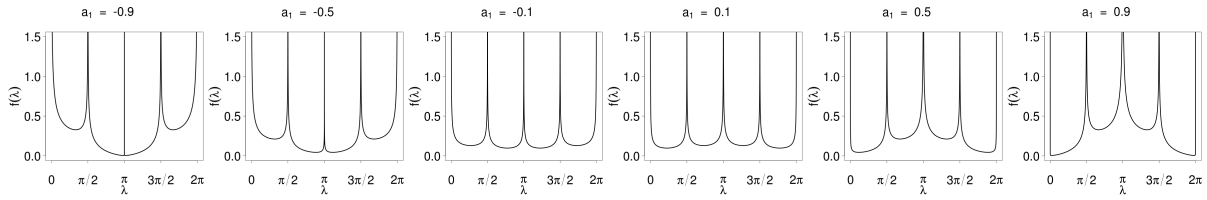


Figura F.6: Função densidade espectral do processo $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$, para $s = 4$, $d = 0.25$, $p = 1$, $q = 0$, com $a_1 \in \{-0.9,\ -0.5,\ -0.1,\ 0.1,\ 0.5,\ 0.9\}$ e $\sigma_g^2 = 1$.

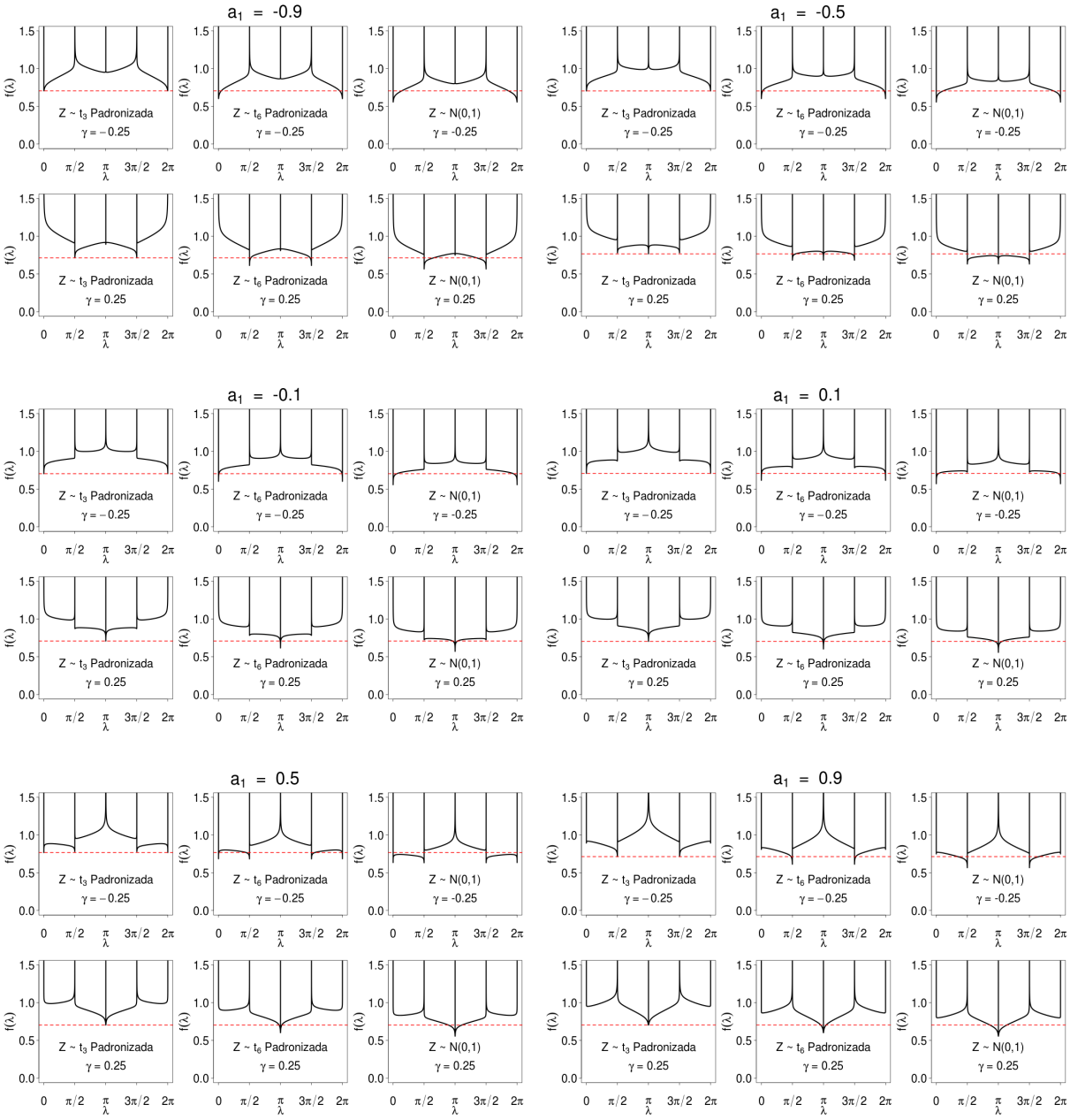


Figura F.7: Função densidade espectral do processo $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, para $s = 4$, $d = 0.25$, $p = 1$, $q = 0$. Para cada valor de $a_1 \in \{-0.9,\ -0.5,\ -0.1,\ 0.1,\ 0.5,\ 0.9\}$: $Z_t = \sqrt{\frac{\nu-2}{\nu}} Z_t^*$, onde $Z_t^* \sim t_\nu$, para $\nu \in \{3, 6\}$ e $Z_t \sim \mathcal{N}(0,1)$, $\theta = 0.15$ e $\gamma \in \{-0.25, 0.25\}$.

Pela Figura F.7, observamos que a região onde a $f_{\ln(X_t^2)}(\cdot)$ atinge o seu mínimo depende, não

somente do sinal de γ, mas também, do sinal de a_1. Sendo que a diferença mais marcante está nos gráficos para $a_1 = -0.9$ e 0.9, para qualquer γ fixo. Assim como nos exemplos anteriores, a função $f_{\ln(X_t^2)}(\cdot)$ apresenta comportamento semelhante ao caso $p = q = 0$ (Caso 1), quando $|a_1| = 0.1$. Novamente, para a distribuição t-Student, os valores de $f_{\ln(X_t^2)}(\cdot)$ são mais distantes da origem do que para a distribuição Gaussiana. Ao contrário dos casos anteriores, o traçado da função varia lentamente, quando a_1 cresce/decresce, para qualquer γ fixo.

Caso 4: $p = 1$ e $q = 1$

Nas Figuras F.8 e F.9 - F.12 apresentamos, respectivamente, a função densidade espectral do processo $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$, fixando-se $\sigma_g^2 = 1$,

$$f_{\ln(h_t^2)}(\lambda) = \frac{1}{2\pi}\left|\frac{1 - a_1 e^{-\mathrm{i}\lambda}}{1 - b_1 e^{-\mathrm{i}\lambda}}\right|^2 \left[2\sin\left(\frac{s\lambda}{2}\right)\right]^{-2d}, \quad \text{para todo } \lambda \in [0, \pi],$$

e a função densidade espectral do processo estocástico $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, considerando-se diferentes distribuições para as inovações,

$$f_{\ln(X_t^2)}(\lambda) = \frac{\sigma_g^2}{2\pi}\left|\frac{1 - a_1 e^{-\mathrm{i}\lambda}}{1 - b_1 e^{-\mathrm{i}\lambda}}\right|^2 \left[2\sin\left(\frac{s\lambda}{2}\right)\right]^{-2d} + \frac{C_1}{\pi}\mathfrak{Re}\left(e^{-\mathrm{i}\lambda}\frac{1 - a_1 e^{-\mathrm{i}\lambda}}{1 - b_1 e^{-\mathrm{i}\lambda}}(1 - e^{-\mathrm{i}s\lambda})^{-d}\right) + \frac{\sigma^2_{\ln(Z_t^2)}}{2\pi},$$

para todo $\lambda \in [0, \pi]$, com $\sigma^2_{\ln(Z_t^2)}$, σ_g^2 e C_1 dados na Tabela F.1.

Pela Figura F.8, observamos que a mudança no traçado de $f_{\ln(h_t^2)}(\cdot)$ ocorre tanto com a variação de a_1 quanto de b_1. Para $a_1 < 0$ e $b_1 < a_1$, a função tem comportamento semelhante ao caso $p = 0$ e $q = 1$ (Caso 2), com os mesmos valores de b_1. Para $a_1 < 0$ e $b_1 > a_1$, a função tem comportamento semelhante ao caso $p = 1$, $q = 0$ (Caso 3) com os mesmos valores de a_1. Para $a_1 > 0$ e $b_1 < a_1$, a função tem comportamento semelhante ao caso $p = 1$, $q = 0$ (Caso 3). Para $a_1 > 0$ e $b_1 > a_1$, a função tem comportamento semelhante ao caso $p = 0$, $q = 1$ (Caso 2).

Nas Figuras F.9 - F.12, observamos que o comportamento de $f_{\ln(X_t^2)}(\cdot)$ pode ser sempre comparado aos casos $p = 0$ e $q = 1$ ou $p = 1$ e $q = 0$. Obviamente, ao contrário de $f_{\ln(h_t^2)}(\cdot)$, não só os coeficientes a_1 e b_1 devem ser levados em conta, mas também o sinal de γ. A conclusão mais importante desta análise é a de que não existe nenhuma característica marcante nos gráficos das Figuras F.9 - F.12 que possibilite distinguir um processo SFIEGARCH$(1, d, 1)_s$ de um SFIEGARCH$(0, d, 1)_s$ ou de um SFIEGARCH$(1, d, 0)_s$.

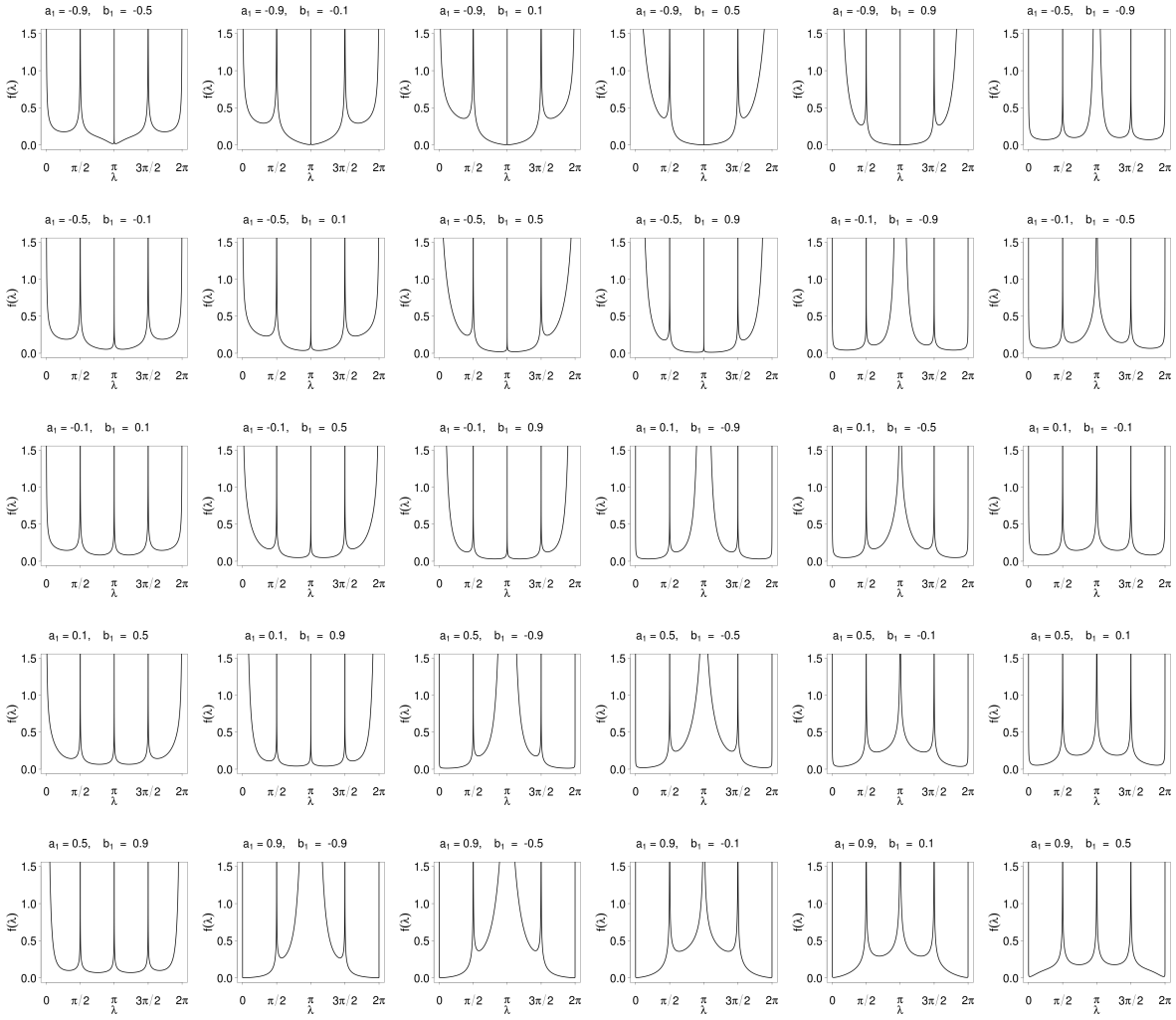


Figura F.8: Função densidade espectral do processo $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$, para $s = 4$, $d = 0.25$, $p = 1$, $q = 1$, com $a_1, b_1 \in \{-0.9, -0.5, -0.1, 0.1, 0.5, 0.9\}$ tais que $a_1 \neq b_1$, e $\sigma_g^2 = 1$.

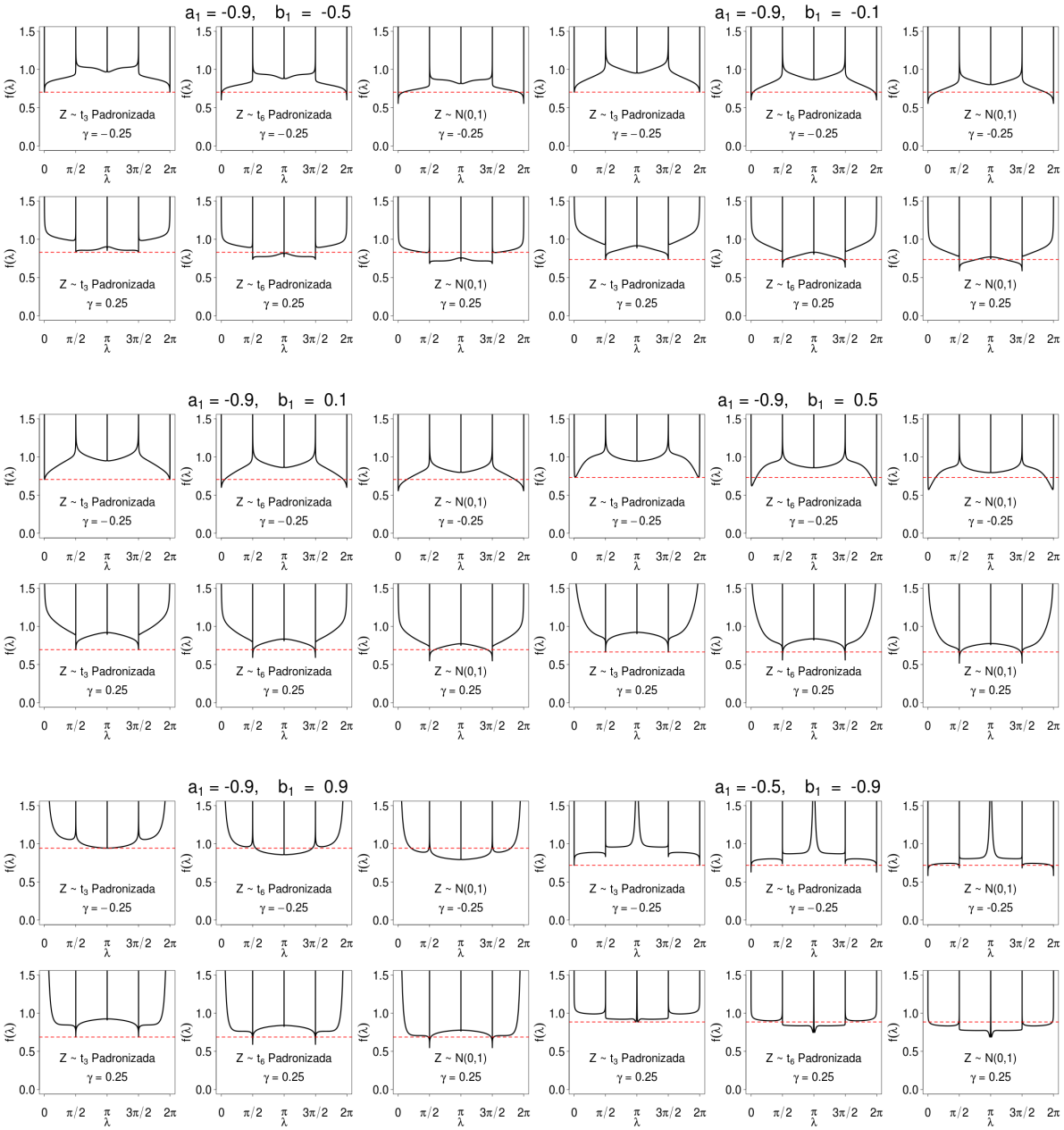


Figura F.9: Função densidade espectral do processo $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$, para $s = 4$, $d = 0.25$, $p = 1$, $q = 0$. Para cada valor de $a_1, b_1 \in \{-0.9, -0.5, -0.1, 0.1, 0.5, 0.9\}$, com $a_1 \neq b_1$: $Z_t = \sqrt{\frac{\nu-2}{\nu}} Z_t^*$, onde $Z_t^* \sim t_\nu$, para $\nu \in \{3, 6\}$ e $Z_t \sim \mathcal{N}(0, 1)$, $\theta = 0.15$ e $\gamma \in \{-0.25, 0.25\}$. **(Continua)**

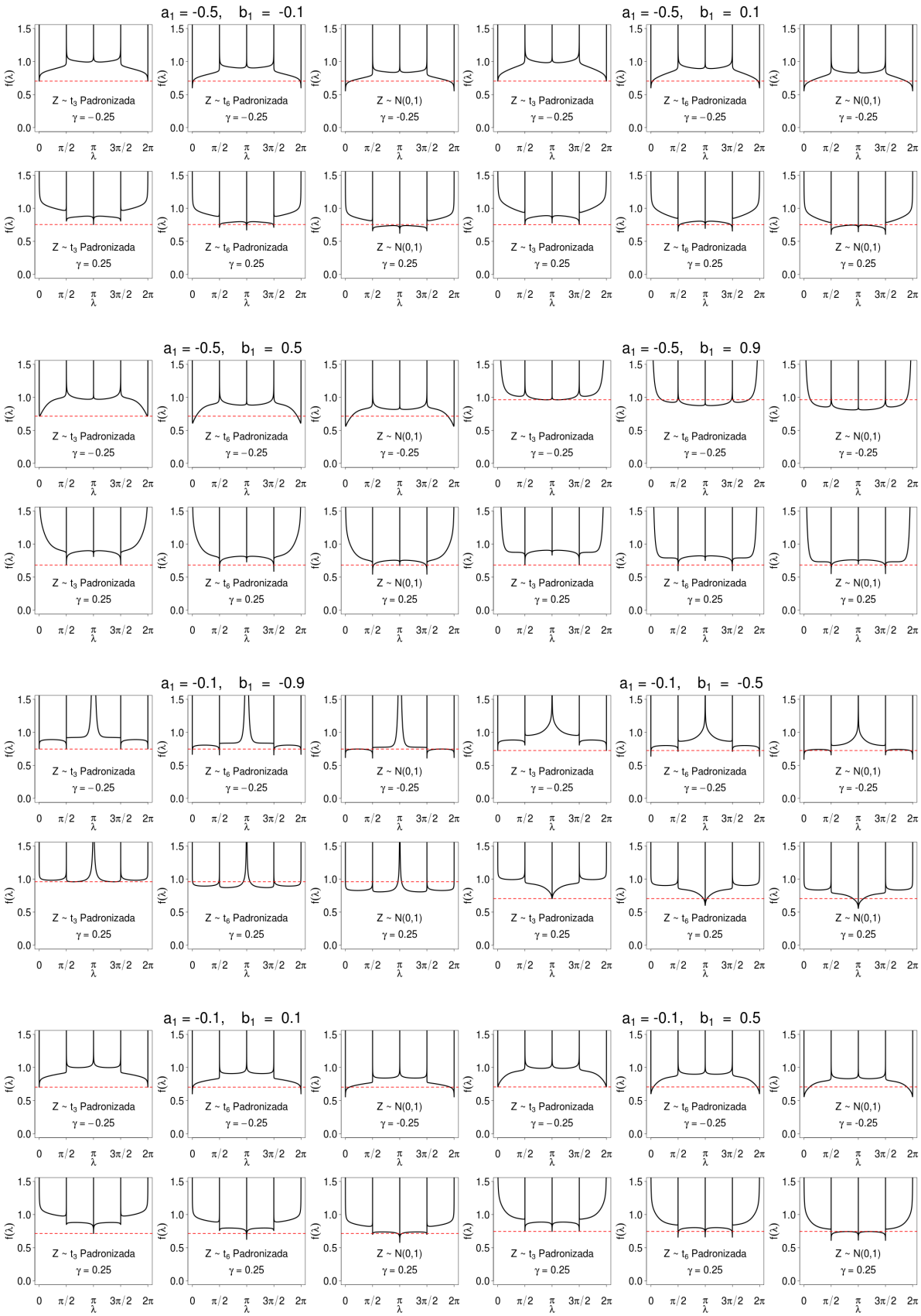


Figura F.10: (Continuação da Figura F.9). Função densidade espectral do processo $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$

.

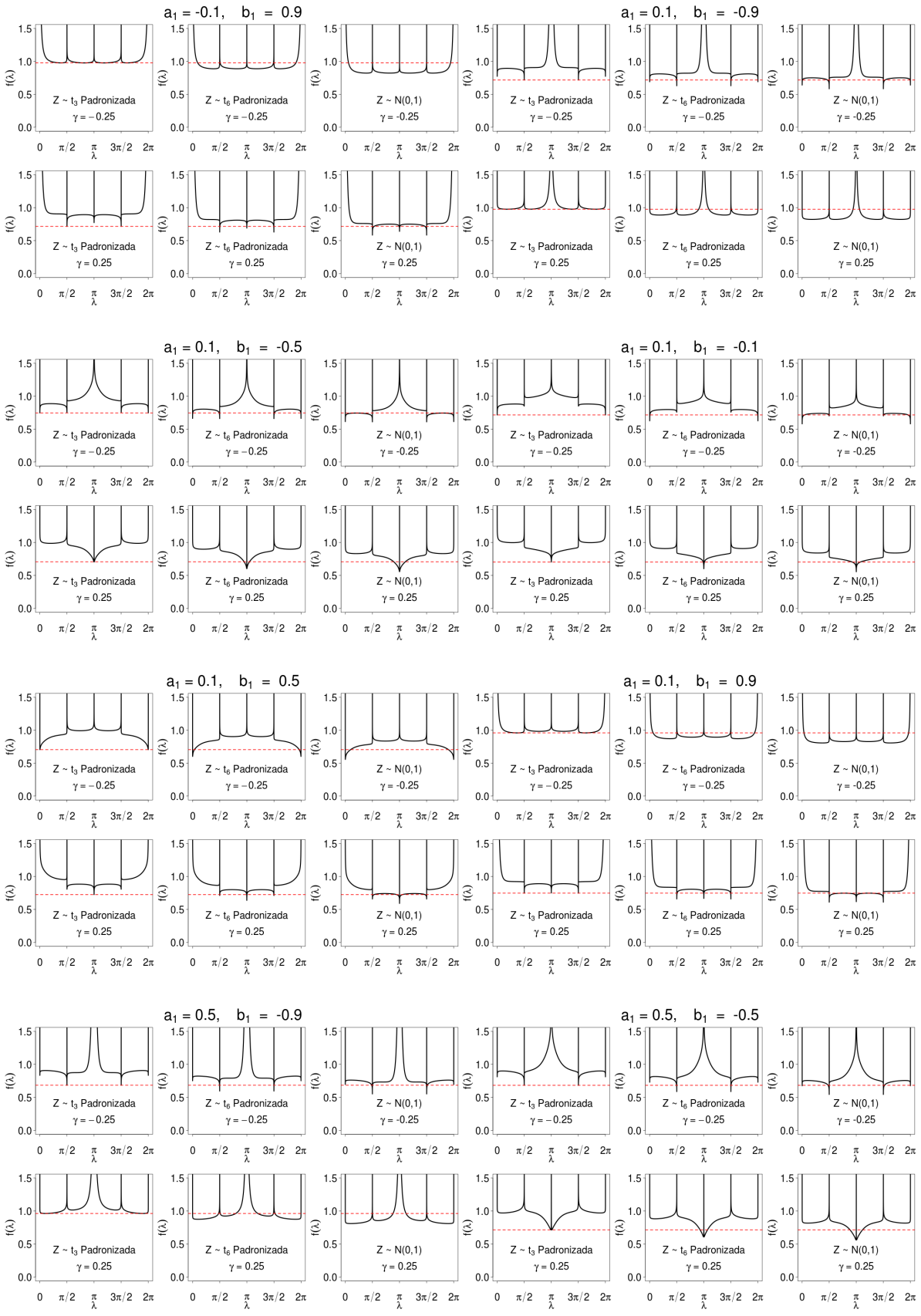


Figura F.11: (Continuação da Figura F.9). Função densidade espectral do processo $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$

.

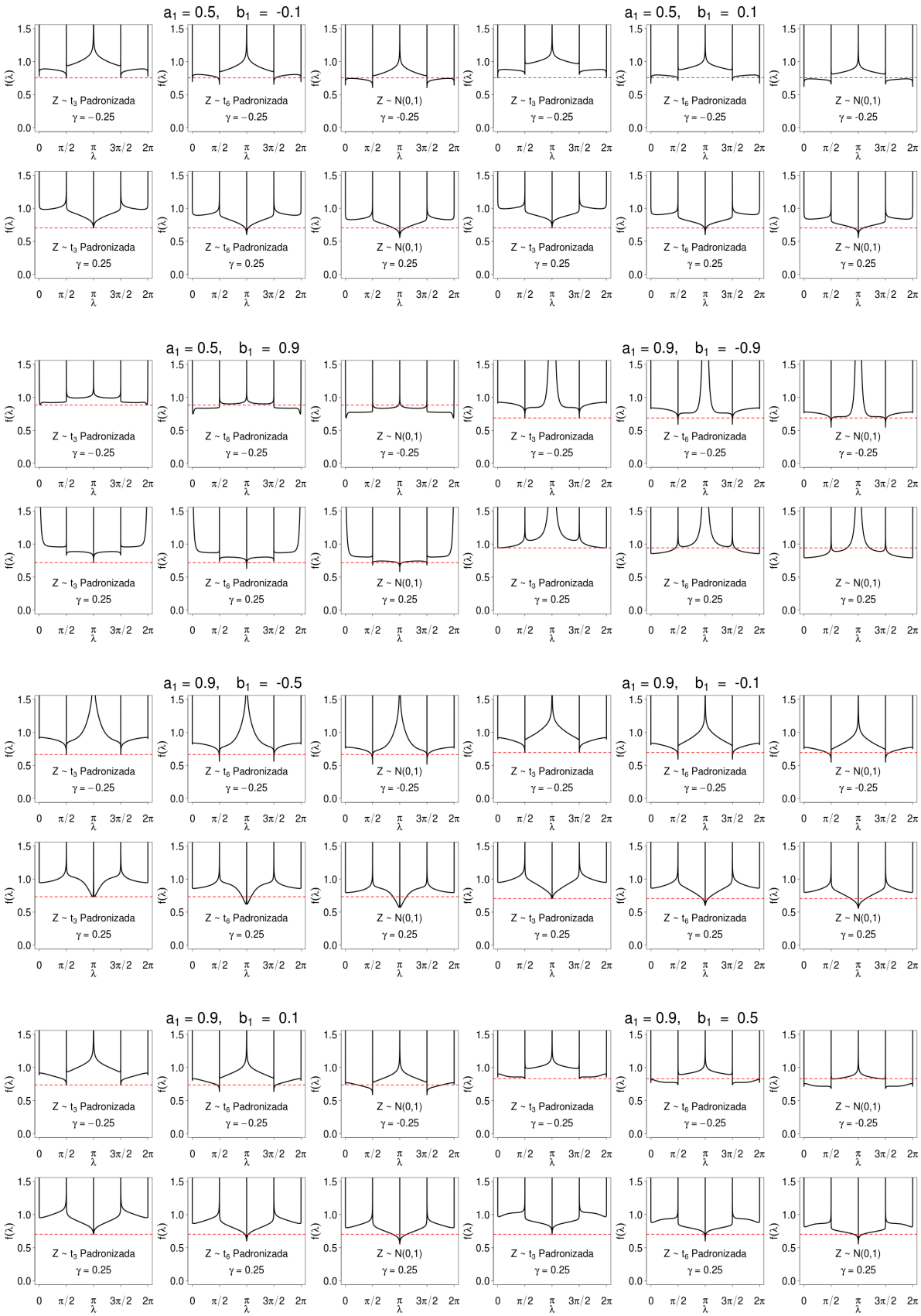


Figura F.12: (Continuação da Figura F.9). Função densidade espectral do processo $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$

.

Apêndice G

Artigo Prass e Lopes (2012)

Autores: Taiane S. Prass e Sílvia R. C. Lopes

Título: Quasi-likelihood Estimation on SFIEGARCH Process: a Simulation Study

Revista: Proceedings of the Joint Statistical Meetings. Alexandria, VA: American Statistical Association, vol. 1, 1313-1323

Ano: 2012

Quasi-likelihood Estimation on SFIEGARCH Process: a Simulation Study

Taiane S. Prass *† Sílvia R.C. Lopes

Abstract
The theoretical properties of the quasi-maximum likelihood estimate (QMLE) for SFIEGARCH models are still an open issue. This work presents a simulation study to access the finite sample performance of this estimator for misspecified SFIEGARCH$(0, d, 0)_s$ process. The simulation study presented here considers the Student's t and the generalized error distribution (GED) as underlying distributions for the innovation process. In order to obtain lighter and heavier tails than the Gaussian distribution, different values for the nuisance parameter ν are considered for both underlying distributions.

Key Words: SFIEGARCH Process; QMLE Method; Misspecification.

1. Introduction

In the last years economists have noticed that volatility of high frequency financial time series shows long range dependence merged with periodic behavior. According to Bordignon et al. (2009), in the case of exchange rate returns, this pattern is generally attributed to different openings of European, Asian and North American markets superimposed each other. Stock markets present a similar pattern, mainly due to the so-called time-of-day phenomena, such as market opening, closing operations, lunch-hour and overlapping effects.

To account for the long memory periodic behavior observed in financial time series, Bordignon et al. (2009) and Lopes and Prass (2012) introduce, respectively, the PLM-EGARCH and the SFIEGARCH processes. Both processes are extensions of the well known FIEGARCH process (Nelson, 1991). A discussion on their similarities and differences can be found in Lopes and Prass (2012).

This work presents a simulation study to access the finite sample performance of quasi-maximum likelihood estimate (QMLE) for SFIEGARCH$(0, d, 0)_s$ processes when the underlying distribution functions are misspecified. The asymptotic properties for the quasi-likelihood method are well established for ARCH/GARCH models (see, for instance, Lee and Hansen, 1994; Lumsdaine, 1996; Berkes et al., 2003; Berkes and Horváth, 2003; Hall and Yao, 2003) and also for EGARCH models (see, for instance, Straumann and Mikosch, 2006). However, at our best knowledge, there are no theoretical results regarding SFIEGARCH processes in the literature.

The paper is organized as follows. Section 2 introduces SFIEGARCH processes, presents some results regarding stationarity and ergodicity of these processes and gives a recurrence formula to calculate the coefficients for the polynomial needed to generate samples from these processes. Section 3 describes the quasi-maximum likelihood procedure. Section 4 describes the simulation study and results. Section 5 concludes the paper.

*Corresponding author e-mail: taianeprass@gmail.com
†Authors' affiliation: Mathematics Institute - Federal University of Rio Grande do Sul - Porto Alegre, RS, Brazil.

2. SFIEGARCH Process

Let $d \in \mathbb{R}$ and $(1-\mathcal{B}^s)^{-d}$ be the operator defined by its Maclaurin series expansion, namely,

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

where $\Gamma(\cdot)$ is the Gamma function, $\mathcal{B}$ is the backward shift operator defined by $\mathcal{B}^{sk}(X_t) = X_{t-sk}$, for all $t \in \mathbb{Z}$ and $s, k \in \mathbb{N}$, $\pi_{d,sj+r} := 0$, for all $r \in \{1, \cdots, s-1\}$, and $\pi_{d,sj} := \delta_{-d,j} := \Gamma(j+d)\big[\Gamma(j+1)\Gamma(d)\big]^{-1}$, for all $j \in \mathbb{N}$.

Consider the order p and q polynomials, respectively, defined by

$$\alpha(z) := \sum_{i=0}^{p} (-\alpha_i) z^i \quad \text{and} \quad \beta(z) := \sum_{j=0}^{q} (-\beta_j) z^j, \quad \text{with} \quad \alpha_0 = \beta_0 = -1.$$

Assuming that $\alpha(\cdot)$ and $\beta(\cdot)$ have no common roots and that $\beta(z) \neq 0$ in the closed disk $\{z : |z| \leq 1\}$, define $\lambda(\cdot)$ by

$$\lambda(z) := \frac{\alpha(z)}{\beta(z)} (1-z^s)^{-d} = \sum_{k=0}^{\infty} \lambda_{d,k} z^k, \quad |z| < 1. \tag{2}$$

Definition 1. Let $\{Z_t\}_{t\in\mathbb{Z}}$ be a sequence of independent and identically distributed random variables (i.i.d.), with zero mean and variance equal to one, $\theta, \gamma \in \mathbb{R}$ and

$$g(Z_t) := \theta Z_t + \gamma\big[|Z_t| - \mathbb{E}\big(|Z_t|\big)\big], \quad \text{for all} \quad t \in \mathbb{Z}. \tag{3}$$

Moreover, for $\omega \in \mathbb{R}$, $d < 0.5$ and $s \in \mathbb{N}^*$, let $\{X_t\}_{t\in\mathbb{Z}}$ be the stochastic process defined by

$$X_t = \sigma_t Z_t, \tag{4}$$

$$\ln(\sigma_t^2) = \omega + \frac{\alpha(\mathcal{B})}{\beta(\mathcal{B})} (1-\mathcal{B}^s)^{-d} g(Z_{t-1}) = \omega + \lambda(\mathcal{B}) g(Z_{t-1}), \tag{5}$$

for all $t \in \mathbb{Z}$. Then, $\{X_t\}_{t\in\mathbb{Z}}$ is called a *seasonal* FIEGARCH process and it is denoted by SFIEGARCH$(p, d, q)_s$.

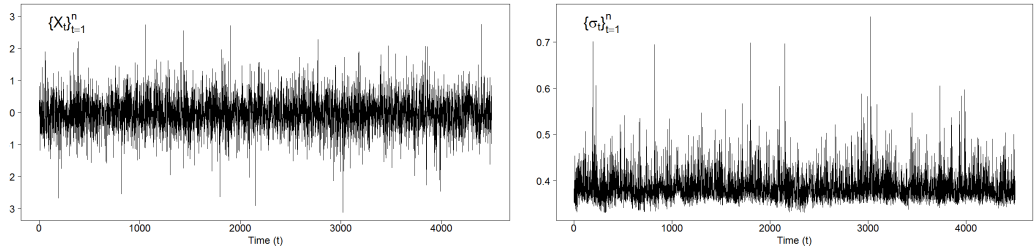


Figure 1: Simulated SFIEGARCH$(0, d, 0)_s$ time series $\{X_t\}_{t=1}^n$ (left) and its conditional standard deviation $\{\sigma_t\}_{t=1}^n$ (right), with $n = 4,500$ observations.

Example 1. Figure 1 shows an SFIEGARCH$(0, d, 0)_s$ time series $\{X_t\}_{t=1}^{4,500}$ (left) and its conditional standard deviation $\{\sigma_t\}_{t=1}^{4,500}$ (right), obtained from expression (5), with $\omega = -2.5$, $\theta = -0.05$, $\gamma = 0.14$, $d = 0.25$, $s = 6$ and $Z_0 \sim t_{6.5}$. For this example, polynomial (2) was truncated at $m = 100,000$.

Remark 1. Notice that,

- if $d = 0$, we have the so-called EGARCH(p, q) process (Nelson, 1991);
- if $s = 1$, we have the FIEGARCH(p, d, q) process (Bollerslev and Mikkelsen, 1996);
- an equivalent definition for SFIEGARCH process is given if one replaces expression (5) by

$$\beta(\mathcal{B})(1 - \mathcal{B}^s)^d(\ln(\sigma_t^2) - \omega) = \alpha(\mathcal{B})g(Z_{t-1}), \quad \text{for all} \quad t \in \mathbb{Z}. \tag{6}$$

Notice that expression (6) is similar (not necessarily equivalent) to the one presented by Bordignon et al. (2009) for a PLM-EGARCH(p, m, d, q, s) model.

The following proposition summarizes the main results regarding the stationarity and ergodicity of SFIEGARCH$(p, d, q)_s$ processes.

Proposition 1. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a SFIEGARCH(p, d, q) process and suppose that γ and θ, given in* (3)*, are not both equal to zero. Then, the following properties hold*

- $\ln(\sigma_t^2) - \omega = \sum_{k=0}^{\infty} \lambda_{d,k} g(Z_{t-1-k})$*, for all $t \in \mathbb{Z}$, is well defined and converges almost surely (a.s.) if and only if $d < 0.5$. If $d \leq 0$ the series converges absolutely a.s.*
- *If $d < 0.5$, $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is a causal* SARFIMA$(p, 0, q) \times (0, D, 0)_s$ *process, with $D = d$.*
- *If $d < 0.5$, the random variable X_t is finite with probability one, for all $t \in \mathbb{Z}$.*
- *If $d < 0.5$, the stochastic processes $\{\ln(\sigma_t^2) - \omega\}_{t\in\mathbb{Z}}$, $\{\sigma_t^2\}_{t\in\mathbb{Z}}$ and $\{X_t\}_{t\in\mathbb{Z}}$ are stationary (strictly and weakly) and ergodic.*
- *If $d < 0.5$ and $\mathbb{E}(|\ln(Z_0^2)|) < \infty$, then the process $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is well defined and it is strictly stationary and ergodic. Moreover, if $\mathbb{E}([\ln(Z_0^2)]^2) < \infty$ then it is also weakly stationary.*

Proof. See Lopes and Prass (2012). □

Proposition 2 gives a recurrence formula to calculate the coefficients of the polynomial $\lambda(\cdot)$. This expression is used to generate the SFIEGARCH time series in the simulation study presented in Section 4.

Proposition 2. *Let $\lambda(\cdot)$ be the polynomial defined by* (5)*. Suppose $\alpha(\cdot)$ and $\beta(\cdot)$ have no common roots and $\beta(z) \neq 0$ in the closed disk $\{z : |z| \leq 1\}$. Then, the coefficients $\lambda_{d,k}$, for all $k \in \mathbb{N}$, are given by*

$$\lambda_{d,0} = 1 \quad \textit{and} \quad \lambda_{d,k} = -\alpha_k^* + \sum_{i=0}^{k-1} \lambda_{d,i} \Big[\sum_{j=0}^{\min\{k-i,q\}} \delta^*_{d,\frac{k-i-j}{s}} \beta_j \Big], \quad \textit{for all } k > 0, \tag{7}$$

where $\alpha_k^* = \alpha_k$, if $k \leq p$, and 0 otherwise, and $\delta^*_{d,m} = \delta_{d,m}$, if $m \in \mathbb{N}$, and 0 otherwise, with $\delta_{d,m}$ defined in (1)*, for all $m \in \mathbb{N}$.*

Proof. See Lopes and Prass (2012). □

3. Quasi Maximum Likelihood Estimator (QMLE)

Let $\boldsymbol{\psi} \in \mathbb{R}^{\mathfrak{p}}$, with $\mathfrak{p} \leq \mathfrak{q} + p + q + 4$, be the vector of unknown parameters in expression (5) and observe that, from (4), the standardized residual of a SFIEGARCH process is given by

$$Z_t := \frac{X_t}{\sigma_t}, \quad \text{where} \quad \sigma_t := \sigma_t(\boldsymbol{\psi}) = \exp\left\{\frac{1}{2}\left[\omega + \sum_{k=0}^{\infty} \lambda_{d,k} g(Z_{t-1-k})\right]\right\}, \quad \text{for all } t \in \mathbb{Z}.$$

Denote by $f_Z(\cdot; \boldsymbol{\nu})$ the probability density function of Z_0, with nuisance parameters $\boldsymbol{\nu} \in \mathbb{R}^{\mathfrak{q}}$, where the value $\mathfrak{q}$ depends on the distribution considered. Thus, given an SFIEGARCH time series $\{x_t\}_{t=1}^n$, the maximum likelihood estimator (MLE) of $\boldsymbol{\psi} = (\boldsymbol{\eta}', \boldsymbol{\nu}')'$ is obtained by maximizing the log-likelihood function

$$\mathscr{L}_n(x_1, \cdots, x_n; \boldsymbol{\psi}) = \sum_{t=1}^{n} \ell_t(x_t; \boldsymbol{\psi}), \tag{8}$$

where $\ell_t(x_t; \boldsymbol{\psi}) = \ln\big(f_Z(z_t; \boldsymbol{\nu})\big) - 0.5\ln(\sigma_t^2(\boldsymbol{\psi}))$ is the log-likelihood function for the t-th observation, for all $t = 1, \cdots, n$.

Remark 2. Note that σ_t^2 is a $\mathcal{F}_t$-measurable function, where $\mathcal{F}_t$ is the information obtained until the instant t. Consequently, $\ell_t(\cdot; \boldsymbol{\psi})$ is the conditional distribution of $X_t|\mathcal{F}_{t-1}$, for all $t \in \mathbb{Z}$.

Obviously, the implementation of the maximum likelihood procedure requires the conditional density function $\ell_t(\cdot; \boldsymbol{\psi})$ to be known, which usually does not hold in practice. If the true distribution of Z_0 is unknown, an estimator for $\boldsymbol{\eta}$ is obtained by assuming a given distribution function for $X_t|\mathcal{F}_{t-1}$, for all $t \in \mathbb{Z}$, or equivalently, for Z_0. The estimator obtained under this condition is called pseudo-maximum likelihood estimator (PMLE) and it coincides with the MLE whenever the correct distribution function of $X_t|\mathcal{F}_{t-1}$ is specified, for all $t \in \mathbb{Z}$. Under the hypothesis that Z_0 has Gaussian distribution, it is called quasi-maximum likelihood estimator (QMLE) and the estimation procedure is called quasi-likelihood method (QLM).

Although the Student's t and the Generalized Error Distribution (GED) functions are also considered in the literature, the QLM is usually preferred. This is so because this method considers the standard Gaussian distribution and hence, the nuisance parameters do not need to be estimated. Moreover, under this assumption, $\sigma_t = \sigma_t(\boldsymbol{\eta})$, for all $t \in \mathbb{Z}$.

4. Simulation Study

We generate samples from SFIEGARCH$(0, d, 0)_s$ processes by considering two distributions for the random variable Z_0, namely, the t_{ν_1} and the GED(ν_2), with different values for the nuisance parameters ν_1 and ν_2. To estimate $\boldsymbol{\eta} = (d,\, \theta,\, \gamma)' \in \mathbb{R}^3$ we apply the quasi-maximum likelihood procedure as explained below.

4.1 Data Generating Process

To generate samples from SFIEGARCH processes we set the following:

1. the seasonal parameter $s \in \{2,\, 6\}$;
2. the differencing parameter $d \in \{0.10,\, 0.25,\, 0.35,\, 0.45\}$;

3. the nuisance parameter $\nu_1 \in \{2.5,\, 3.5,\, 5.0\}$, for the Student's t distribution, and $\nu_2 \in \{1.2,\, 2.5,\, 5.0\}$, for the GED distribution;

4. for all models, we fixed $\omega = 0$ since it is only a scaling parameter (we will assume that this parameter is known). We also considered $\theta = -0.25$ and $\gamma = 0.24$, which are values close to the ones observed in practical applications.

5. for each model, we consider the sample size $n \in \{2,000;\, 5,000\}$ with $re = 1,000$ replications;

6. when $s = 2$, the infinite sum (5) is truncated at $m = 50,000$ and, for $s = 6$, $m = 100,000$.

The sample $\{x_t\}_{t=1}^n$ is then obtained through the relation

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

and $\{z_t\}_{t=-m}^n$ is a sample of size $m + n + 1$ from the underlying distribution.

4.2 Parameter Estimation

To estimate the parameters of the model we assume that s and ω are known and hence the vector of unknown parameters is $\boldsymbol{\eta} = (d,\, \theta,\, \gamma)' \in \mathbb{R}^3$. From (8), the estimator $\hat{\boldsymbol{\eta}}$ of $\boldsymbol{\eta}$ is the value that maximizes

$$\mathscr{L}(x_1, \cdots, x_n; \boldsymbol{\eta}) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\left[\ln(\sigma_t^2) + \frac{x_t^2}{\sigma_t^2}\right]. \tag{9}$$

In order to solve (9), for each candidate $\boldsymbol{\eta} = (d,\, \theta,\, \gamma)' \in \mathbb{R}^3$, the coefficients $\{\lambda_{d,k}\}_{k=0}^n$ are obtained through the recurrence formula (7) and σ_t is obtained recursively by assuming that $g(z_t) = 0$, whenever $t < 1$ (initial conditions),

$$\sigma_1 = e^{0.5} \quad \text{and} \quad z_1 = \frac{x_1}{\sigma_1};$$

$$\sigma_t = \exp\left\{\frac{1}{2}\sum_{k=0}^{n-1}\lambda_{d,k} g(z_{t-1-k})\right\} \quad \text{and} \quad z_t = \frac{x_t}{\sigma_t}, \quad \text{for all } t = 2, \cdots, n.$$

4.3 Performance Measures

For any model, given $\boldsymbol{\eta} = (d, \theta, \gamma)' := (\eta_1, \eta_2, \eta_3)'$ we denote by $\hat{\eta}_i^{(k)}$ the estimate for η_i in the k-th replication, for $k \in \{1, \cdots, re\}$, $re = 1,000$ and $i \in \{1, 2, 3\}$. Moreover, given $\hat{\boldsymbol{\eta}}$, 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,

$$\bar{\eta}_i := \frac{1}{re}\sum_{k=1}^{re}\hat{\eta}_i^{(k)}, \quad sd := \sqrt{\frac{1}{re}\sum_{k=1}^{re}(\hat{\eta}_i^{(k)} - \bar{\eta}_i)^2}, \quad bias := \frac{1}{re}\sum_{k=1}^{re} e_i^{(k)},$$

$$mae := \frac{1}{re}\sum_{k=1}^{re}|e_i^{(k)}| \quad \text{and} \quad mse := \frac{1}{re}\sum_{k=1}^{re}\left(e_i^{(k)}\right)^2,$$

where $e_i^{(k)} := \hat{\eta}_i^{(k)} - \eta_i$, for $k \in \{1, \cdots, re\}$. The values of these statistics for this simulation study are reported in Tables 1-4.

4.4 Results

Tables 1 and 2 present the estimation results for $s = 2$, when the underlying distribution functions are, respectively, the Student's t and the GED. The estimation results for $s = 6$ are shown in Tables 3 and 4.

From Tables 1-4 one observes that,

- as expected, the estimation improves (in terms of sd, mae and mse) as the sample size increases;
- generally, the parameter θ seems to be better estimated than d and γ;
- for all parameters, the $bias$ is usually negative, in the GED case, and it varies in the Student's t case;
- for the Student's t distribution, as ν_1 increases, the $bias$ and the mae (mse) decreases. This result is expected since for $\nu_1 \leq 2$, Z_0 does not have finite variance and, as $\nu_1 \to \infty$, the t_{ν_1} distribution converges to the Gaussian one;
- in the GED case, the parameters are better estimated for $\nu_2 = 2.5$ (for $\nu_2 = 2$ we have the Gaussian distribution).

5. Conclusions

In this work we considered seasonal FIEGARCH (SFIEGARCH) processes which account for both, the long memory and seasonal behavior observed in financial time series. Properties of these processes where discussed and a Monte Carlo simulation study was conducted to access the finite sample performance of the quasi-maximum likelihood (QML) procedure.

The samples from the SFIEGARCH model were obtained by considering a recurrence formula to calculate the coefficients $\{\lambda_{d,k}\}_{k\in\mathbb{N}}$. Two distributions were considered for the random variable Z_0, namely, the t_{ν_1} and the GED(ν_2), with different values for the nuisance parameters ν_1 and ν_2, respectively. The values of ν_1 and ν_2 were selected so we obtained lighter and heavier tails than the Gaussian distribution.

We conclude that, given the complexity of SFIEGARCH models and the misspecification of the underlying distribution, the quasi-likelihood method performs relatively well, which is indicated by the small $bias$, mae and mse values for the estimates even when $s = 6$ and the sample size is $n = 2,000$.

Acknowledgments

T.S. Prass was supported by CNPq-Brazil. 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. The authors are also grateful to the (Brazilian) National Center of Super Computing (CESUP-UFRGS) for the computational resources.

Table 1: Estimation results for $s = 2$ and Student's t as the underlying distribution.

		$n = 2,000$					$n = 5,000$				
ν_1	η	$\bar{\eta}$	sd	$bias$	mae	mse	$\bar{\eta}$	sd	$bias$	mae	mse
	$d = 0.10$	0.1868	0.2743	0.0868	0.2301	0.0828	0.1467	0.2235	0.0467	0.1698	0.0521
2.5	$\theta = -0.25$	-0.2073	0.1720	0.0427	0.1318	0.0314	-0.2223	0.1404	0.0277	0.0987	0.0205
	$\gamma = 0.24$	0.2446	0.1801	0.0046	0.1356	0.0325	0.2274	0.1479	-0.0126	0.1035	0.0220
	$d = 0.10$	0.0677	0.1986	-0.0323	0.1503	0.0405	0.0590	0.1221	-0.0410	0.0948	0.0166
3.5	$\theta = -0.25$	-0.2451	0.0849	0.0049	0.0633	0.0072	-0.2462	0.0573	0.0038	0.0431	0.0033
	$\gamma = 0.24$	0.2218	0.1128	-0.0182	0.0877	0.0131	0.2240	0.0796	-0.0160	0.0602	0.0066
	$d = 0.10$	0.0740	0.1344	-0.0260	0.1021	0.0187	0.0748	0.0784	-0.0252	0.0642	0.0068
5.0	$\theta = -0.25$	-0.2507	0.0555	-0.0007	0.0425	0.0031	-0.2501	0.0361	-0.0001	0.0281	0.0013
	$\gamma = 0.24$	0.2279	0.0801	-0.0121	0.0639	0.0066	0.2326	0.0511	-0.0074	0.0410	0.0027
	$d = 0.25$	0.2497	0.2355	-0.0003	0.1673	0.0555	0.2108	0.1795	-0.0392	0.1220	0.0337
2.5	$\theta = -0.25$	-0.2122	0.1650	0.0378	0.1211	0.0287	-0.2254	0.1381	0.0246	0.0917	0.0197
	$\gamma = 0.24$	0.2288	0.1760	-0.0112	0.1346	0.0311	0.2091	0.1516	-0.0309	0.1075	0.0239
	$d = 0.25$	0.2005	0.1592	-0.0495	0.1189	0.0278	0.1831	0.0987	-0.0669	0.0893	0.0142
3.5	$\theta = -0.25$	-0.2462	0.0825	0.0038	0.0608	0.0068	-0.2488	0.0548	0.0012	0.0408	0.0030
	$\gamma = 0.24$	0.2122	0.1149	-0.0278	0.0908	0.0140	0.2101	0.0797	-0.0299	0.0663	0.0073
	$d = 0.25$	0.2233	0.1031	-0.0267	0.0772	0.0113	0.2157	0.0592	-0.0343	0.0538	0.0047
5.0	$\theta = -0.25$	-0.2495	0.0540	0.0005	0.0418	0.0029	-0.2507	0.0344	-0.0007	0.0267	0.0012
	$\gamma = 0.24$	0.2248	0.0795	-0.0152	0.0639	0.0065	0.2241	0.0498	-0.0159	0.0412	0.0027
	$d = 0.35$	0.2899	0.2197	-0.0601	0.1576	0.0519	0.2542	0.1713	-0.0958	0.1377	0.0385
2.5	$\theta = -0.25$	-0.2238	0.1572	0.0262	0.1111	0.0254	-0.2391	0.1246	0.0109	0.0820	0.0157
	$\gamma = 0.24$	0.2076	0.1804	-0.0324	0.1409	0.0336	0.1878	0.1448	-0.0522	0.1167	0.0237
	$d = 0.35$	0.2874	0.1407	-0.0626	0.1147	0.0237	0.2663	0.1007	-0.0837	0.1036	0.0171
3.5	$\theta = -0.25$	-0.2522	0.0799	-0.0022	0.0584	0.0064	-0.2550	0.0603	-0.0050	0.0406	0.0037
	$\gamma = 0.24$	0.1925	0.1214	-0.0475	0.1013	0.0170	0.1775	0.0974	-0.0625	0.0876	0.0134
	$d = 0.35$	0.3105	0.0994	-0.0395	0.0821	0.0114	0.2974	0.0701	-0.0526	0.0723	0.0077
5.0	$\theta = -0.25$	-0.2530	0.0545	-0.0030	0.0418	0.0030	-0.2563	0.0357	-0.0063	0.0279	0.0013
	$\gamma = 0.24$	0.2031	0.0878	-0.0369	0.0756	0.0091	0.1936	0.0629	-0.0464	0.0624	0.0061
	$d = 0.45$	0.3492	0.2373	-0.1008	0.1828	0.0665	0.3311	0.2038	-0.1189	0.1677	0.0557
2.5	$\theta = -0.25$	-0.2428	0.1545	0.0072	0.1062	0.0239	-0.2510	0.1242	-0.0010	0.0799	0.0154
	$\gamma = 0.24$	0.1598	0.1755	-0.0802	0.1615	0.0372	0.1267	0.1471	-0.1133	0.1578	0.0345
	$d = 0.45$	0.3848	0.1719	-0.0652	0.1357	0.0338	0.3776	0.1411	-0.0724	0.1170	0.0252
3.5	$\theta = -0.25$	-0.2555	0.0846	-0.0055	0.0631	0.0072	-0.2569	0.0605	-0.0069	0.0457	0.0037
	$\gamma = 0.24$	0.1240	0.1304	-0.1160	0.1475	0.0305	0.0945	0.1059	-0.1455	0.1575	0.0324
	$d = 0.45$	0.4121	0.1360	-0.0379	0.1074	0.0199	0.4108	0.1044	-0.0392	0.0841	0.0124
5.0	$\theta = -0.25$	-0.2506	0.0600	-0.0006	0.0465	0.0036	-0.2540	0.0416	-0.0040	0.0327	0.0017
	$\gamma = 0.24$	0.1220	0.1154	-0.1180	0.1402	0.0273	0.0937	0.0910	-0.1463	0.1516	0.0297

Table 2: Estimation results for $s = 2$ and GED as the underlying distribution.

		$n = 2,000$					$n = 5,000$				
ν_2	η	$\bar{\eta}$	sd	$bias$	mae	mse	$\bar{\eta}$	sd	$bias$	mae	mse
	$d = 0.10$	0.0818	0.1006	-0.0182	0.0778	0.0105	0.0828	0.0601	-0.0172	0.0482	0.0039
1.2	$\theta = -0.25$	-0.2493	0.0437	0.0007	0.0351	0.0019	-0.2481	0.0268	0.0019	0.0213	0.0007
	$\gamma = 0.24$	0.2355	0.0603	-0.0045	0.0480	0.0037	0.2340	0.0389	-0.0060	0.0319	0.0015
	$d = 0.10$	0.0942	0.0706	-0.0058	0.0556	0.0050	0.0969	0.0416	-0.0031	0.0331	0.0017
2.5	$\theta = -0.25$	-0.2515	0.0281	-0.0015	0.0226	0.0008	-0.2508	0.0177	-0.0008	0.0141	0.0003
	$\gamma = 0.24$	0.2406	0.0496	0.0006	0.0395	0.0025	0.2419	0.0304	0.0019	0.0242	0.0009
	$d = 0.10$	0.0861	0.0536	-0.0139	0.0431	0.0031	0.0865	0.0323	-0.0135	0.0276	0.0012
5.0	$\theta = -0.25$	-0.2531	0.0231	-0.0031	0.0185	0.0005	-0.2527	0.0145	-0.0027	0.0119	0.0002
	$\gamma = 0.24$	0.2396	0.0425	-0.0004	0.0336	0.0018	0.2388	0.0280	-0.0012	0.0224	0.0008
	$d = 0.25$	0.2270	0.0787	-0.0230	0.0613	0.0067	0.2217	0.0470	-0.0283	0.0433	0.0030
1.2	$\theta = -0.25$	-0.2486	0.0431	0.0014	0.0345	0.0019	-0.2491	0.0257	0.0009	0.0206	0.0007
	$\gamma = 0.24$	0.2299	0.0587	-0.0101	0.0475	0.0035	0.2248	0.0380	-0.0152	0.0329	0.0017
	$d = 0.25$	0.2465	0.0544	-0.0035	0.0405	0.0030	0.2432	0.0317	-0.0068	0.0256	0.0011
2.5	$\theta = -0.25$	-0.2525	0.0286	-0.0025	0.0228	0.0008	-0.2533	0.0175	-0.0033	0.0144	0.0003
	$\gamma = 0.24$	0.2407	0.0491	0.0007	0.0392	0.0024	0.2393	0.0302	-0.0007	0.0238	0.0009
	$d = 0.25$	0.2304	0.0430	-0.0196	0.0372	0.0022	0.2272	0.0273	-0.0228	0.0292	0.0013
5.0	$\theta = -0.25$	-0.2582	0.0227	-0.0082	0.0192	0.0006	-0.2595	0.0142	-0.0095	0.0138	0.0003
	$\gamma = 0.24$	0.2324	0.0425	-0.0076	0.0340	0.0019	0.2293	0.0288	-0.0107	0.0245	0.0009
	$d = 0.35$	0.3137	0.0868	-0.0363	0.0741	0.0088	0.3007	0.0663	-0.0493	0.0681	0.0068
1.2	$\theta = -0.25$	-0.2509	0.0446	-0.0009	0.0351	0.0020	-0.2559	0.0281	-0.0059	0.0229	0.0008
	$\gamma = 0.24$	0.2068	0.0731	-0.0332	0.0634	0.0064	0.1912	0.0547	-0.0488	0.0579	0.0054
	$d = 0.35$	0.3326	0.0740	-0.0174	0.0614	0.0058	0.3205	0.0619	-0.0295	0.0572	0.0047
2.5	$\theta = -0.25$	-0.2567	0.0313	-0.0067	0.0254	0.0010	-0.2615	0.0205	-0.0115	0.0188	0.0006
	$\gamma = 0.24$	0.2099	0.0725	-0.0301	0.0626	0.0062	0.1916	0.0617	-0.0484	0.0614	0.0061
	$d = 0.35$	0.3237	0.0651	-0.0263	0.0569	0.0049	0.3178	0.0534	-0.0322	0.0512	0.0039
5.0	$\theta = -0.25$	-0.2629	0.0248	-0.0129	0.0223	0.0008	-0.2662	0.0166	-0.0162	0.0194	0.0005
	$\gamma = 0.24$	0.1918	0.0707	-0.0482	0.0674	0.0073	0.1749	0.0642	-0.0651	0.0740	0.0084
	$d = 0.45$	0.4147	0.1352	-0.0353	0.1050	0.0195	0.4222	0.0992	-0.0278	0.0787	0.0106
1.2	$\theta = -0.25$	-0.2470	0.0509	0.0030	0.0396	0.0026	-0.2505	0.0365	-0.0005	0.0284	0.0013
	$\gamma = 0.24$	0.1251	0.1096	-0.1149	0.1327	0.0252	0.0875	0.0866	-0.1525	0.1558	0.0307
	$d = 0.45$	0.4365	0.1064	-0.0135	0.0832	0.0115	0.4367	0.0797	-0.0133	0.0646	0.0065
2.5	$\theta = -0.25$	-0.2533	0.0369	-0.0033	0.0289	0.0014	-0.2553	0.0292	-0.0053	0.0234	0.0009
	$\gamma = 0.24$	0.1114	0.1074	-0.1286	0.1440	0.0281	0.0778	0.0823	-0.1622	0.1657	0.0331
	$d = 0.45$	0.4440	0.0923	-0.0060	0.0739	0.0086	0.4419	0.0681	-0.0081	0.0543	0.0047
5.0	$\theta = -0.25$	-0.2540	0.0355	-0.0040	0.0278	0.0013	-0.2551	0.0269	-0.0051	0.0219	0.0007
	$\gamma = 0.24$	0.0952	0.1046	-0.1448	0.1567	0.0319	0.0688	0.0806	-0.1712	0.1732	0.0358

Table 3: Estimation results for $s = 6$ and Student's t as the underlying distribution.

		$n = 2,000$					$n = 5,000$				
ν_1	η	$\bar{\eta}$	sd	$bias$	mae	mse	$\bar{\eta}$	sd	$bias$	mae	mse
	$d = 0.10$	0.2098	0.2774	0.1098	0.2404	0.0890	0.1726	0.2097	0.0726	0.1738	0.0493
2.5	$\theta = -0.25$	-0.2053	0.1711	0.0447	0.1329	0.0313	-0.2177	0.1248	0.0323	0.0949	0.0166
	$\gamma = 0.24$	0.2545	0.1809	0.0145	0.1331	0.0329	0.2310	0.1361	-0.0090	0.0996	0.0186
	$d = 0.10$	0.0690	0.2022	-0.0310	0.1510	0.0419	0.0672	0.1337	-0.0328	0.1018	0.0190
3.5	$\theta = -0.25$	-0.2462	0.0886	0.0038	0.0662	0.0079	-0.2465	0.0591	0.0035	0.0445	0.0035
	$\gamma = 0.24$	0.2259	0.1104	-0.0141	0.0849	0.0124	0.2247	0.0728	-0.0153	0.0574	0.0055
	$d = 0.10$	0.0767	0.1314	-0.0233	0.0990	0.0178	0.0840	0.0799	-0.0160	0.0628	0.0066
5.0	$\theta = -0.25$	-0.2456	0.0540	0.0044	0.0424	0.0029	-0.2474	0.0346	0.0026	0.0276	0.0012
	$\gamma = 0.24$	0.2316	0.0778	-0.0084	0.0615	0.0061	0.2298	0.0501	-0.0102	0.0410	0.0026
	$d = 0.25$	0.2800	0.2279	0.0300	0.1715	0.0528	0.2368	0.1780	-0.0132	0.1222	0.0319
2.5	$\theta = -0.25$	-0.2127	0.1663	0.0373	0.1248	0.0291	-0.2255	0.1150	0.0245	0.0863	0.0138
	$\gamma = 0.24$	0.2390	0.1772	-0.0010	0.1299	0.0314	0.2144	0.1284	-0.0256	0.1003	0.0171
	$d = 0.25$	0.2080	0.1650	-0.0420	0.1216	0.0290	0.2005	0.1031	-0.0495	0.0842	0.0131
3.5	$\theta = -0.25$	-0.2468	0.0848	0.0032	0.0640	0.0072	-0.2472	0.0592	0.0028	0.0433	0.0035
	$\gamma = 0.24$	0.2157	0.1077	-0.0243	0.0857	0.0122	0.2105	0.0755	-0.0295	0.0613	0.0066
	$d = 0.25$	0.2288	0.0990	-0.0212	0.0752	0.0102	0.2278	0.0576	-0.0222	0.0474	0.0038
5.0	$\theta = -0.25$	-0.2456	0.0524	0.0044	0.0410	0.0028	-0.2479	0.0331	0.0021	0.0263	0.0011
	$\gamma = 0.24$	0.2267	0.0745	-0.0133	0.0602	0.0057	0.2226	0.0488	-0.0174	0.0419	0.0027
	$d = 0.35$	0.3270	0.2099	-0.0230	0.1477	0.0446	0.2884	0.1646	-0.0616	0.1213	0.0309
2.5	$\theta = -0.25$	-0.2216	0.1604	0.0284	0.1157	0.0265	-0.2339	0.1067	0.0161	0.0783	0.0117
	$\gamma = 0.24$	0.2221	0.1763	-0.0179	0.1319	0.0314	0.1963	0.1270	-0.0437	0.1073	0.0181
	$d = 0.35$	0.3042	0.1447	-0.0458	0.1082	0.0230	0.2925	0.0949	-0.0575	0.0837	0.0123
3.5	$\theta = -0.25$	-0.2488	0.0810	0.0012	0.0607	0.0066	-0.2517	0.0539	-0.0017	0.0405	0.0029
	$\gamma = 0.24$	0.1997	0.1054	-0.0403	0.0896	0.0127	0.1903	0.0734	-0.0497	0.0711	0.0079
	$d = 0.35$	0.3247	0.0839	-0.0253	0.0664	0.0077	0.3173	0.0528	-0.0327	0.0500	0.0039
5.0	$\theta = -0.25$	-0.2479	0.0510	0.0021	0.0395	0.0026	-0.2517	0.0324	-0.0017	0.0254	0.0011
	$\gamma = 0.24$	0.2141	0.0734	-0.0259	0.0627	0.0061	0.2064	0.0499	-0.0336	0.0486	0.0036
	$d = 0.45$	0.3861	0.2033	-0.0639	0.1512	0.0454	0.3596	0.1705	-0.0904	0.1405	0.0372
2.5	$\theta = -0.25$	-0.2376	0.1594	0.0124	0.1079	0.0256	-0.2484	0.1034	0.0016	0.0719	0.0107
	$\gamma = 0.24$	0.1775	0.1739	-0.0625	0.1425	0.0342	0.1512	0.1275	-0.0888	0.1306	0.0241
	$d = 0.45$	0.3926	0.1492	-0.0574	0.1175	0.0256	0.3926	0.1112	-0.0574	0.0945	0.0157
3.5	$\theta = -0.25$	-0.2524	0.0789	-0.0024	0.0585	0.0062	-0.2546	0.0565	-0.0046	0.0428	0.0032
	$\gamma = 0.24$	0.1540	0.1149	-0.0860	0.1186	0.0206	0.1303	0.0943	-0.1097	0.1224	0.0209
	$d = 0.45$	0.4147	0.1062	-0.0353	0.0867	0.0125	0.4094	0.0821	-0.0406	0.0730	0.0084
5.0	$\theta = -0.25$	-0.2521	0.0533	-0.0021	0.0414	0.0028	-0.2555	0.0372	-0.0055	0.0297	0.0014
	$\gamma = 0.24$	0.1602	0.0993	-0.0798	0.1046	0.0162	0.1350	0.0867	-0.1050	0.1136	0.0185

Table 4: Estimation results for $s = 6$ and GED as the underlying distribution.

		$n = 2,000$					$n = 5,000$				
ν_2	η	$\bar{\eta}$	sd	$bias$	mae	mse	$\bar{\eta}$	sd	$bias$	mae	mse
	$d = 0.10$	0.0887	0.1065	-0.0113	0.0826	0.0115	0.0878	0.0602	-0.0122	0.0486	0.0038
1.2	$\theta = -0.25$	-0.2475	0.0425	0.0025	0.0342	0.0018	-0.2477	0.0271	0.0023	0.0220	0.0007
	$\gamma = 0.24$	0.2321	0.0617	-0.0079	0.0500	0.0039	0.2340	0.0373	-0.0060	0.0304	0.0014
	$d = 0.10$	0.0993	0.0670	-0.0007	0.0517	0.0045	0.0983	0.0421	-0.0017	0.0328	0.0018
2.5	$\theta = -0.25$	-0.2510	0.0288	-0.0010	0.0230	0.0008	-0.2509	0.0184	-0.0009	0.0147	0.0003
	$\gamma = 0.24$	0.2376	0.0485	-0.0024	0.0389	0.0024	0.2393	0.0310	-0.0007	0.0248	0.0010
	$d = 0.10$	0.0875	0.0523	-0.0125	0.0430	0.0029	0.0873	0.0329	-0.0127	0.0278	0.0012
5.0	$\theta = -0.25$	-0.2523	0.0235	-0.0023	0.0185	0.0006	-0.2525	0.0144	-0.0025	0.0115	0.0002
	$\gamma = 0.24$	0.2374	0.0450	-0.0026	0.0360	0.0020	0.2378	0.0279	-0.0022	0.0227	0.0008
	$d = 0.25$	0.2350	0.0785	-0.0150	0.0614	0.0064	0.2306	0.0438	-0.0194	0.0375	0.0023
1.2	$\theta = -0.25$	-0.2474	0.0410	0.0026	0.0330	0.0017	-0.2480	0.0263	0.0020	0.0213	0.0007
	$\gamma = 0.24$	0.2274	0.0592	-0.0126	0.0488	0.0037	0.2270	0.0360	-0.0130	0.0310	0.0015
	$d = 0.25$	0.2484	0.0497	-0.0016	0.0390	0.0025	0.2461	0.0304	-0.0039	0.0241	0.0009
2.5	$\theta = -0.25$	-0.2522	0.0283	-0.0022	0.0228	0.0008	-0.2524	0.0180	-0.0024	0.0145	0.0003
	$\gamma = 0.24$	0.2370	0.0466	-0.0030	0.0371	0.0022	0.2382	0.0298	-0.0018	0.0239	0.0009
	$d = 0.25$	0.2353	0.0417	-0.0147	0.0349	0.0020	0.2320	0.0257	-0.0180	0.0250	0.0010
5.0	$\theta = -0.25$	-0.2557	0.0229	-0.0057	0.0186	0.0006	-0.2573	0.0140	-0.0073	0.0127	0.0003
	$\gamma = 0.24$	0.2330	0.0440	-0.0070	0.0356	0.0020	0.2317	0.0277	-0.0083	0.0234	0.0008
	$d = 0.35$	0.3274	0.0705	-0.0226	0.0573	0.0055	0.3188	0.0444	-0.0312	0.0439	0.0029
1.2	$\theta = -0.25$	-0.2496	0.0399	0.0004	0.0318	0.0016	-0.2525	0.0263	-0.0025	0.0211	0.0007
	$\gamma = 0.24$	0.2154	0.0595	-0.0246	0.0515	0.0041	0.2089	0.0404	-0.0311	0.0409	0.0026
	$d = 0.35$	0.3389	0.0522	-0.0111	0.0425	0.0028	0.3308	0.0401	-0.0192	0.0360	0.0020
2.5	$\theta = -0.25$	-0.2559	0.0289	-0.0059	0.0238	0.0009	-0.2584	0.0185	-0.0084	0.0163	0.0004
	$\gamma = 0.24$	0.2215	0.0521	-0.0185	0.0441	0.0031	0.2149	0.0431	-0.0251	0.0386	0.0025
	$d = 0.35$	0.3300	0.0487	-0.0200	0.0416	0.0028	0.3239	0.0364	-0.0261	0.0359	0.0020
5.0	$\theta = -0.25$	-0.2600	0.0229	-0.0100	0.0199	0.0006	-0.2641	0.0144	-0.0141	0.0167	0.0004
	$\gamma = 0.24$	0.2133	0.0524	-0.0267	0.0461	0.0035	0.2034	0.0436	-0.0366	0.0449	0.0032
	$d = 0.45$	0.4150	0.0961	-0.0350	0.0796	0.0105	0.4130	0.0765	-0.0370	0.0686	0.0072
1.2	$\theta = -0.25$	-0.2519	0.0449	-0.0019	0.0357	0.0020	-0.2553	0.0337	-0.0053	0.0268	0.0012
	$\gamma = 0.24$	0.1607	0.0949	-0.0793	0.1011	0.0153	0.1345	0.0859	-0.1055	0.1140	0.0185
	$d = 0.45$	0.4289	0.0789	-0.0211	0.0647	0.0067	0.4270	0.0639	-0.0230	0.0540	0.0046
2.5	$\theta = -0.25$	-0.2577	0.0334	-0.0077	0.0268	0.0012	-0.2593	0.0243	-0.0093	0.0209	0.0007
	$\gamma = 0.24$	0.1565	0.0995	-0.0835	0.1081	0.0169	0.1295	0.0929	-0.1105	0.1251	0.0209
	$d = 0.45$	0.4339	0.0770	-0.0161	0.0621	0.0062	0.4303	0.0620	-0.0197	0.0519	0.0042
5.0	$\theta = -0.25$	-0.2574	0.0287	-0.0074	0.0237	0.0009	-0.2608	0.0218	-0.0108	0.0196	0.0006
	$\gamma = 0.24$	0.1457	0.1059	-0.0943	0.1195	0.0201	0.1203	0.0964	-0.1197	0.1340	0.0236

REFERENCES

Berkes, I, Horváth, L. and Kokoszka, P. (2003). "GARCH Processes: Structure and Estimation". *Bernoulli*, 9, 201–228.

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

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

Bordignon, S., Caporin, M. and Lisi, F. (2009). "Periodic Long-Memory GARCH models." *Econometric Reviews*, 28(1), 60–82.

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

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

Lopes, S.R.C. and Prass T.S. (2012). "Seasonal FIEGARCH Processes". Working Paper.

Lumsdaine, R. (1996). "Asymptotic Properties of the Maximum Likelihood Estimator in GARCH$(1,1)$ and IGARCH$(1,1)$ Models". *Econometrica*, 64, 575–596.

Nelson, D.B. (1991). "Conditional Heteroskedasticity in Asset Returns: A New Approach". *Econometrica*, 59, 347–370.

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

Apêndice H

Artigo Prass, Lopes e Achcar (2013)

Autores: Taiane S. Prass, Sílvia R. C. Lopes e Jorge A. Achcar

Título: MCMC Bayesian Estimation in FIEGARCH Models

Revista: Em Revisão

Ano: 2013

MCMC Bayesian Estimation in FIEGARCH Models

Taiane S. Prass[(a)]*, Sílvia R.C. Lopes[(a)] and Jorge A. Achcar[(b)]

(a) Mathematics Institute - UFRGS, Porto Alegre, RS, Brazil
(b) Medical School - USP, Ribeirão Preto, SP, Brazil

July 2, 2013

Abstract

Bayesian inference for fractionally integrated exponential generalized autoregressive conditional heteroskedastic (FIEGARCH) models using Markov Chain Monte Carlo (MCMC) methods is described. A simulation study is presented to access the performance of the procedure, under the presence of long-memory in the volatility. Samples from FIEGARCH processes are obtained upon considering the generalized error distribution (GED) for the innovation process. Different values for the tail-thickness parameter ν are considered covering both scenarios, innovation processes with lighter ($\nu < 2$) and heavier ($\nu > 2$) tails than the Gaussian distribution ($\nu = 2$). A sensitivity analysis is performed by considering different prior density functions and by integrating (or not) the knowledge on the true parameter values to select the hyperparameter values.

Key words: Bayesian inference, MCMC, FIEGARCH processes, Long-range dependence.

1 Introduction

ARCH-type (Autoregressive Conditional Heteroskedasticity) and stochastic volatility (Breidt et al., 1998) models are commonly used in financial time series modeling to represent the dynamic evolution of volatilities. By ARCH-type models we mean not only the ARCH model proposed by Engle (1982) but also several generalizations that were lately proposed.

Among the most popular generalizations of the ARCH model is the generalized ARCH (GARCH) model, introduced by Bollerslev (1986), for which the conditional variance depends not only on the p past values of the process (as in the ARCH model), but also on the q past values of the conditional variance. Although the ARCH and GARCH models are widely used in practice, they do not take into account the asymmetry in the volatility, that is, the fact that volatility tends to rise in response to "bad" news and to fall in response to "good" news. As an alternative, Nelson (1991) introduces the exponential GARCH (EGARCH) model. This model not only describes the asymmetry on the volatility, but also have the advantage that the positivity of the conditional variance is always attained since it is defined in terms of the logarithm function.

*Corresponding author: taianeprass@gmail.com

The fractionally integrated EGARCH (FIEGARCH) and fractionally integrated GARCH (FIGARCH) models proposed, respectively, by Bollerslev and Mikkelsen (1996) and Baillie et al. (1996), generalize the EGARCH (Nelson, 1991) and the GARCH (Bollerslev, 1986) models, respectively. FIEGARCH models have not only the capability of modeling clusters of volatility (as ARCH and GARCH models do) and capturing its asymmetry (as the EGARCH model does) but they also take into account the characteristic of long memory in the volatility (as the FIGARCH model does). The non-stationarity of FIGARCH models (in the weak sense) makes this class of models less attractive for practical applications. Another drawback of the FIGARCH models is that we must have $d \geq 0$ and the polynomial coefficients in its definition must satisfy some restrictions so the conditional variance will be positive. FIEGARCH models do not have this problem since the variance is defined in terms of the logarithm function, moreover, they are weak stationary whenever the long memory parameter d is smaller than 0.5 (Lopes and Prass, 2013).

A complete study on the theoretical properties of FIEGARCH processes is presented in Lopes and Prass (2013). The authors also conduct a simulation study to analyze the finite sample performance of the quasi-maximum likelihood (QML) procedure on parameter estimation. The QML procedure became popular for two main reasons. First, the expression for the quasi-likelihood function is simpler for the Gaussian case than when considering, for example, the Student's t or the generalized error distribution (GED). Second, since the parameters of the distribution function are not estimated, the dimension of the optimization problem is reduced. On the other hand, the results in Lopes and Prass (2013) indicate that, although the QML presents a relatively good performance when the sample size is 2000 and the estimation improves as the sample size increases, it does so very slowly.

In this work we propose the use of Bayesian methods using Monte Carlo simulation techniques on the estimation of the FIEGARCH model parameters. This procedure is usually considered to analyze financial time series assuming stochastic volatility models (see, for example, Meyer and Yu, 2000), mostly because of the difficulty on applying traditional statistical techniques due to the complexity of the likelihood function. To generate samples from the joint posterior distribution for the parameters of interest we use MCMC (Markov Chain Monte Carlo) methods as the Gibbs Sampling algorithm (see, for example, Gelfand and Smith, 1990; Casela and George, 1992) or the Metropolis-Hastings algorithm (see, for example, Smith and Roberts, 1993; Chib and Greenberg, 1995). These samples are generated from all conditional posterior distributions for each parameter given all the other parameters and the data.

A simulation study is conducted to access the finite sample performance of the procedure proposed here, under the presence of long-memory in the volatility. The samples from FIEGARCH processes are obtained upon considering the GED for the innovation process. Taking into account that financial time series are usually characterized by heavy tailed distributions, different values for the tail-thickness parameter ν are considered covering both scenarios: innovation processes with lighter and heavier tails than the Gaussian distribution. A sensitivity analysis is performed by considering different prior density functions and by integrating (or not) the knowledge on the true parameter values to select the hyperparameter values.

The paper is organized as follows. In Section 2 a review on the definition and main properties of FIEGARCH processes is presented. Section 3 describes the parameter estimation procedure when Bayesian inference using MCMC is considered. Section 4 describes the steps used in the simulation study, such as the data generating process, the prior selection procedure and the performance measures considered. This section also reports the simulation results. Section 5 concludes the paper.

2 FIEGARCH Processes

Let $(1-\mathcal{B})^{-d}$ be the operator defined by its Maclaurin series expansion, namely,

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

where $\tau_{d,k} := \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)}$, for all $k \geq 1$, $\Gamma(\cdot)$ is the gamma function and $\mathcal{B}$ is the backward shift operator defined by $\mathcal{B}^k(X_t) = X_{t-k}$, for all $k \in \mathbb{N}$.

Assume that $\alpha(\cdot)$ and $\beta(\cdot)$ are polynomials of order p and q, respectively, defined by

$$\alpha(z) = \sum_{i=0}^{p} (-\alpha_i) z^i \quad \text{and} \quad \beta(z) = \sum_{j=0}^{q} (-\beta_j) z^j, \tag{2}$$

with $\alpha_0 = \beta_0 = -1$. If $\alpha(\cdot)$ and $\beta(\cdot)$ have no common roots and $\beta(z) \neq 0$ in the closed disk $\{z : |z| \leq 1\}$, then the function $\lambda(\cdot)$, defined by

$$\lambda(z) = \frac{\alpha(z)}{\beta(z)}(1-z)^{-d} := \sum_{k=0}^{\infty} \lambda_{d,k} z^k, \quad \text{for all } |z| < 1, \tag{3}$$

is analytic in the open disc $\{z : |z| < 1\}$, for any $d > 0$, and in the closed disk $\{z : |z| \leq 1\}$, whenever $d \leq 0$. Therefore, $\lambda(\cdot)$ is well defined and the power series representation in (3) is unique. More specifically, the coefficients $\lambda_{d,k}$, for all $k \in \mathbb{N}$, are given by (see Lopes and Prass, 2013)

$$\lambda_{d,0} = 1 \quad \text{and} \quad \lambda_{d,k} = -\alpha_k^* + \sum_{i=0}^{k-1} \lambda_i \sum_{j=0}^{k-i} \beta_j^* \delta_{d,k-i-j}, \quad \text{for all} \quad k \geq 1, \tag{4}$$

where

$$\alpha_m^* := \begin{cases} \alpha_m, & \text{if} \quad 0 \leq m \leq p; \\ 0, & \text{if} \quad m > p; \end{cases} \qquad \beta_m^* := \begin{cases} \beta_m, & \text{if} \quad 0 \leq m \leq q; \\ 0, & \text{if} \quad m > q; \end{cases} \tag{5}$$

and $\delta_{d,j} := \tau_{-d,j}$, for all $j \in \mathbb{N}$, are the coefficients obtained upon replacing $-d$ by d in (1), that is

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

Let $\theta, \gamma \in \mathbb{R}$ and $\{Z_t\}_{t\in\mathbb{Z}}$ be a sequence of independent and identically distributed (i.i.d.) random variables, with zero mean and variance equal to one. Assume that θ and γ are not both equal to zero and define $\{g(Z_t)\}_{t\in\mathbb{Z}}$ by

$$g(Z_t) = \theta Z_t + \gamma[|Z_t| - \mathbb{E}(|Z_t|)], \quad \text{for all } t \in \mathbb{Z}. \tag{6}$$

It follows that (see Lopes and Prass, 2013) $\{g(Z_t)\}_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic process. Moreover, since $\mathbb{E}(Z_0^2) < \infty$, then $\{g(Z_t)\}_{t\in\mathbb{Z}}$ is also weakly stationary with mean zero (therefore a white noise process) and variance σ_g^2 given by

$$\sigma_g^2 = \theta^2 + \gamma^2 - [\gamma \mathbb{E}(|Z_0|)]^2 + 2\,\theta\,\gamma\,\mathbb{E}(Z_0|Z_0|). \tag{7}$$

Now, for any $d < 0.5$ and $\omega \in \mathbb{R}$, let $\{X_t\}_{t\in\mathbb{Z}}$ be the stochastic process defined by

$$X_t = \sigma_t Z_t, \tag{8}$$

$$\begin{aligned}\ln(\sigma_t^2) &= \omega + \frac{\alpha(\mathcal{B})}{\beta(\mathcal{B})}(1-\mathcal{B})^{-d} g(Z_{t-1}) \\ &= \omega + \sum_{k=0}^{\infty} \lambda_{d,k} g(Z_{t-1-k}), \qquad \text{for all } t \in \mathbb{Z}.\end{aligned} \tag{9}$$

Then $\{X_t\}_{t\in\mathbb{Z}}$ is a *Fractionally Integrated* EGARCH *process*, denoted by FIEGARCH(p, d, q) (Bollerslev and Mikkelsen, 1996).

The properties of FIEGARCH(p, d, q) processes, with $d < 0.5$, are given below (the proofs of these properties can be found in Lopes and Prass, 2013). Henceforth $GED(\nu)$ denotes the generalized error distribution with tail thickness parameter ν.

Proposition 1. *Let $\{X_t\}_{t\in\mathbb{Z}}$ FIEGARCH(p, d, a) process. Then the following properties hold:*

1. *$\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ is a stationary (weakly and strictly) and an ergodic process and the random variable $\ln(\sigma_t^2)$ is almost surely finite, for all $t \in \mathbb{Z}$;*
2. *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;*
3. *$\{X_t\}_{t\in\mathbb{Z}}$ and $\{\sigma_t^2\}_{t\in\mathbb{Z}}$ are strictly stationary and ergodic processes;*
4. *if $\{Z_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. GED(ν) random variables, with $v > 1$, zero mean and variance equal to one, then $\mathbb{E}(X_t^r) < \infty$ and $\mathbb{E}(\sigma_t^{2r}) < \infty$, for all $t \in \mathbb{Z}$ and $r > 0$.*

3 Parameter Estimation: Bayesian Inference using MCMC

Let ν be the parameter (or vector of parameters) associated to the probability density function of Z_0 and denote by

- $\boldsymbol{\eta} = (\nu, d, \theta, \gamma, \omega, \alpha_1, \cdots, \alpha_p, \beta_1, \cdots, \beta_q)' := (\eta_1, \eta_2, \cdots, \eta_{5+p+q})'$ the vector of unknown parameters in (9);
- $\boldsymbol{\eta}_{(-i)}$ the vector containing all parameters in $\boldsymbol{\eta}$ except η_i, for each $i \in \{1, \cdots, 5+p+q\}$;
- $p_Z(\cdot|\nu)$ the probability density function of Z_0 given ν;
- $\mathcal{F}_t$ the σ-algebra generated by $\{Z_s\}_{s\leq t}$;
- $p_{X_t}(\cdot|\boldsymbol{\eta}, \mathcal{F}_{t-1})$ the probability density function of X_t given $\boldsymbol{\eta}$ and $\mathcal{F}_{t-1}$, for all $t \in \mathbb{Z}$.

From (9) it is evident that, given $\boldsymbol{\eta}$, σ_t is a $\mathcal{F}_{t-1}$-measurable random variable. Moreover, since $X_t = \sigma_t Z_t$ and $p_Z(\cdot|\nu, \mathcal{F}_{t-1}) = p_Z(\cdot|\nu)$, the following equality holds

$$p_{X_t}(x_t|\boldsymbol{\eta}, \mathcal{F}_{t-1}) = \frac{1}{\sigma_t} p_Z\big(x_t\sigma_t^{-1}|\nu\big), \quad \text{with} \quad \sigma_t = \exp\left\{\frac{1}{2}\left[\omega + \sum_{k=0}^{\infty} \lambda_{d,k} g(z_{t-1-k})\right]\right\}, \tag{10}$$

for all $x_t \in \mathbb{R}$ and $t \in \mathbb{Z}$. Furthermore, from (10), the conditional probability of $\boldsymbol{X} := (X_1, \cdots, X_n)'$ given $\boldsymbol{\eta}$ and $\mathcal{F}_0$ can be written as

$$\begin{aligned} p_{\boldsymbol{X}}\big(x_1, \cdots, x_n|\boldsymbol{\eta}, \mathcal{F}_0\big) &= p_{X_n}(x_n|\boldsymbol{\eta}, x_{n-1}, \cdots, x_1, \mathcal{F}_0) \times \cdots \times p_{X_1}(x_1|\boldsymbol{\eta}, \mathcal{F}_0) \\ &= \prod_{t=1}^{n} \frac{1}{\sigma_t} p_Z\big(x_t \sigma_t^{-1}|\nu\big). \end{aligned} \tag{11}$$

Given any $I_0 \in \mathcal{F}_0$, select a prior conditional density function $p_{I_0}(\cdot|\boldsymbol{\eta})$ for I_0 given $\boldsymbol{\eta}$. Also, select a prior[1] density function $\pi_i(\cdot)$ for η_i and a prior conditional probability density function $p_{(-i)}(\cdot|\eta_i)$ for $\boldsymbol{\eta}_{(-i)}$ given η_i, for each $i \in \{1, \cdots, 5+p+q\}$.

Observe that, by applying the Bayes' rule, the conditional probability density function of η_i given $\boldsymbol{X}$, $\boldsymbol{\eta}_{(-i)}$ and any I_0, can be written as

$$p\big(\eta_i|\boldsymbol{X}, \boldsymbol{\eta}_{(-i)}, I_0\big) = \frac{p_{\boldsymbol{X}}\big(\boldsymbol{X}|\boldsymbol{\eta}, I_0\big) \times p_{I_0}\big(I_0|\boldsymbol{\eta}\big) \times p_{(-i)}(\boldsymbol{\eta}_{(-i)}|\eta_i) \times \pi_i(\eta_i)}{p_{(-i)}\big(\boldsymbol{X}, \boldsymbol{\eta}_{(-i)}, I_0\big)}, \tag{12}$$

for each $i \in \{1, \cdots, 5+p+q\}$, where $p_{\boldsymbol{X}}\big(\,\cdot\,|\boldsymbol{\eta}, \mathcal{F}_0\big)$ is given in (11) and $p_{(-i)}\big(\cdot, \cdot, \cdot\big)$ is the joint probability density function of $\boldsymbol{X}$, $\boldsymbol{\eta}_{(-i)}$ and I_0, which does not depend on η_i.

The parameter estimation is then carried out by using the MCMC method as described below.

3.1 Gibbs Sampling with Metropolis Steps

Gibbs sampling (Geman and Geman, 1984; Gelfand and Smith, 1990) is a popular MCMC algorithm for obtaining a sequence of random samples from multivariate probability distribution when direct sampling is difficult. The algorithm assumes that the conditional distribution of each random variable is known and it is easy to sample from it. The steps of the sampling procedure are the following.

Step 1. Set an arbitrary initial value for the vector of parameters $\boldsymbol{\eta}$, namely, $\boldsymbol{\eta}^{(0)} = (\eta_1^{(0)}, \cdots, \eta_{5+p+q}^{(0)})'$. Let $m = 0$;

Step 2. Given the sample $\boldsymbol{\eta}^{(m)} = (\eta_1^{(m)}, \cdots, \eta_{5+p+q}^{(m)})'$,

- generate $\eta_1^{(m+1)}$ from $p\big(\eta_1|\boldsymbol{X}, \eta_2^{(m)}, \eta_3^{(m)}, \cdots, \eta_{5+p+q}^{(m)}, I_0\big)$;
- generate $\eta_2^{(m+1)}$ from $p\big(\eta_2|\boldsymbol{X}, \eta_1^{(m+1)}, \eta_3^{(m)}, \cdots, \eta_{5+p+q}^{(m)}, I_0\big)$;

$$\vdots \qquad \vdots \qquad \vdots$$

- generate $\eta_{5+p+q}^{(m+1)}$ from $p\big(\eta_{5+p+q}|\boldsymbol{X}, \eta_1^{(m+1)}, \cdots \eta_{4+p+q}^{(m+1)}, I_0\big)$;

Step 3. Once the vector $\boldsymbol{\eta}^{(m+1)} = (\eta_1^{(m+1)}, \cdots, \eta_{5+p+q}^{(m+1)})'$ is obtained, return to step 2, with $m = m+1$, until $m = N$, where N is the desired sample size.

When it is not possible to sample directly from $p\big(\eta_i|\boldsymbol{X}, \boldsymbol{\eta}_{(-i)}, I_0\big)$, for one or more $i \in \{1, \cdots, 5+p+q\}$, an alternative option is to consider a combination of Gibbs sampler and

[1] In fact, the priors $\pi_i(\cdot)$ are not necessarily probability density functions. For instance, $\pi(x) = 1$ and $\pi(x) = 1/x$, are examples of improper priors (i.e., they do not integrate to 1) used in practice.

Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970) algorithms. This method is usually referred to as Gibbs sampler with Metropolis steps. In this case, to draw the random variate η_i, one shall follow the same steps 1-3 just described. However, instead of sampling directly from $p(\eta_i|\boldsymbol{X}, \boldsymbol{\eta}_{(-i)}, \mathcal{F}_0)$, one shall consider the Metropolis-Hastings algorithm with $p(\eta_i|\boldsymbol{X}, \boldsymbol{\eta}_{(-i)}, \mathcal{F}_0)$ as the invariant (target) distribution.

Metropolis-Hastings algorithm is easy to implement since it does not require knowing the normalization constant $p_{(-i)}(\boldsymbol{X}, \boldsymbol{\eta}_{(-i)}, I_0)$, defined in (12). For simplicity of notation, in what follows $p_*(\cdot)$ shall denote any one of the non-normalized probability density function which corresponds to $p(\eta_i|\boldsymbol{X}, \boldsymbol{\eta}_{(-i)}, \mathcal{F}_0)$, for $i \in \{1, \cdots, 5+p+q\}$. The Metropolis-Hastings sampling procedure consists if the following steps.

Step 1. Select a transition kernel[2](also called proposal distribution) $q(\cdot|\cdot)$ for which the sampling procedure is known.

Step 2. Set an arbitrary initial value y_0 for the chain. Let $m = 0$.

Step 3. Generate a draw ξ from $q(\cdot|y_m)$.

Step 4. Calculate $\alpha(y_m, \xi) = \min\left\{1, \dfrac{p_*(\xi)q(y_m|\xi)}{p_*(y_m)q(\xi|y_m)}\right\}$.

Step 5. Draw $u \sim \mathcal{U}[0, 1]$.

Step 6. Define $y_{m+1} = \begin{cases} \xi, & \text{if } u < \alpha(y_m, \xi); \\ y_m, & \text{otherwhise.} \end{cases}$

Step 7. If $m + 1 < N$ (where N is the desired sample size), let $m = m + 1$ and go to Step 3.

Remark 1.

1. When considering Gibbs sampler with Metropolis steps only one iteration of Metropolis-Hastings algorithm is performed for each Gibbs sampler iteration.
2. In both cases, Gibbs sampler and Metropolis-Hastings algorithm, it is advised to discard the first B (for some $B < N$) observations (that is, the burn-in sample) to assure the chain convergence.
3. The sample obtained from the algorithm described above is not independent. An alternative is to run parallel chains instead. Another common strategy to reduce sample autocorrelations is thinning the Markov chain, that is, to keep only every k-th simulated draw from each sequence. There is some controversy surrounding the question of whether or not it is better to run one long chain or several shorter ones (Gelman and Rubin, 1992; Geyer, 1992). Also, MacEachern and Berliner (1994) show that one always get more precise posterior estimates if the entire Markov chain is used instead of the thinned one.

4 Simulation Study

This simulation study considers FIEGARCH$(0, d, 0)$ processes. Under this scenario, the vector of unknown parameters is $\boldsymbol{\eta} = (\nu, d, \theta, \gamma, \omega)' := (\eta_1, \cdots, \eta_5)'$. The Bayesian inference approach, using MCMC to obtain posterior density functions, is used to estimate the parameters of the model.

[2] A transition kernel is a function $q(x|y)$ which is a probability measure with respect to x, so $\int q(x|y)dx = 1$.

4.1 Data Generating Process

The samples from FIEGARCH$(0, d, 0)$ processes are obtained by setting the following.

- $Z_0 \sim \text{GED}(\nu)$, with zero mean and variance equal to one. Thus,

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

- $d \in \{0.10, 0.25, 0.35, 0.45\}$ and $\nu \in \{1.1, 1.5, 1.9, 2.5, 5\}$;
- for all models, $\omega = -5.40$, $\theta = -0.15$ and $\gamma = 0.24$. These values are close to the ones already observed in practical applications (see, for instance, Nelson, 1991; Bollerslev and Mikkelsen, 1996; Ruiz and Veiga, 2008; Lopes and Prass, 2013).
- the infinite sum in (9) is truncated at $m^* = 50,000$.

For each combination of d and ν, a sample $\{z_t\}_{t=-m^*}^n$, of size $m^* + n + 1$, is drawn from the GED(ν) distribution and then the sample $\{x_t\}_{t=1}^n$, from the FIEGARCH$(0, d, 0)$ process, is obtained through the relation

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

4.2 Parameter Estimation Settings

The samples from the posterior distributions are obtained by considering the Gibbs sampler algorithm with Metropolis steps as described in Section 3. The transition kernel $q(\cdot|\cdot)$ considered in the Metropolis-Hastings algorithm is the function defined as

$$q(x|y) = f(x; y, \sigma, a, b),$$

where $f(\cdot; \cdot, \cdot, \cdot, \cdot)$ is the truncated normal density function, defined as

$$f(x; \mu, \sigma, a, b) = \begin{cases} \dfrac{1}{\sigma}\dfrac{\phi\left(\frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}, & \text{if } a \le x \le b, \\ 0, & \text{otherwise,} \end{cases}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are, respectively, the probability density and cumulative distribution functions of the standard normal distribution; $a, b \in \mathbb{R}$ are, respectively, the lower and upper limits of the distribution's support; μ and σ denote, respectively, the distribution's (non-truncated version) mean and standard deviation.

To select a reasonable $\boldsymbol{\eta}^{(0)}$, $p_{\boldsymbol{X}}\left(\boldsymbol{X}|\boldsymbol{\eta}, \mathcal{F}_0\right)$ is calculated for different combinations of ν, d, θ, γ and ω. Then $\boldsymbol{\eta}^{(0)}$ is defined as the vector $\boldsymbol{\eta} = (\nu, d, \theta, \gamma, \omega)'$ with higher likelihood function value. To eliminate any dependence on the initial $\boldsymbol{\eta}^{(0)}$ a burn-in of size 1000 is considered.

A sample obtained by the method being described will probably present significative correlation[3]. However, due to the ergodicity property of the Markov chain, the estimation of the

[3]In fact, for the parameter d, this correlation could only be removed when the thinning parameter t was set to 200.

mean is not affect by the correlation in the sample. Therefore, to avoid unnecessary computational work, which ultimately would not lead to improvement in terms of parameter estimation, thinning is not implemented. Nevertheless, an example showing the influence of using the entire chain, the thinned chain or only the first 1000 observations of the entire chain (after burn-in) is provided in the following.

Example 1. Let $\boldsymbol{\eta} = (\nu, d, \theta, \gamma, \omega)' := (\eta_1, \cdots, \eta_5)'$ and assume that

$$\pi_i(\eta_i) = \begin{cases} c_i, & \text{if } \eta_i \in I_i; \\ 0, & \text{otherwise;} \end{cases} \qquad \text{for each} \quad i \in \{1, \cdots, 5\}, \tag{13}$$

with $c_1 = 1$, $c_2 = c_3 = c_4 = 2$, $c_5 = 1/30$, $I_1 = (0, \infty)$, $I_2 = [0, 0.5]$, $I_3[-0.5, 0]$, $I_4 = I_2$ and $I_5 = [-15, 15]$.

In the sequel, $\{\eta_i^{(k)}\}_{k=1}^{\mathfrak{n}}$ denotes the chain of size $\mathfrak{n}$ obtained from the posterior distribution of η_i, upon considering the prior $\pi_i(\eta_i)$ defined in (13), for each $i \in \{1, \cdots, 5\}$. Also, $\mathfrak{b}$, $\mathfrak{t}$ and N denote, respectively, the burn-in size, the thinning parameter and the sample size of the thinned chain[4] obtained from $\{\eta_i^{(k)}\}_{k=1}^{\mathfrak{n}}$, for any $i \in \{1, \cdots, 5\}$.

Figure 1 presents the graph of $\{\eta_i^{(k)}\}_{k=1}^{\mathfrak{n}}$, for each $i \in \{1, \cdots, 5\}$, with $\mathfrak{n} = 200{,}801$. Figure 1 also shows the thinned chain of size $N = 1000$ obtained by considering $\mathfrak{b} = 1000$ and $\mathfrak{t} = 200$. Furthermore, Figure 1 gives the sample of size 1000, obtained from $\{\eta_i^{(k)}\}_{k=1}^{\mathfrak{n}}$ by considering a burn-in equal to 1000 and no thinning, for each $i \in \{1, \cdots, 5\}$. The true parameter values of the FIEGARCH$(0, d, 0)$ model corresponding to these graphs are $\nu_0 = 1.9$, $d_0 = 0.25$, $\theta_0 = -0.15$, $\gamma_0 = 0.24$ and $\omega_0 = -5.4$. Figure 2 gives the histogram and kernel density functions corresponding to each sample in Figure 1. The graphs of the prior $\pi_i(\eta_i)$ defined in (13), for $i \in \{1, \cdots, 5\}$, are represented in Figure 2 by the dashed lines. For a better visualization of the posterior distributions, in Figure 2, the range for the x-axis was restricted to the intervals $[-1.5, 2.5]$, $[-0.5, 0]$, $[0, 0.5]$ and $[-5.6, -5.1]$, respectively, for the parameters ν, θ, γ and ω.

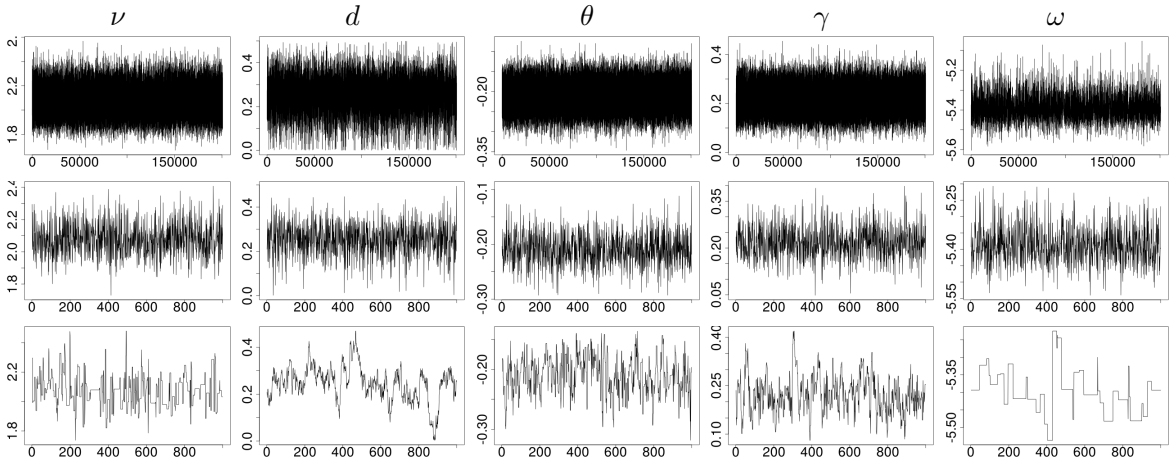


Figure 1: Original chain with sample size 200801 (top row). Thinned chain with sample size 1000 and thinning parameter equal to 200 (middle row). Unthinned chain with sample size 1000 (bottom row). For the middle and bottom rows the burn-in size is equal to 1000. The true parameter values of the FIEGARCH$(0, d, 0)$ model corresponding to these graphs are $\nu_0 = 1.9$, $d_0 = 0.25$, $\theta_0 = -0.15$, $\gamma_0 = 0.24$ and $\omega_0 = -5.4$.

[4] Observe that, by setting $\mathfrak{b} = 1000$ and $\mathfrak{t} = 200$, then a thinned chain of size $N = 1000$ can only be obtained from $\{\eta_i^{(k)}\}_{k=1}^{\mathfrak{n}}$ when $\mathfrak{n} \geq \mathfrak{b} + 1 + \mathfrak{t}(N-1) = 200{,}801$.

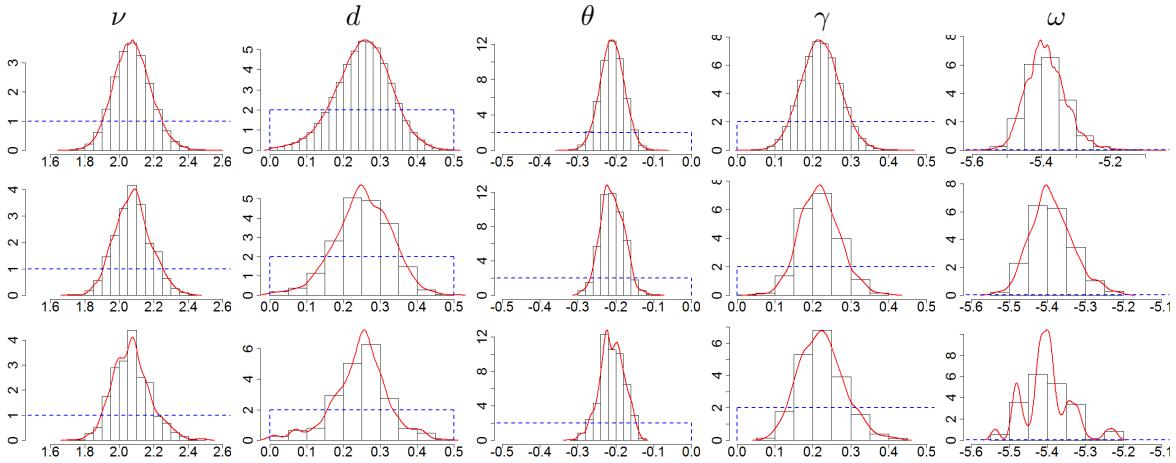


Figure 2: Histogram and kernel density functions for the original chain with sample size 200801 (top row); the thinned chain with sample size 1000 and thinning parameter equal to 200 (middle row) and the unthinned chain with sample size 1000 (bottom row). For the middle and bottom rows the burn-in size is equal to 1000. The dashed lines correspond to the graphs of the priors $\pi_i(\eta_i)$ defined in (13), for $i \in \{1, \cdots, 5\}$. The range for the x-axis was restricted to the intervals $[-1.5, 2.5]$, $[-0.5, 0]$, $[0, 0.5]$ and $[-5.6, -5.1]$, respectively, for the parameters ν, θ, γ and ω. The true parameter values of the FIEGARCH$(0, d, 0)$ model corresponding to these graphs are $\nu_0 = 1.9$, $d_0 = 0.25$, $\theta_0 = -0.15$, $\gamma_0 = 0.24$ and $\omega_0 = -5.4$.

As shown in Figure 1 (see also Table 1), the mean of the posterior distribution does not change significantly when the entire sample or the thinned chain is considered instead of the unthinned one. On the other hand, Figure 2 reinforces the idea that the entire chain gives better estimates for the density function (notice that the curves in the graphs are smoother). Although the thinned chain is not as efficient as the entire chain, it still provides better estimates for the density function than the unthinned one.

Table 1 presents the summary statistics for the samples obtained from the posterior distribution for each parameter of the FIEGARCH$(0, d, 0)$ model. This table considers the entire, thinned and unthinned chains. The statistics reported in this table are the sample mean $(\bar{\eta}_i)$, the sample standard deviation (sd_{η_i}) and the 95% credibility interval $CI_{0.95}(\eta_i)$ for the parameter η_i in $\boldsymbol{\eta} = (\nu, d, \theta, \gamma, \omega)' := (\eta_1, \cdots, \eta_5)'$, for each $i \in \{1, \cdots, 5\}$. The true parameter values considered for this illustration are $\nu_0 \in \{1.1, 1.5, 2.5, 5.0\}$, $d_0 = 0.25$, $\theta_0 = -0.15$, $\gamma_0 = 0.24$ and $\omega_0 = -5.4$.

From Table 1 it is clear that, for any η_i, with $i \in \{1, \cdots, 5\}$, the use of the entire or the thinned (thinning parameter $\mathfrak{t} = 200$) does not yield significant improvement in terms of parameter estimation. Not even the differences in the sample standard deviations or in the credibility intervals, which are the statistics affected by the sample correlation, justify the computational effort to obtain a sample of size 200,801. The same conclusions are obtained when considering $d_0 \in \{0.10, 0.35, 0.45\}$. This concludes the example.

Different prior distributions are tested as explained in the sequel. Since the conditional probability density function of I_0 given $\boldsymbol{\eta}$ is difficult to obtain, in all scenarios, it is assumed that $g(Z_s) = 0$, for all $s \leq 1$, and it is fixed $p_{I_0}(\cdot|\boldsymbol{\eta}) = 1$. Moreover, since (9) is well defined regardless the relation among the parameters of the model, it is assumed that

$$p_{(-i)}(\boldsymbol{\eta}_{(-i)}|\eta_i) \propto \prod_{j \neq i} \pi_j(\eta_j), \quad \text{for any } i \in \{1, \cdots, 5\}.$$

Table 1: Summary for the entire, thinned (thinning parameter $\mathfrak{t} = 200$) and unthinned sample from posterior distributions considering all prior uniforms: mean $\bar{\eta}_i$, standard deviation sd_{η_i} and the 95% credibility interval $CI_{0.95}(\eta_i)$ for the parameter η_i in $\boldsymbol{\eta} = (\nu, d, \theta, \gamma, \omega)' := (\eta_1, \cdots, \eta_5)'$, for each $i \in \{1, \cdots, 5\}$. The true parameter values considered in this simulation are $\nu_0 \in \{1.1, 1.5, 2.5, 5.0\}$, $d_0 = 0.25$, $\theta_0 = -0.15$, $\gamma_0 = 0.24$ and $\omega_0 = -5.4$. For the thinned and unthinned samples, the burn-in size is $\mathfrak{b} = 1000$.

Chain	ν_0	$\bar{\nu}$ (sd_ν) $CI_{0.95}(\nu)$	$\bar{d}$ (sd_d) $CI_{0.95}(d)$	$\bar{\theta}$ (sd_θ) $CI_{0.95}(\theta)$	$\bar{\gamma}$ (sd_γ) $CI_{0.95}(\gamma)$	$\bar{\omega}$ (sd_ω) $CI_{0.95}(\omega)$
Entire	1.1	1.095 (0.047) [1.006; 1.188]	0.263 (0.112) [0.035; 0.464]	-0.087 (0.038) [-0.164; -0.016]	0.233 (0.062) [0.114; 0.358]	-5.469 (0.078) [-5.612; -5.299]
	1.5	1.478 (0.067) [1.351; 1.612]	0.220 (0.081) [0.056; 0.375]	-0.184 (0.037) [-0.257; -0.111]	0.240 (0.057) [0.130; 0.355]	-5.408 (0.058) [-5.520; -5.291]
	1.9	2.077 (0.107) [1.874; 2.297]	0.252 (0.075) [0.092; 0.392]	-0.209 (0.032) [-0.272; -0.148]	0.220 (0.051) [0.122; 0.322]	-5.386 (0.058) [-5.487; -5.261]
	2.5	2.727 (0.153) [2.441; 3.040]	0.298 (0.056) [0.184; 0.405]	-0.203 (0.025) [-0.253; -0.153]	0.205 (0.045) [0.118; 0.296]	-5.361 (0.053) [-5.463; -5.252]
	5.0	5.227 (0.366) [4.548; 5.978]	0.220 (0.051) [0.115; 0.317]	-0.173 (0.019) [-0.211; -0.135]	0.294 (0.036) [0.224; 0.366]	-5.303 (0.039) [-5.384; -5.230]
Thinned	1.1	1.095 (0.047) [1.006; 1.193]	0.264 (0.110) [0.038; 0.462]	-0.087 (0.038) [-0.161; -0.015]	0.233 (0.062) [0.116; 0.362]	-5.469 (0.080) [-5.609; -5.288]
	1.5	1.480 (0.067) [1.353; 1.613]	0.221 (0.077) [0.070; 0.367]	-0.184 (0.037) [-0.257; -0.104]	0.240 (0.057) [0.129; 0.352]	-5.405 (0.060) [-5.524; -5.291]
	1.9	2.079 (0.103) [1.888; 2.292]	0.253 (0.075) [0.098; 0.387]	-0.210 (0.031) [-0.269; -0.152]	0.218 (0.052) [0.120; 0.327]	-5.388 (0.056) [-5.486; -5.261]
	2.5	2.718 (0.154) [2.427; 3.064]	0.298 (0.058) [0.181; 0.410]	-0.202 (0.026) [-0.255; -0.153]	0.205 (0.044) [0.121; 0.291]	-5.361 (0.053) [-5.466; -5.254]
	5.0	5.224 (0.363) [4.534; 5.991]	0.218 (0.052) [0.112; 0.316]	-0.173 (0.019) [-0.211; -0.137]	0.295 (0.037) [0.224; 0.370]	-5.302 (0.039) [-5.391; -5.230]
Unthinned	1.1	1.108 (0.039) [1.028; 1.198]	0.265 (0.129) [0.016; 0.467]	-0.089 (0.038) [-0.176; -0.019]	0.230 (0.060) [0.108; 0.345]	-5.476 (0.052) [-5.553; -5.366]
	1.5	1.474 (0.067) [1.352; 1.644]	0.250 (0.071) [0.123; 0.393]	-0.175 (0.035) [-0.248; -0.100]	0.240 (0.053) [0.135; 0.353]	-5.426 (0.057) [-5.535; -5.344]
	1.9	2.072 (0.109) [1.885; 2.303]	0.245 (0.074) [0.068; 0.400]	-0.210 (0.032) [-0.273; -0.152]	0.223 (0.059) [0.116; 0.349]	-5.400 (0.062) [-5.537; -5.244]
	2.5	2.720 (0.148) [2.412; 3.016]	0.308 (0.060) [0.192; 0.426]	-0.200 (0.026) [-0.257; -0.155]	0.198 (0.042) [0.119; 0.281]	-5.367 (0.058) [-5.462; -5.274]
	5.0	5.311 (0.356) [4.676; 6.070]	0.226 (0.045) [0.137; 0.316]	-0.176 (0.019) [-0.215; -0.140]	0.293 (0.036) [0.225; 0.368]	-5.291 (0.035) [-5.346; -5.246]

In a first moment the prior distributions for ν, d, θ, γ and ω are selected by considering only the basic set of information usually available in practice. The information on each parameter and the corresponding prior selected are given in Table 2. This scenario shall be referred to as Case 1. Table 3 presents the mean, standard deviation, lower and upper limits for the transition kernel considered at iteration m of the Gibbs sampler with Metropolis steps, when the prior for η_i in $\boldsymbol{\eta} = (\nu, d, \theta, \gamma, \omega)' := (\eta_1, \cdots, \eta_5)'$, for each $i \in \{1, \cdots, 5\}$, is defined according to Case 1.

In a second moment the knowledge on the true parameter values is gradually incorporated to provide more informative priors for d, θ and/or γ. This analysis, combined with the first scenario, provides information on the sensitivity of the estimates with respect to the priors functions and hyperparameters. In all cases, the priors for ν and ω are the same and are the

Table 2: Information available in practice for the parameter η_i in $\boldsymbol{\eta} = (\nu, d, \theta, \gamma, \omega)' := (\eta_1, \cdots, \eta_5)'$ and the corresponding prior considered, for each $i \in \{1, \cdots, 5\}$.

Information Available	**Prior**
The generalized error distribution is well defined for any $\nu > 0$.	$\nu \sim \mathbb{I}_{(0,\infty)}(\nu)$ *
Long-memory in volatility is observed if and only if $d \in (0, 0.5)$. This characteristic can be detected, for example, through the periodogram function of the time series $\{\ln(X_t^2)\}_{t=1}^n$ (see Lopes and Prass, 2013).	$d \sim \mathcal{U}(0, 0.5)$
Empirical evidence suggests that $\theta \in [-1, 0]$. **	$\theta \sim \mathcal{U}(-1, 0)$
Empirical evidence suggests that $\gamma \in [0, 1]$. **	$\gamma \sim \mathcal{U}(0, 1)$
$\omega = \mathbb{E}(\ln(h_t^2)) = \mathbb{E}(\ln(X_t^2)) + \mathbb{E}(\ln(Z_t^2))$. The choice of the interval for ω will depend on the magnitude of the data. The sample mean of $\{\ln(X_t^2)\}_{t=1}^n$ or $\ln(\hat{\sigma}_X^2)$, where $\hat{\sigma}_X^2$ is the sample variance of $\{X_t\}_{t=1}^n$, can be used to obtain a rough approximation for ω	$\omega \sim \mathcal{U}(-15, 15)$.

Notes: * Given $A \subset \mathbb{R}$, the symbol $\mathbb{I}_A(x)$ denotes the improper prior defined as 1, if $x \in A$, and 0, if $x \notin A$.
** See, for instance, Nelson (1991); Bollerslev and Mikkelsen (1996); Ruiz and Veiga (2008); Lopes and Prass (2013). To the best of our knowledge, a FIEGARCH model for which θ or γ are not in the intervals, respectively, $[-1, 0]$ and $[0, 1]$ has never been reported in the literature.

Table 3: Parameters of the truncated normal distribution (transition kernel) considered, at iteration m of the Gibbs sampler, to obtain the sample from the posterior distribution of the parameter η_i in $\boldsymbol{\eta} = (\nu, d, \theta, \gamma, \omega)' := (\eta_1, \cdots, \eta_5)'$, for each $i \in \{1, \cdots, 5\}$.

Parameter	ν	d	θ	γ	ω
Mean (y)	$\nu^{(m-1)}$	$d^{(m-1)}$	$\theta^{(m-1)}$	$\gamma^{(m-1)}$	$\omega^{(m-1)}$
Standard Deviation (σ)	0.500	0.025	0.050	0.050	1.500
Lower Limit (a)	0.000	0.000	-1.000	0.000	-15.000
Upper Limit (b)	10.000	0.500	0.000	1.000	15.000

Note: $\eta_i^{(m-1)}$, for any $i \in \{1, \cdots, 5\}$, denotes the parameter value obtained in the $(m-1)$th iteration. Different combinations of standard deviation, lower and upper limits were tested for the parameter η_i in $\boldsymbol{\eta} = (\nu, d, \theta, \gamma, \omega)'$, for each $i \in \{1, \cdots, 5\}$. The values presented here correspond to the final choice.

ones defined in Table 2. The scenarios considered in this second step are described in the following and shall be referred to as Case 2 - Case 5.

Case 2: Gaussian Prior for $x = \phi^{-1}(d)$ and Uniform Priors for θ and γ.

In this case θ and γ remain with the same priors as in Case 1. For the parameter d it is assumed that $x \sim \mathcal{N}(\mu_\phi, \sigma_\phi^2)$ and $d = \phi(x)$, where $\phi : \mathbb{R} \to (0, 0.5)$ is given by

$$\phi(x) = \frac{e^x}{2(1 + e^x)}, \quad \text{for all } x \in \mathbb{R}. \tag{14}$$

First, the knowledge of d_0 is applied to set $\mu_\phi = \phi^{-1}(d_0)$, so $\mu_\phi \in \{-1.386, 0.000, 0.847, 2.197\}$,

respectively, for $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$. This scenario shall be referred to as C2.1. Second, the knowledge on d_0 is ignored and the parameter μ_ϕ is assumed to be equal to zero. This scenario shall be referred to as C2.2. For both, C2.1 and C2.2, different values of σ_ϕ are tested. Third, the approaches considered in C2.1 and C2.2 are combined by setting $\mu_\phi = \phi^{-1}(\bar{d})$, where $\bar{d}$ is the estimate of d obtained in C2.2. This scenario shall be referred to as C2.3. The value of σ_ϕ considered in C2.3 is the one which provides better estimates for d in C2.1.

The kernel parameter values for ν, θ, γ and ω are the same as in the Case 1. For $x = \phi^{-1}(d)$, at iteration m of the Gibbs sampler, the kernels mean (y), standard deviation (sd), lower (a) and upper limits (b) are set, respectively, as $x^{(m-1)}$, 1, -10 and 10, where $x^{(m-1)}$ is the parameter value obtained at iteration $m-1$.

Case 3: Gaussian Prior for $x = \phi^{-1}(d)$, Beta Prior for $-\theta$ and Uniform Prior for γ.

In this case, the priors of γ and d are the same ones considered, respectively, in Case 1 and in scenario C2.1 of Case 2. It is also assumed that $-\theta \sim \text{Beta}(a_1, b_1)$, which is equivalent to set

$$\pi_3(\theta) = (-\theta)^{a_1-1}(1+\theta)^{b_1-1}B(a_1, b_1), \quad \theta \in [-1, 0],$$

where $B(\cdot, \cdot)$ is the beta function.

First, the fact that $X \sim \text{Beta}(a, b)$ implies $\mathbb{E}(X) = a(a+b)^{-1}$, is applied to set $b_1 = a_1(1+\theta_0)(-\theta_0)^{-1}$, where $\theta_0 = -0.15$ is the true parameter value considered in this simulation study. Different values of a_1 are tested. This scenario shall be referred to as C3.1. Second, the knowledge on θ_0 is ignored and different combinations of a_1 and b_1 are tested. This scenario shall be referred to as C3.2. Third, the approaches considered in C3.1 and C3.2 are combined by setting $b_1 = a_1(1+\bar{\theta}_0)(-\bar{\theta}_0)^{-1}$, where $-\bar{\theta}_0$ is the estimate of θ obtained in C3.2. The value of a_1 considered in this case is the one which provides better estimates for θ in C3.1. This scenario shall be referred to as C3.3.

The kernel parameter values are the same as in Case 2.

Case 4: Gaussian Prior for $x = \phi^{-1}(d)$, Beta Priors for $-\theta$ and γ.

In this case, the priors of d and $-\theta$ are the same ones considered, respectively, in scenario C2.1 of Case 2 and in scenario C3.1 of Case 3. It is also assumed that $\gamma \sim \text{Beta}(a_2, b_2)$. Two scenarios, denoted by C4.1 and C4.2 are considered. With the obvious identifications, the construction of C4.1 and C4.2 is analogous, respectively, to the construction of scenarios C3.1 and C3.2 in Case 3.

The kernel parameter values are the same as in Case 2.

Case 5: Beta Priors for $2d$, $-\theta$ and γ.

In this case, the priors of $-\theta$ and γ are the same ones considered, respectively, in scenario C3.1 of Case 3 and in scenario C4.1 of Case 4. Moreover, for each $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$ considered in this simulation study, it is assumed that $2d \sim \text{Beta}(a_3, b_3)$, which is equivalent to set

$$\pi_2(d) = 2(2d)^{a_3-1}(1-2d)^{b_3-1}B(a_3, b_3), \quad d \in [0, 0.5],$$

where $B(\cdot, \cdot)$ is the beta function.

In this case, only two scenarios are considered. First, it is assumed that $b_3 = a_3(1 - 2d_0)(2d_0)^{-1}$ and different values of a_3 are tested. This scenario shall be referred to as C5.1. Second, an approach similar to scenarios C3.3 and C4.3, respectively, in Case 3 and Case 4, is considered. However, in this case, it is assumed that $b_3 = a_3(1 - 2\bar{d})(2\bar{d})^{-1}$, with $\bar{d}$ obtained in Case 1. The value of a_3 considered in this case is the one which provides better estimates for d in C5.1. This scenario shall be referred to as C5.2.

The kernel parameter values are the same as in Case 1.

4.3 Estimates and Performance Measures

Let $\{\eta_i^{(k)}\}_{k=1}^M$ be a sample of size M from the posteriori distribution of η_i in $\boldsymbol{\eta} = (\nu, d, \theta, \gamma, \omega)' := (\eta_1, \cdots, \eta_5)'$, for any $i \in \{1, \cdots, 5\}$. Denote by $\bar{\eta}_i$ and sd_{η_i}, respectively, the sample mean and standard deviation of $\{\eta_i^{(k)}\}_{k=1}^M$, namely,

$$\bar{\eta}_i = \frac{1}{M}\sum_{k=1}^{M}\eta_i^{(k)} \quad \text{and} \quad \text{sd}_{\eta_i} = \sqrt{\frac{1}{M}\sum_{k=1}^{M}(\eta_i^{(k)} - \bar{\eta}_i)^2}, \quad \text{for any} \quad i \in \{1, \cdots, 5\}.$$

Then the estimate $\hat{\eta}_i$ of η_i is defined as $\hat{\eta}_i := \bar{\eta}_i$.

Moreover, let $\hat{q}_i(\alpha)$ denote the α quantile[5] for the posterior sample distribution of η_i, for any $\alpha \in [0, 1]$ and $i \in \{1, \cdots, 5\}$. Then a $100(1 - \alpha)\%$ credibility interval for η_i is given by

$$CI_{1-\alpha}(\eta_i) = \left[\hat{q}_i\left(\frac{\alpha}{2}\right), \hat{q}_i\left(1 - \frac{\alpha}{2}\right)\right], \quad \text{for any} \quad i \in \{1, \cdots, 5\}.$$

Furthermore, the estimation bias and the absolute percentage error (ape) of estimation are given, respectively, by

$$\text{bias}_{\eta_i} = \bar{\eta}_i - \eta_i \quad \text{and} \quad \text{ape}_{\eta_i} = \left|\frac{\text{bias}_{\eta_i}}{\eta_i}\right|, \quad \text{for any} \quad i \in \{1, \cdots, 5\}.$$

4.4 Results

The results obtained in this simulation study, by considering the scenarios described in Section 4.2, are the following.

Case 1: The Priors as Defined in Table 2.

Table 4 present the summary statistics for the samples obtained from the posterior distribution for each parameter of the FIEGARCH$(0, d, 0)$ model. The statistics reported in this table (the same applies to Table 5) are the sample mean ($\bar{\eta}_i$), the sample standard deviation (sd_{η_i}) and the 95% credibility interval $CI_{0.95}(\eta_i)$ for the parameter η_i in $\boldsymbol{\eta} = (\nu, d, \theta, \gamma, \omega)' := (\eta_1, \cdots, \eta_5)'$, for each $i \in \{1, \cdots, 5\}$. The bold-face font for the mean indicates that the absolute percentage error of estimation (ape_{η_i}) in the corresponding case is higher than 0.10 (that is, 10%). The bold-face font for the credibility interval indicates that the true parameter value is not contained in the interval.

[5] In this work, the following definition is adopted (Brockwell and Davis, 1991). Given any $0 \leq \alpha \leq 1$, the number $q(\alpha)$ satisfying $\mathbb{P}(X \leq q(\alpha)) \geq \alpha$ and $\mathbb{P}(X \geq q(\alpha)) \geq 1 - \alpha$, is called a quantile of order α (or α quantile) for the random variable X (or for the distribution function of X).

Table 4: Summary for the sample obtained from posterior distributions considering all prior uniforms: mean $\bar{\eta}_i$, standard deviation sd_{η_i} and the 95% credibility interval $CI_{0.95}(\eta_i)$ for the parameter η_i in $\boldsymbol{\eta} = (\nu, d, \theta, \gamma, \omega)' := (\eta_1, \cdots, \eta_5)'$, for each $i \in \{1, \cdots, 5\}$. The true parameter values considered in this simulation are $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$, $\nu_0 \in \{1.1, 1.5, 2.5, 5.0\}$, $\theta_0 = -0.15$, $\gamma_0 = 0.24$ and $\omega_0 = -5.4$.

d_0	ν_0	$\bar{\nu}$ (sd_ν) $CI_{0.95}(\nu)$	$\bar{d}$ (sd_d) $CI_{0.95}(d)$	$\bar{\theta}$ (sd_θ) $CI_{0.95}(\theta)$	$\bar{\gamma}$ (sd_γ) $CI_{0.95}(\gamma)$	$\bar{\omega}$ (sd_ω) $CI_{0.95}(\omega)$
0.10	1.1	1.093 (0.044) [0.989; 1.195]	**0.181** (0.123) [0.005; 0.458]	**-0.084** (0.041) [-0.171; -0.013]	0.236 (0.066) [0.093; 0.357]	-5.438 (0.058) [-5.551; -5.337]
	1.5	1.480 (0.069) [1.353; 1.635]	**0.147** (0.079) [0.020; 0.330]	**-0.177** (0.038) [-0.258; -0.106]	0.232 (0.052) [0.122; 0.340]	-5.420 (0.036) [-5.510; -5.338]
	1.9	2.088 (0.111) **[1.908; 2.296]**	0.093 (0.055) [0.004; 0.201]	**-0.220** (0.032) **[-0.286; -0.154]**	**0.216** (0.060) [0.105; 0.337]	-5.410 (0.035) [-5.486; -5.340]
	2.5	2.724 (0.140) [2.491; 3.027]	**0.192** (0.076) [0.038; 0.330]	**-0.201** (0.025) **[-0.256; -0.153]**	**0.198** (0.045) [0.116; 0.287]	-5.388 (0.031) [-5.448; -5.333]
	5.0	5.297 (0.364) [4.641; 6.014]	0.101 (0.051) [0.015; 0.217]	**-0.174** (0.020) [-0.215; -0.133]	**0.297** (0.036) [0.232; 0.374]	-5.336 (0.028) **[-5.383; -5.287]**
0.25	1.1	1.108 (0.039) [1.028; 1.198]	0.265 (0.129) [0.016; 0.467]	**-0.089** (0.038) [-0.176; -0.019]	0.230 (0.060) [0.108; 0.345]	-5.476 (0.052) [-5.553; -5.366]
	1.5	1.474 (0.067) [1.352; 1.644]	0.250 (0.071) [0.123; 0.393]	**-0.175** (0.035) [-0.248; -0.100]	0.240 (0.053) [0.135; 0.353]	-5.426 (0.057) [-5.535; -5.344]
	1.9	2.072 (0.109) [1.885; 2.303]	0.245 (0.074) [0.068; 0.400]	**-0.210** (0.032) **[-0.273; -0.152]**	0.223 (0.059) [0.116; 0.349]	-5.400 (0.062) [-5.537; -5.244]
	2.5	2.720 (0.148) [2.412; 3.016]	**0.308** (0.060) [0.192; 0.426]	**-0.200** (0.026) **[-0.257; -0.155]**	**0.198** (0.042) [0.119; 0.281]	-5.367 (0.058) [-5.462; -5.274]
	5.0	5.311 (0.356) [4.676; 6.070]	0.226 (0.045) [0.137; 0.316]	**-0.176** (0.019) [-0.215; -0.140]	**0.293** (0.036) [0.225; 0.368]	-5.291 (0.035) **[-5.346; -5.246]**
0.35	1.1	1.099 (0.040) [1.009; 1.194]	0.349 (0.108) [0.093; 0.492]	**-0.097** (0.038) [-0.178; -0.027]	0.230 (0.056) [0.121; 0.330]	-5.495 (0.090) [-5.674; -5.318]
	1.5	1.479 (0.065) [1.352; 1.639]	0.329 (0.065) [0.204; 0.461]	**-0.178** (0.036) [-0.246; -0.106]	0.246 (0.052) [0.143; 0.340]	-5.423 (0.076) [-5.561; -5.302]
	1.9	2.064 (0.110) [1.843; 2.299]	0.364 (0.054) [0.227; 0.461]	**-0.199** (0.030) [-0.265; -0.139]	0.233 (0.050) [0.139; 0.330]	-5.377 (0.090) [-5.535; -5.199]
	2.5	2.732 (0.150) [2.481; 3.031]	0.380 (0.052) [0.283; 0.479]	**-0.201** (0.024) **[-0.254; -0.161]**	**0.200** (0.043) [0.110; 0.285]	-5.307 (0.066) [-5.410; -5.149]
	5.0	5.229 (0.321) [4.603; 5.864]	0.318 (0.040) [0.243; 0.409]	**-0.177** (0.019) [-0.216; -0.140]	**0.289** (0.036) [0.227; 0.366]	-5.227 (0.046) **[-5.298; -5.127]**
0.45	1.1	1.096 (0.039) [1.010; 1.161]	0.436 (0.053) [0.313; 0.499]	**-0.115** (0.034) [-0.187; -0.047]	0.241 (0.054) [0.136; 0.338]	-5.453 (0.128) [-5.716; -5.174]
	1.5	1.475 (0.073) [1.353; 1.627]	0.411 (0.048) [0.311; 0.494]	**-0.179** (0.034) [-0.246; -0.110]	0.257 (0.048) [0.158; 0.353]	-5.411 (0.123) [-5.600; -5.133]
	1.9	2.052 (0.111) [1.846; 2.279]	0.450 (0.032) [0.385; 0.497]	**-0.191** (0.026) [-0.238; -0.141]	0.243 (0.043) [0.165; 0.320]	-5.367 (0.130) [-5.614; -5.125]
	2.5	2.725 (0.152) [2.461; 3.021]	0.447 (0.032) [0.381; 0.495]	**-0.206** (0.021) **[-0.247; -0.167]**	**0.211** (0.041) [0.133; 0.296]	-5.150 (0.081) **[-5.310; -4.985]**
	5.0	5.177 (0.322) [4.553; 5.832]	0.417 (0.032) [0.348; 0.480]	**-0.177** (0.019) [-0.220; -0.140]	**0.286** (0.032) [0.228; 0.350]	-5.041 (0.068) **[-5.182; -4.929]**

Note: The bold-face font for the estimated **mean** indicates that the absolute percentage error is higher than 10%. The bold-face font for the **credibility interval** indicates that the interval does not contain the true parameter value.

From Table 4 one observes that the parameters ν and ω are always well estimated, in terms of absolute percentage error (ape), regardless the combination of $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$ and $\nu_0 \in \{1.1, 1.5, 2.5, 5.0\}$ considered (the error is less than 10% in all cases). The credibility interval $CI_{0.95}(\nu)$ contains the true parameter value ν_0 in all cases, except when $d_0 = 0.10$ and $\nu_0 = 1.9$. Also, the estimation bias for ν is always negative when $\nu_0 < 1.9$ and positive when $\nu_0 \geq 1.9$, except for the combination $(\nu_0, d_0) = (1.1, 0.25)$. For the parameter ω, the credibility interval $CI_{0.95}(\omega)$ does not contain the true parameter value ($\omega_0 = -5.4$) in 5 out of 20 combinations of ν_0 and d_0 (see $\nu_0 = 5$ and all d_0; $\nu_0 = 2.5$ and $d_0 = 0.45$). Moreover, the estimation bias for ω is always negative when $\nu_0 \leq 1.5$ and always positive when $\nu_0 \geq 2.5$.

Table 4 also reports that $\text{ape}_\theta > 10\%$ for all combinations of d_0 and ν_0. On the other hand, in most cases (14 out of 20), the credibility interval $CI_{0.95}(\theta)$ contains the true parameter value $\theta_0 = -0.15$. The cases for which $\theta_0 \notin CI_{0.95}(\theta)$ are $\nu = 1.9$ and $d_0 \in \{0.10, 0.25\}$ and $\nu_0 = 2.5$ and any d_0. The bias for θ is always positive when $\nu_0 = 1.1$ (for any d_0) and negative in all other cases.

Furthermore, Table 4 shows that the parameter γ seems to be better estimated when the GED distribution presents heavy tails ($\nu_0 < 2$), except when $d = 0.10$, in which case $\text{ape}_\gamma > 10\%$ when $\nu_0 = 1.9$. Also, with exception of four cases ($d_0 = 0.10$ and $\nu_0 \in \{1.1, 1.5, 2.5\}$; $d_0 = 0.25$ and $\nu_0 = 2.5$), the parameter d is always well estimated. The bias for parameters d and γ does not seem to follow any pattern and both, $d_0 \in CI_{0.95}(d)$ and $\gamma_0 \in CI_{0.95}(\gamma)$, for any combination of ν_0 and d_0.

Case 2: Gaussian Prior for $x = \phi^{-1}(d)$ and Uniform Priors for θ and γ.

Changing the prior for d does not yield significant difference on the estimation of ν, θ, γ and ω.

When the true value of d_0 is used to set $\mu_\phi = \phi^{-1}(d_0)$ (scenario C2.1), the best performance is observed by letting $\sigma_\phi = 0.15$. In this case, the absolute percentage error of estimation (ape_d) is smaller than 10% for all combinations of ν_0 and d_0, with $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$ and $\nu_0 \in \{1.1, 1.5, 2.5, 5.0\}$. If $\sigma_\phi = 0.10$ the chain takes too long to move from the initial point when $d_0 = 0.10$. When $\sigma_\phi = 0.25$, there is only one case for which $\text{ape}_d > 10\%$ ($d_0 = 0.10$ and $\nu_0 = 2.5$). In fact, in this case, $\text{ape}_d = 0.103$, which is still acceptable ($\sigma_\phi = 0.15$ still seems to be the best choice). Furthermore, as σ_ϕ increases, the number of cases for which $\text{ape}_d > 10\%$ also increases. For instance, when $\sigma_\phi \in \{0.50, 1.00, 3.00\}$, $\text{ape}_d > 10\%$ in 2, 4 and 10 cases, respectively.

When d_0 is assumed unknown and μ_ϕ is set to zero (scenario C2.2), $\sigma_\phi = 3$ seems to provide better results than smaller values of σ_ϕ. Under this scenario, $\text{ape}_d > 10\%$ for 8 out of 20 combinations of ν_0 and d_0. Therefore, $d \sim \mathcal{U}(0, 0.5)$ still provides better estimates for the parameter d (see Table 4). Higher values of σ_ϕ do not improve the estimation of d. Too high values of σ_ϕ actually make the estimation worst. In particular, when $\sigma_\phi = 4$ the results are similar to $\sigma_\phi = 3$, if $d_0 > 0.1$. If $d_0 = 0.1$ then $\sigma_\phi = 3$ is slightly better than $\sigma_\phi = 4$. When $\sigma_\phi = 5$, ape_d is, in most cases, higher than when $\sigma_\phi = 3$. For σ_ϕ smaller than 3 the estimation bias is much higher. For instance, when $\sigma_\phi = 0.15$, $\text{ape}_d \leq 10\%$ only for $d_0 = 0.25$ (for all ν_0). For all other combinations of d_0 and ν_0 $\text{ape}_d > 20\%$. Also, when $b = 1$, $\text{ape}_d > 20\%$ in 12 out of 20 cases. In particular, $\text{ape}_d > 20\%$ for $d_0 = 0.10$ and all ν_0. As it should be expected, C2.1 performs much better than C2.2.

Upon considering a two step estimator (scenario C2.3), no improvement is observed, when compared to scenario C2.2. In fact, the estimates obtained by letting $\mu_\phi = \phi^{-1}(\bar{d})$ (where $\bar{d}$ is the estimate of d obtained in C2.2) and $\sigma_\phi = 0.15$ (the parameter which leads to the best performance in C2.1) are very close to $\bar{d}$ itself.

Case 3: Gaussian Prior for $x = \phi^{-1}(d)$, Beta Prior for $-\theta$ and Uniform Prior for γ.

The estimation of ν, d, γ and ω is not significantly affected by the change in the prior for $-\theta$.

When the knowledge on the true parameter values d_0 and θ_0 is applied to set $\mu_\phi = \phi^{-1}(d_0)$, $\sigma_\phi = 0.15$, for each $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$ (the best scenario in Case 2), and $b_1 = a_1(1 + \theta_0)(-\theta_0)^{-1}$ (scenario C3.1), it is observed that larger values of a_1 lead to better estimates for θ. Although any $a_1 \in \{110, 150, 200\}$ leads to $\text{ape}_\theta \leq 10\%$, for all combinations of $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$ and $\nu_0 \in \{1.1, 1.5, 2.5, 5.0\}$, the best performance is obtained by setting $a_1 = 110$ ($b_1 \approx 623.333$). As a_1 decreases, the estimation performance decreases. For instance, when $a_1 = 100$, one case for which $\text{ape}_\theta > 10\%$ is observed. When $a_1 = 20$ the number of cases for which $\text{ape}_\theta > 10\%$ increases to 10 and no case for which $\text{ape}_\theta \leq 10\%$ is observed if $a_1 \in \{2.0, 0.1\}$. More specifically, for any $a_1 \in \{2.0, 0.1\}$, $10\% < \text{ape}_\theta \leq 20\%$ for $\nu_0 \in \{1.5, 5.0\}$ and all values of d_0 (8 out of 20 cases) and, in the remaining 12 cases, $\text{ape}_\theta > 20\%$.

By assuming θ_0 unknown (scenario C3.2) or by considering a two step estimator (scenario C3.3), no case for which $\text{ape}_\theta < 10\%$ is observed. The combinations of a_1 and b_1 tested in scenario C3.2 are: $(a_1, b_1) \in \{(2, 3), (2, 5), (2, 9), (4, 7), (5, 7), (10, 40), (10, 60), (10, 70), (100, 500), (100, 600)\}$. Among these values, the best performance is obtained when $a_1 = 10$ and $b_1 = 50$. In this case, $10\% < \text{ape}_\theta \leq 20\%$ in 12 out of 20 cases, which is slightly better than the performance obtained assuming $\theta \sim \mathcal{U}(0, 1)$ (in this case, $10\% < \text{ape}_\theta \leq 20\%$ in 8 out of 20 cases).

Case 4: Gaussian Prior for $x = \phi^{-1}(d)$, Beta Priors for $-\theta$ and γ.

Analogously to Case 2 and Case 3, the estimation of ν, d, θ and ω is not significantly affected by the change in the prior for γ.

By considering the true parameter values d_0, θ_0 and γ_0 and setting $\mu_\phi = \phi^{-1}(d_0)$, $\sigma_\phi = 0.15$, for each $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$ (the best scenario in Case 2), $a_1 = 110$, $b_1 = a_1(1 + \theta_0)(-\theta_0)^{-1}$ (the best scenario in Case 3) and $b_2 = a_2(1 - \gamma_0)\gamma_0^{-1}$ (scenario C4.1), it is observed the following: larger values of a_2 (smaller than a_1, however) lead to better estimates for γ and as a_2 decreases, the estimation performance decays. For instance, when $a_2 = 40$ only one case for which $\text{ape}_\theta > 10\%$ is observed and when $a_2 \in \{10, 25, 30\}$, the number of cases increases to 5 ($d_0 = 0.45$ and all ν_0). On the other hand, any $a_2 \in \{50, 100\}$ gives $\text{ape}_\theta \leq 10\%$, for all combinations of $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$ and $\nu_0 \in \{1.1, 1.5, 2.5, 5.0\}$. The simulation results for $a_2 = 50$ ($b_2 \approx 158.333$) are illustrated in Figure 3.

Figure 3 shows the sample mean (solid circle) and the 95% credibility interval (solid line) for the sample obtained from the posterior distribution of ν, d, θ, γ and ω (respectively, from top to bottom), for each combination of d_0 and ν_0. The true parameter values $\nu_0, d_0, \theta_0, \gamma_0$ and ω_0 are represented in the corresponding row by the dashed line. The graphs related to θ, γ and ω (respectively, the third, fourth and fifth rows, from top to bottom) consider the same scale for all $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$. Also, for the parameters θ, γ and ω, there is one graph for each d_0 and, for each one of these graphs, the true value of ν_0 is indicated in the x-axis.

From Figure 3 one observes that, for ν and ω, the conclusion regarding the estimation bias and the credibility intervals are basically the same as in Case 1 (see Table 4). On the other hand, under C4.1 of Case 4), $\text{ape}_{\eta_i} < 10\%$, for all $i \in \{1, \cdots, 5\}$ and any combination of $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$ and $\nu_0 \in \{1.1, 1.5, 2.5, 5.0\}$ (compare the parameter θ in Table 4 and Figure 3). As in Case 1, the bias for θ is always positive when $\nu_0 = 1.1$ (for any d_0) and negative when $\nu_0 > 1.1$ and the bias for the parameters d and γ does not seem to follow any pattern. Under the current scenario, d_0, θ_0 and γ_0 are all contained in the respective credibility

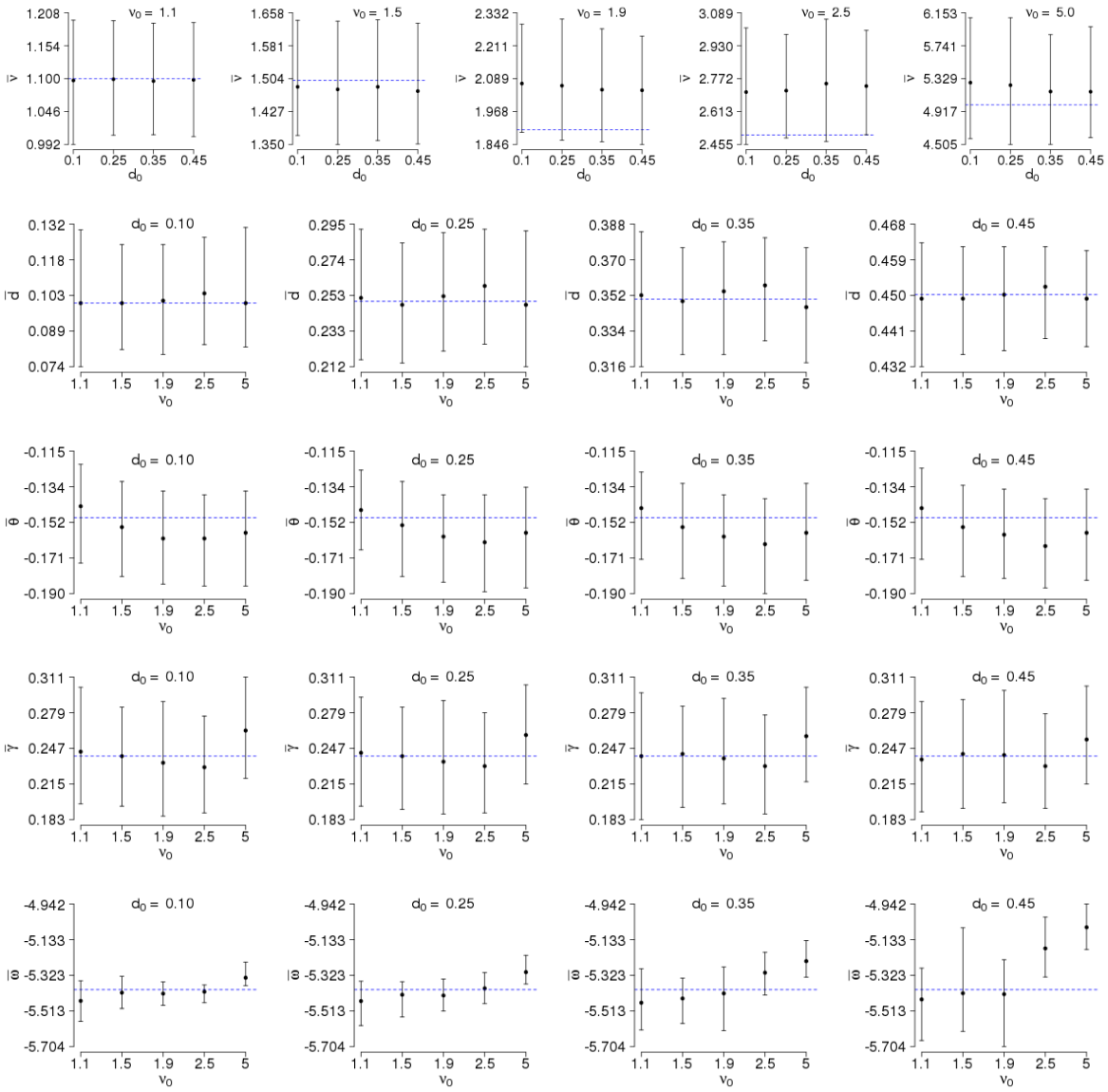


Figure 3: Posterior mean (solid circle), the true parameter value (dashed line) and the 95% credibility interval (solid line) for the parameters ν, d, θ, γ and ω (from top to bottom), for each combination of d_0 and ν_0. The posterior distributions were obtained by considering an improper prior for ν, a Gaussian prior for $\phi^{-1}(d)$, Beta priors for $-\theta$ and γ and a uniform prior for ω. The true parameters values considered in this simulation are $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$, $\nu_0 \in \{1.1, 1.5, 2.5, 5.0\}$, $\theta_0 = -0.15$, $\gamma_0 = 0.24$ and $\omega_0 = -5.4$.

intervals, for any combination of $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$ and $\nu_0 \in \{1.1, 1.5, 2.5, 5.0\}$.

When the true value of γ_0 is not used to choose b_2 (scenario C4.2) similar results to the ones in Figure 3 are still obtained for some combinations of (a_2, b_2). Not surprisingly, the pairs (a_2, b_2) which lead to good estimates are such that $a_2(a_2 + b_2)^{-1}$ (the mean μ_B of the prior distribution) is close to γ_0. For instance, when $(a_2, b_2) \in \{(100, 300), (100, 350)\}$ (μ_B is, respectively, equal to 0.25 and 0.22, while $\gamma_0 = 0.24$) it is obtained $\text{ape}_\gamma < 10\%$ for all combinations of $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$ and $\nu_0 \in \{1.1, 1.5, 2.5, 5.0\}$. The pair $(a_2, b_2) = (100, 350)$ provides slightly better results than $(a_2, b_2) = (100, 300)$ only when $\nu_0 = 5$. When $(a_2, b_2) = (100, 280)$ (so $\mu_B \approx 0.26$), $\text{ape}_\gamma \leq 10\%$ in 16 out of 20 cases (in the remaining 4 cases,

ape_γ does not exceed 13.4%).

On the other hand, choosing a_2 and b_2 such that $a_2(a_2+b_2)^{-1}$ is close to the true γ_0 does not necessarily lead to good estimates. For instance, if $(a_2, b_2) = (5, 15)$ then $\mu_B = 0.25$, as it is when $(a_2, b_2) = (100, 300)$, but $\text{ape}_\gamma > 10\%$ in 6 out of 20 cases. Also, it is not evident that the more distant $a_2(a_2+b_2)^{-1}$ is from γ_0, the worst is the estimation. For instance, by letting $(a_2, b_2) \in \{(3, 15), (100, 440), (10, 30), (10, 40), (20, 80), (100, 400), (100, 270), (5, 10)\}$ then, respectively, $\mu_B \in \{0.167, 0.185, 0.200, 0.200, 0.200, 0.200, 0.270, 0.330\}$ and it is observed that $\text{ape}_\gamma > 10\%$ in 14, 20, 4, 10, 12, 16, 13 and 5 out of 20 cases, respectively.

Case 5: Beta Priors for $2d$, $-\theta$ and γ.

Analogously to all other cases, the estimation of ν, θ, γ and ω is not significantly affected by the change in the prior for d.

Upon assuming $-\theta \sim \text{Beta}(a_1, b_1)$ and $\gamma \sim \text{Beta}(a_2, b_2)$, with the same a_1, a_2, b_1 and b_2 as in scenario C4.1 of Case 4, and letting $2d \sim \text{Beta}(a_3, b_3)$, with $b_3 = a_3(1-2d_0)(2d_0)^{-1}$, for each $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$ (scenario C5.1 of Case 5), the following is concluded. If $a_3 \in \{25, 50\}$ then $\text{ape}_d < 10\%$ for all $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$. By increasing or decreasing too much the a_3 values the estimation performance decays. For instance, $a_3 \in \{0.10, 0.20, 2.00\}$ yields $\text{ape}_d > 10\%$ in 3, 1 and 7 cases, respectively.

Table 5 reports the simulation results for $a_3 = 25$ and $b_3 = a_3(1-2d_0)(2d_0)^{-1}$, which gives $b_1 \in \{100.000, 25.000, 10.714, 2.778\}$, respectively, for $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$. The conclusions on the results presented in this table are the same as in Figure 3. Although the credibility intervals for d are slightly wider in Table 5 than in Figure 3, in both tables $d_0 \in CI_{0.95}(d)$ for any combination of $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$ and $\nu_0 \in \{1.1, 1.5, 2.5, 5.0\}$.

As in Case 2, when considering a two step estimator (scenario C5.2), no improvement is observed, when compared to Case 1. In fact, once again, the estimates obtained by letting $a_3 = 25$ and $b_3 = a_3(1-2\bar{d})(2\bar{d})^{-1}$, where $\bar{d}$ is the estimate of d obtained in Case 1, are very close to $\bar{d}$ itself.

5 Conclusions

The Bayesian inference approach for parameter estimation on FIEGARCH models was described and a Monte Carlo simulation study was conducted to analyze the performance of the method under the presence of long-memory in volatility. The samples from FIEGARCH processes were obtained by considering the infinite sum representation for the logarithm of the volatility. A recurrence formula was used to obtain the coefficients for this representation. The generalized error distribution, with different tail-thickness parameters was considered so both innovation processes with lighter and heavier tails than the Gaussian distribution, were covered.

Markov Chain Monte Carlo (MCMC) methods where used to obtain samples from the posterior distribution of the parameters. A sensitivity analysis was performed by considering the following steps. First, an improper prior for ν and uniform priors d, θ, γ and ω were selected. In this case, only the basic set of information usually available in practice was considered. Second, non-uniform priors were selected for one or more parameters in $\{d, \theta, \gamma\}$. A Gaussian prior for $\phi^{-1}(d)$, with $\phi(\cdot)$ defined in (14), combined with uniform or Beta priors for θ ($-\theta$ in the Beta case) and γ was considered. In the sequel, a comparison was made by assuming Beta priors for $2d$, $-\theta$ and γ. The sensitivity analysis was completed by integrating (or not) the knowledge on

Table 5: Summary for the sample obtained from posterior distributions considering Beta priors for $2d, -\theta$ and γ: mean $\bar{\eta}_i$, standard deviation sd_{η_i} and the 95% credibility interval $CI_{0.95}(\eta_i)$ for the parameter η_i in $\boldsymbol{\eta} = (\nu, d, \theta, \gamma, \omega)' := (\eta_1, \cdots, \eta_5)'$, for each $i \in \{1, \cdots, 5\}$. The true parameter values considered in this simulation are $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$, $\nu_0 \in \{1.1, 1.5, 2.5, 5.0\}$, $\theta_0 = -0.15$, $\gamma_0 = 0.24$ and $\omega_0 = -5.4$.

d_0	ν_0	$\bar{\nu}$ (sd_ν) $CI_{0.95}(\nu)$	$\bar{d}$ (sd_d) $CI_{0.95}(d)$	$\bar{\theta}$ (sd_θ) $CI_{0.95}(\theta)$	$\bar{\gamma}$ (sd_γ) $CI_{0.95}(\gamma)$	$\bar{\omega}$ (sd_ω) $CI_{0.95}(\omega)$
0.10	1.1	1.102 (0.045) [1.012; 1.198]	0.100 (0.018) [0.069; 0.139]	-0.144 (0.011) [-0.170; -0.124]	0.243 (0.026) [0.199; 0.294]	-5.472 (0.063) [-5.569; -5.347]
	1.5	1.483 (0.063) [1.364; 1.607]	0.102 (0.019) [0.069; 0.141]	-0.154 (0.013) [-0.178; -0.132]	0.239 (0.024) [0.190; 0.285]	-5.412 (0.047) [-5.524; -5.333]
	1.9	2.067 (0.108) [1.879; 2.309]	0.101 (0.016) [0.070; 0.134]	-0.160 (0.012) [-0.186; -0.137]	0.234 (0.027) [0.184; 0.291]	-5.422 (0.034) [-5.498; -5.360]
	2.5	2.702 (0.143) [2.442; 2.977]	0.106 (0.019) [0.074; 0.150]	-0.160 (0.012) [-0.186; -0.136]	0.231 (0.023) [0.187; 0.277]	-5.407 (0.022) [-5.470; -5.375]
	5.0	5.251 (0.391) [4.514; 6.080]	0.099 (0.017) [0.070; 0.132]	-0.158 (0.012) [-0.186; -0.134]	0.263 (0.024) [0.220; 0.312]	-5.343 (0.036) **[-5.390; -5.265]**
0.25	1.1	1.095 (0.044) [0.986; 1.190]	0.255 (0.031) [0.197; 0.315]	-0.145 (0.012) [-0.173; -0.123]	0.242 (0.028) [0.191; 0.295]	-5.455 (0.057) [-5.559; -5.348]
	1.5	1.485 (0.066) [1.357; 1.628]	0.252 (0.030) [0.193; 0.313]	-0.154 (0.013) [-0.180; -0.132]	0.240 (0.022) [0.195; 0.281]	-5.421 (0.046) [-5.529; -5.324]
	1.9	2.058 (0.113) [1.864; 2.316]	0.260 (0.030) [0.194; 0.309]	-0.159 (0.012) [-0.185; -0.138]	0.236 (0.028) [0.189; 0.293]	-5.431 (0.048) [-5.525; -5.340]
	2.5	2.719 (0.146) [2.452; 3.008]	0.271 (0.029) [0.216; 0.326]	-0.163 (0.012) [-0.188; -0.137]	0.229 (0.021) [0.186; 0.275]	-5.386 (0.038) [-5.469; -5.304]
	5.0	5.208 (0.326) [4.548; 5.871]	0.244 (0.028) [0.190; 0.299]	-0.159 (0.013) [-0.185; -0.134]	0.260 (0.023) [0.213; 0.309]	-5.313 (0.034) **[-5.381; -5.256]**
0.35	1.1	1.097 (0.038) [1.014; 1.175]	0.355 (0.034) [0.283; 0.413]	-0.145 (0.012) [-0.169; -0.126]	0.241 (0.027) [0.186; 0.298]	-5.469 (0.080) [-5.630; -5.289]
	1.5	1.481 (0.064) [1.359; 1.628]	0.349 (0.030) [0.285; 0.403]	-0.154 (0.012) [-0.178; -0.131]	0.239 (0.024) [0.192; 0.285]	-5.447 (0.077) [-5.587; -5.310]
	1.9	2.070 (0.102) [1.870; 2.288]	0.370 (0.029) [0.306; 0.421]	-0.160 (0.012) [-0.183; -0.136]	0.236 (0.027) [0.186; 0.293]	-5.414 (0.082) [-5.587; -5.237]
	2.5	2.720 (0.147) [2.449; 3.009]	0.375 (0.028) [0.321; 0.426]	-0.163 (0.011) [-0.185; -0.143]	0.228 (0.023) [0.186; 0.274]	-5.337 (0.063) [-5.440; -5.221]
	5.0	5.147 (0.346) [4.598; 5.965]	0.344 (0.027) [0.287; 0.395]	-0.159 (0.012) [-0.185; -0.137]	0.258 (0.023) [0.214; 0.305]	-5.255 (0.048) **[-5.321; -5.171]**
0.45	1.1	1.101 (0.042) [1.024; 1.191]	0.454 (0.024) [0.402; 0.489]	-0.145 (0.012) [-0.169; -0.122]	0.238 (0.026) [0.189; 0.286]	-5.424 (0.128) [-5.682; -5.160]
	1.5	1.493 (0.073) [1.362; 1.645]	0.450 (0.024) [0.395; 0.488]	-0.154 (0.012) [-0.177; -0.132]	0.243 (0.024) [0.195; 0.291]	-5.414 (0.132) [-5.681; -5.139]
	1.9	2.045 (0.109) [1.844; 2.241]	0.464 (0.019) [0.419; 0.491]	-0.158 (0.011) [-0.181; -0.134]	0.239 (0.026) [0.193; 0.291]	-5.424 (0.126) [-5.717; -5.216]
	2.5	2.742 (0.142) **[2.507; 3.010]**	0.466 (0.017) [0.431; 0.493]	-0.164 (0.012) [-0.191; -0.141]	0.228 (0.022) [0.189; 0.275]	-5.170 (0.084) **[-5.308; -5.002]**
	5.0	5.164 (0.346) [4.558; 5.883]	0.447 (0.022) [0.396; 0.486]	-0.158 (0.011) [-0.182; -0.137]	0.256 (0.022) [0.214; 0.300]	-5.070 (0.076) **[-5.227; -4.942]**

Note: The bold-face font for the **credibility interval** indicates that the interval does not contain the true parameter value.

the true parameter values to select the hyperparameter values.

An example was presented to illustrate the similarities or differences on the mean, standard deviation and credibility intervals estimated by considering a chain of size $N = 200801$, a thinned chain (thinning parameter 200 and burn-in size 1000) and a sample of size 1000 (obtained from the larger chain, after the burn-in of size 1000). Given the ergodicity of the Markov chain, the posterior means for all three chains were very close. The differences on the standard deviations and credibility intervals are not significant enough to justify the use of the entire or thinned chains. Although the example only presents the case $d_0 = 0.25$, the same conclusions apply to $d_0 \in \{0.10, 0.35, 0.45\}$.

The simulation study showed that if the prior of one or more parameters is changed, the estimation of the other parameters is not significantly affected. The parameters ν and ω are always well estimated, in terms of absolute percentage error, regardless priors considered for d, θ and γ, for any combination of $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$ and $\nu_0 \in \{1.1, 1.5, 2.5, 5.0\}$. With a few exceptions, the true parameter value ν_0 was contained in the 95% credibility interval, for any combination of $\nu_0 \in \{1.1, 1.5, 1.9, 2.5, 5.0\}$ and $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$ considered. The true parameter value ω_0 was not contained in any credibility interval when $\nu_0 = 5$.

Regardless the prior considered, the parameter d is usually better estimated when $d \in \{0.35, 0.45\}$. The Gaussian prior for $\phi^{-1}(d)$ only provided better results (globally) when the knowledge on the true parameter value d_0 was used to set $\mu_\phi = \phi^{-1}(d_0)$ and σ_ϕ was set to some value smaller or equal than 1. In particular, only when $b = 0.15$ the absolute percentage error of estimation (ape) became smaller than 10% for all $d_0 \in \{0.10, 0.25, 0.35, 0.45\}$. Although the credibility intervals for d are slightly wider when a Beta prior is considered, the use of the Beta prior for $2d$ neither improves nor degrades the estimation performance, compared to the Gaussian prior for $\phi^{-1}(d)$.

The absolute percentage error of estimation for θ (ape_θ) only became smaller than 10% when the Beta prior was considered and the true value of the parameter was used to select the hyperparameter. When θ_0 was assumed unknown the ape_θ was always between 10% and 38.1%. The parameter γ is always better estimated than θ, for any priors considered. Similar to d and θ, the best performance is obtained when the true parameter value is used to select the hyperparameters. On the other hand, γ is the only parameter for which there are hyperparameter values that do not yield $\mu_B = \gamma_0$ (μ_B is the mean of the prior distribution and γ_0 is the true parameter value) while still providing good estimates.

Acknowledgments

T.S. Prass was supported by CNPq-Brazil. 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. The authors are grateful to the (Brazilian) National Center of Super Computing (CESUP-UFRGS) for the computational resources.

References

Baillie, R.; T. Bollerslev and H.O. Mikkelsen (1996). "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity". *Journal of Econometrics*, vol. 74, 3-30.

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

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

Breidt, F.; N. Crato and P.J.F. de Lima (1998). "On the Detection and Estimation of Long Memory in Stochastic Volatility". *Journal of Econometrics*, vol. 83, 325-348.

Brockwell, P.J. and R.A. Davis (1991). *Time Series: Theory and Methods*. Second Edition. New York: Springer-Verlag.

Casella, G. and E.I. George (1992). "Explaining the Gibbs Sampler", *The American Statistician*, vol 46(3), 167-174.

Chib, S. and E. Greenberg (1995). "Understanding the Metropolis-Hastings Algorithm", *The American Statistician*, vol. 49(4), 327-335.

Engle, R.F. (1982). "Autoregressive Conditional Heteroskedasticity with Estimates of Variance of U.K. Inflation". *Econometrica*, vol. 50, 987-1008.

Gelfand, A.E. and A.F.M Smith (1990). "Sampling-Based Approaches to Calculating Marginal Densities". *Journal of the American Statistical Association*, vol. 85(410), 398-409.

Geman, S. and D. Geman (1984). "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, 609-628.

Gelman, A. and D. Rubin (1992). "Inference from Iterative Simulation Using Multiple Sequences". *Statistical Science*, vol. 7, 457-511.

Geyer, C. J. (1992). "Practical Markov chain Monte Carlo". *Statistical Science*, vol. 7, 473-511.

Hastings, W.K. (1970). "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". *Biometrika*, vol. 57(1), 97-109.

Lopes, S.R.C and T.S. Prass (2013). "Theoretical Results on Fractionally Integrated Exponential Generalized Autoregressive Conditional Heteroskedastic Processes." Working Paper.

MacEachern, S.N. and L.M. Berliner (1994). "Subsampling the Gibbs Sampler". *The American Statistician*, vol. 48, 188-190.

Metropolis, N.; A.W. Rosenbluth; M.N. Rosenbluth; A.H. Teller and E. Teller (1953). "Equations of State Calculations by Fast Computing Machines". *Journal of Chemical Physics*, vol. 21(6), 1087-1092.

Meyer, R. and T. Yu (2000). "Bugs for a Bayesian Analysis of Stochastic Volatility Models". *Econometrics Journal*, vol. 3, 198-215.

Nelson, D.B. (1991). "Conditional Heteroskedasticity in Asset Returns: A New Approach". *Econometrica*, vol. 59, 347-370.

Ruiz E. and H. Veiga (2008). "Modelling long-memory volatilities with leverage effect: A-LMSV versus FIEGARCH". *Computational Statistics and Data Analysis*, vol. 52(6), 2846-2862.

Smith, A.F.M. and G.O. Roberts (1993). "Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods". *Journal of the Royal Statistical Society*, Ser. B, vol. 55, 3-23.

Apêndice I

Artigo Prass, Lopes e Crato (2013)

Autores: Taiane S. Prass, Sílvia R. C. Lopes e Nuno Crato

Título: SFIEGARCH Processes with α-Stable Innovations: a Bayesian Approach for Parameter Estimation

Revista: Em Fase de Preparação

Ano: 2013

SFIEGARCH Processes with α-Stable Innovations: a Bayesian Approach for Parameter Estimation

Taiane S. Prass*, Sílvia R.C. Lopes and Nuno Crato

September 10, 2013

Abstract

Here we develop the theory of seasonal FIEGARCH processes with stable innovations, denoted by α-SFIEGARCH, establishing conditions for their existence, invertibility, stationarity and ergodicity properties. Bayesian inference for α-SFIEGARCH models using Markov Chain Monte Carlo (MCMC) methods is described. A simulation study is presented to access the performance of the procedure, under the presence of long-range dependence in the volatility. Different values for the tail index parameter α are considered, including $\alpha = 2$, which corresponds to the Gaussian distribution. By considering $s \in \{1, 7\}$ both scenarios, non-seasonal and seasonal long-range dependence behaviors are covered. Empirical applications, considering two observed time series is also provided. The first data set corresponds to the São Paulo stock exchange index log-return time series, observed in the period from January 02, 1995 to December 10, 2001. The second data set is the one-hour log-return time series corresponding to the S&P500 stock exchange index observed in the period from March 21, 2007 to August 13, 2009.

Key words: Bayesian inference, MCMC, SFIEGARCH processes, Stable innovations, Long-range dependence.

1 Introduction

Stable laws (also called α-stable, stable Paretian or Lévy stable) were introduced by Lévy (1925) when the author was investigating the behavior of sums of independent random variables. A random variable X is said to be stable (or it has a stable distribution) if it has the stability property, that is, for any $a, b > 0$, there exist $c > 0$ and $d \in \mathbb{R}$ such that

$$aX_1 + bX_2 \stackrel{d}{=} cX + d, \tag{1}$$

where X_1 and X_2 are two independent copies of X and "$\stackrel{d}{=}$" denotes equality of distributions. The most popular stable laws are the Gaussian, Cauchy and Lévy distributions. Furthermore, according to the generalized central limit theorem, stable laws are the only possible limit distributions for properly normalized and centered sums of independent, identically distributed random variables (Laha and Rohatgi, 1979).

Alternative definitions for stable random variables can be found in the literature (see, for instance, Samorodnitsky and Taqqu, 1994). In particular, Feller (1971) states that X is a stable random variable if for n independent copies X_i of X, there exist constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that

$$X_1 + \cdots + X_n \stackrel{d}{=} c_n X + d_n. \tag{2}$$

*Corresponding Author: taianeprass@gmail.com

Stable laws are usually called α-stable because $c_n = n^{1/\alpha}$, for some $0 < \alpha \leq 2$ (Feller, 1971).

Probability density functions of α-stable distributions have no analytical representation, except for the Gaussian, Cauchy and Lévy distributions. On the other hand, the characteristic function for any stable law is given by

$$\varphi(t) = \begin{cases} \exp\left\{\mathrm{i}\mu t - \sigma^\alpha |t|^\alpha \left[1 - \mathrm{i}\beta \operatorname{sgn}(t) \tan(\frac{\pi\alpha}{2})\right]\right\}, & \text{if } \alpha \neq 1, \\ \exp\left\{\mathrm{i}\mu t - \sigma|t| \left[1 + \mathrm{i}\beta \frac{2}{\pi} \operatorname{sgn}(t) \ln(|t|)\right]\right\}, & \text{if } \alpha = 1, \end{cases} \tag{3}$$

where $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\sigma \geq 0$, $\mu \in \mathbb{R}$, $\mathrm{i} = \sqrt{(-1)}$ and $\operatorname{sgn}(\cdot)$ is the sign function defined by

$$\operatorname{sgn}(t) = \begin{cases} 1, & \text{se} \quad t > 0, \\ 0, & \text{se} \quad t = 0, \\ -1, & \text{se} \quad t < 0. \end{cases}$$

Expression (3) shows that a stable law is, in fact, a four-parameter distribution. The parameter α is called the *characteristic exponent* or *tail index* and describes the tail of the distribution. The parameter β, σ and μ are known, respectively, as *skewness*, *scale* and *location* parameters.

The class of stable laws became very popular in financial modeling for two main reasons. First, α-stable distributions have the stability property and domains of attraction and hence they provide very good approximations for a wide range of observable data (Kozubowski et al., 2003). Second, not only stable laws generalize the Gaussian distribution, but also they allow for heavy tails and skewness, which are frequently observed in financial data. The Student's t distribution also allows for heavier tails than the Gaussian one but it lacks the stability property (Panorska et al., 1995).

With the recent advances in computer technology, researchers are having the opportunity to analyze financial data sets collected at a finer time scale. Among the discoveries, economists have reported that volatility of high frequency financial time series shows long-range dependence merged with periodic behavior. In the case of exchange rate returns, these patterns are generally attributed to different openings of European, Asian and North American markets superimposed each other (Bordignon et al., 2007). Similar behavior is observed in stock markets, mainly due to the so-called time-of-day phenomena, such as market opening, closing operations, lunch-hour and overlapping effects.

To model the long-range dependence and periodic behavior of the volatility, some processes, all considering finite variance innovations, were proposed in the literature. In particular, Bordignon et al. (2007, 2009) introduced new GARCH-type models characterized by long memory behavior of periodic type. The generalized long memory GARCH (G-GARCH) introduces generalized periodic long-memory filters, based on Gegenbauer polynomials, into the equation describing the time-varying volatility of standard GARCH models. Results on the weak stationarity of these processes are unknown (Bordignon et al., 2007). The periodic long-memory GARCH (PLM-GARCH) process represents a natural extension of the FIGARCH model proposed for modeling the volatility long-range persistence. According to Bordignon et al. (2009), PLM-GARCH are not weak stationary but it can be shown that, under certain conditions, strict stationarity holds. Periodic long memory log-GARCH (PLM-LGARCH) and EGARCH (PLM-EGARCH) models were also mentioned, without further discussion on their theoretical properties, in Bordignon et al. (2009).

More recently, Lopes and Prass (2013b) introduced the seasonal FIEGARCH (SFIEGARCH) processes which (as the PLM-EGARCH model) is an extension of the well known FIEGARCH process (Bollerslev and Mikkelsen, 1996). Among the results presented in Lopes and Prass (2013b), the authors provide the conditions for existence and stationarity of the SFIEGARCH$(p, d, q)_s$ process, for any $p, q > 0$, and the asymptotic dependence structure of both $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ and $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ processes, where $\{X_t\}_{t\in\mathbb{Z}}$ is an SFIEGARCH process and $\{\sigma_t^2\}_{t\in\mathbb{Z}}$ is its conditional variance. The spectral representation of both processes $\{\ln(\sigma_t^2)\}_{t\in\mathbb{Z}}$ and $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ is also provided. The theoretical results given in that paper are then applied to describe the volatility of the S&P500 US

stock index log-return time series in the period from December 13, 2004 to October 02, 2009. The residuals analysis provided in Lopes and Prass (2013b) reinforces the evidence that financial returns have heavy-tailed distributions.

In this work we consider symmetric α-stable distributions ($\mu = \beta = 0$), denoted by $S\alpha S$ and introduce α-SFIEGARCH$(p, d, q)_s$ processes, for $\alpha \in (1, 2]$. The assumption $\alpha > 1$ is not very restrictive in the context of financial modeling, because most financial time series of interest appear to have finite means. When $\alpha = 2$, we obtain the Gaussian SFIEGARCH(p, d, q) model covered in Lopes and Prass (2013b). The conditions for the existence and stationarity of α-SFIEGARCH processes are presented. The Bayesian inference approach to estimate the model parameters, which uses Markov Chain Monte Carlo (MCMC) to obtain posterior probability density functions, is also described. The paper is organized as follows: in the next section we present some basic ideas on α-FIEGARCH processes. In Section 3 we describe the parameter estimation procedure. In Section 4 we present a simulated study showing how to generate α-FIEGARCH models and how the proposed Bayesian approach performs. In Section 5 the Bayesian inference procedure is applied to estimate the parameters of α-SFIEGARCH models for two observed time series. Section 6 concludes the paper.

2 SFIEGARCH Processes with Stable Innovations

In this section we present the definition and properties of the SFIEGARCH process with stable innovations. We denote by $S_\alpha(\beta, \sigma, \mu)$ the stable distribution with parameters α, β, σ and μ. The symmetric stable distribution ($\mu = \beta = 0$) is referred to as $S\alpha S$.

Assume that $\{Z_t\}_{t\in\mathbb{Z}}$ be a sequence of independent and identically distributed (i.i.d.) random variables, with $Z_0 \sim S\alpha S$, for some $\alpha \in (1, 2]$ and $\sigma > 0$. Let $\theta, \gamma \in \mathbb{R}$, with $|\theta| + |\gamma| > 0$, and define $\{g(Z_t)\}_{t\in\mathbb{Z}}$ as

$$\begin{aligned} g(Z_t) &:= \theta Z_t + \gamma[|Z_t| - \mathbb{E}(|Z_t|)] \\ &= \theta Z_t + \gamma\left[|Z_t| - \frac{2\sigma}{\pi}\Gamma\left(1 - \frac{1}{\alpha}\right)\right], \quad \text{for all } t \in \mathbb{Z}, \end{aligned} \tag{4}$$

where $\Gamma(\cdot)$ is the gamma function.

Let $a(\cdot)$ and $b(\cdot)$ be two polynomials, respectively, of order p and q, with no common roots, given by

$$a(z) = \sum_{i=0}^{p}(-a_i)z^i \quad \text{and} \quad b(z) = \sum_{j=0}^{q}(-b_j)z^j, \tag{5}$$

with $a_0 = b_0 = -1$ and $b(z) \neq 0$ in the closed disk $\{z : |z| \leq 1\}$. Moreover, let $\mathcal{B}$ be the backward shift operator defined by $\mathcal{B}^k(X_t) = X_{t-k}$, for all $k \in \mathbb{N}$, and $(1 - \mathcal{B}^s)^{-d}$, for $s \in \mathbb{N} \setminus \{0\}$, be the operator defined by

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

Definition 1. A stochastic process $\{X_t\}_{t\in\mathbb{Z}}$ is an SFIEGARCH *process with stable innovations*, denoted by α-SFIEGARCH$(p, d, q)_s$, if it can be written as

$$X_t = h_t Z_t \tag{7}$$

$$\ln(h_t^2) = \omega + \frac{a(\mathcal{B})}{b(\mathcal{B})}(1 - \mathcal{B}^s)^{-d}g(Z_{t-1}), \quad \text{for all } t \in \mathbb{Z}, \tag{8}$$

where $\{Z_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. $S\alpha S$ random variables with scaling parameter $\sigma = 1$ and tail index $\alpha \in (1, 2]$; $\omega \in \mathbb{R}$; $s \in \mathbb{N}\setminus\{0\}$; $d < 1 - \frac{1}{\alpha}$; $g(\cdot)$ is the function defined in (4); $a(\cdot)$ and $b(\cdot)$ are the polynomials given in (5) and $(1-\mathcal{B}^s)^{-d}$ is the operator defined in (6).

Remark 1. Simple calculations show that, in Definition 1, if σ is allowed to assume any non-negative value and θ, γ, ω, $\{Z_t\}_{t\in\mathbb{Z}}$ and $\{h_t\}_{t\in\mathbb{Z}}$, are replaced, respectively, by

$$\theta^* := \sigma\theta, \quad \gamma^* := \sigma\gamma, \quad \omega^* := \omega + \ln(\sigma^2), \quad Z_t^* := \sigma^{-1}Z_t \quad \text{and} \quad h_t^* := \sigma h_t,$$

for all $t \in \mathbb{Z}$, then the same stochastic process $\{X_t\}_{t\in\mathbb{Z}}$ is obtained. Therefore, the assumption $\sigma = 1$ is necessary if one desires the random variables $\{Z_t\}_{t\in\mathbb{Z}}$ and $\{h_t\}_{t\in\mathbb{Z}}$ and the parameters θ, γ and ω to be uniquely defined.

Remark 2. Long-range dependence is usually defined in terms of the asymptotic behavior of the autocorrelation (equivalently, autocovariance) and/or spectral density functions. However, for infinite variance stochastic process such as the α-SFIEGARCH just defined, those functions are meaningless. Therefore, following the ideas in Kokoszka and Taqqu (1995), long-range dependence for infinite variance processes shall be defined considering the equivalent Gaussian process. In another words, we shall say that an α-SFIEGARCH process has long-range dependence in the volatility if the stochastic process obtained upon replacing the $S\alpha S$ random variables $\{Z_t\}_{t\in\mathbb{Z}}$ by an i.i.d. Gaussian sequence does.

It is easy to see that, in Definition 1, if we replace the stable innovations $\{Z_t\}_{t\in\mathbb{Z}}$ by an i.i.d. sequence of random variables, with mean zero and variance equal to one, then we obtain the SFIEGARCH process defined in Lopes and Prass (2013b). In particular, when $s = 1$ we obtain a FIEGARCH process (Bollerslev and Mikkelsen, 1996; Lopes and Prass, 2013a) and, when $d = 0$, we have the EGARCH process (Nelson, 1991).

Alternatively, let $\lambda(\cdot)$ be the function defined by

$$\lambda(z) := \frac{a(z)}{b(z)}(1-z^s)^{-d} = \sum_{k=0}^{\infty}\lambda_{d,k}z^k, \quad \text{for all } |z| < 1, \tag{9}$$

and observe that this expansion is unique (see Lopes and Prass, 2013b). Moreover, $\{\lambda_{d,k}\}_{k\in\mathbb{Z}}$ can be obtained recursively by (see Lopes and Prass, 2013b)

$$\lambda_{d,k} = \begin{cases} 1, & \text{if} \quad k = 0, \\ -a_k + \displaystyle\sum_{i=0}^{k-1}\lambda_{d,i}\Big(\sum_{j=0}^{(k-i)\wedge q}\delta^*_{d,\frac{k-i-j}{s}}b_j\Big), & \text{if} \quad k \le p; \\ \displaystyle\sum_{i=0}^{k-1}\lambda_{d,i}\Big(\sum_{j=0}^{(k-i)\wedge q}\delta^*_{d,\frac{k-i-j}{s}}b_j\Big), & \text{if} \quad k > p, \end{cases} \tag{10}$$

where $(k-i)\wedge q = \min\{k-i, q\}$ and

$$\delta^*_{d,m} = \begin{cases} \delta_{d,m}, & \text{if} \quad m \in \mathbb{N} \\ 0, & \text{if} \quad m \notin \mathbb{N} \end{cases}$$

with $\delta_{d,m}$, for all $m \in \mathbb{N}$, given in (6). Thus (8) may be rewritten as

$$\ln(h_t^2) = \omega + \sum_{k=0}^{\infty}\lambda_{d,k}g(Z_{t-1-k}), \quad \text{for all } t \in \mathbb{Z}. \tag{11}$$

Furthermore, proceeding as in the proof of theorem 2.1 in Kokoszka and Taqqu (1995), it can be shown that the stochastic process $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$ satisfying (11) is also the unique solution of

$$b(\mathcal{B})\left(\ln(h_t^2)-\omega\right)=a(\mathcal{B})(1-\mathcal{B}^s)^{-d}g(Z_{t-1}),\quad \text{for all } t\in\mathbb{Z}. \tag{12}$$

Therefore, definitions (8), (11) and (12) are all equivalent.

From (8) and/or (11) one concludes that, if the random variables $\{h_t\}_{t\in\mathbb{Z}}$ are finite with probability one, then h_t is a $\mathcal{F}_{t-1}$-measurable random variable, where $\mathcal{F}_t$ denotes the σ-algebra generated by $\{Z_k\}_{k\leq t}$. However, since Z_0 does not have finite variance, the random variable h_t no longer is the conditional standard deviation of $X_t|\mathcal{F}_{t-1}$, for $t\in\mathbb{Z}$, as it would be in the finite variance case (Lopes and Prass, 2013b). In fact, from the properties of stable distributions and the $\mathcal{F}_{t-1}$-measurability of h_t, it follows that, if $Z_t\sim S_\alpha(0,\sigma,0)$, with $\alpha\in(1,2]$, then $X_t|\mathcal{F}_{t-1}\sim S_\alpha(0,\sigma h_t,0)$, for all $t\in\mathbb{Z}$. In particular, if $\sigma=1$ then h_t itself is the scaling parameter for the distribution of $X_t|\mathcal{F}_{t-1}$.

Observe that Z_t is finite with probability one, $h_t=\exp\left\{\frac{1}{2}\omega+\frac{1}{2}\sum_{k=0}^{\infty}\lambda_{d,k}g(Z_{t-1-k})\right\}$ and $X_t=h_tZ_t$, for all $t\in\mathbb{Z}$. Therefore, once we have proved that h_t, for all $t\in\mathbb{Z}$, is finite with probability one, it will follow immediately that $\{X_t\}_{t\in\mathbb{Z}}$ is well defined. Also, to prove that the random variable h_t, for all $t\in\mathbb{Z}$, is finite with probability one, it is enough to show that the sum (11) is almost surely convergent. To prove the almost sure convergence of (11) we need the results given in Lemma 1 and Theorem 1.

Lemma 1 shows that, if $Z_0\sim S\alpha S$, with $\alpha\in(1,2]$, then the r-th moment of $\{g(Z_t)\}_{t\in\mathbb{Z}}$ is finite, for any $0<r<\alpha$. This result shall be used to prove a sufficient condition for the convergence of (11). Notice that the i.i.d., the strict stationarity and the ergodicity properties of $\{g(Z_t)\}_{t\in\mathbb{Z}}$ follow immediately from the i.i.d. property of $\{Z_t\}_{t\in\mathbb{Z}}$.

Lemma 1. *Let $\{Z_t\}_{t\in\mathbb{Z}}$ be an i.i.d. sequence of $S\alpha S$ random variables, with $\alpha\in(1,2]$, and $\{g(Z_t)\}_{t\in\mathbb{Z}}$ be the stochastic process defined by (4). Then*

$$\mathbb{E}(|g(Z_t)|^r)<\infty,\quad \textit{for all}\quad 0<r<\alpha.$$

Proof. Notice that $\mathbb{E}(|Z_0|^r)<\infty$, for all $0<r<\alpha$ (Samorodnitsky and Taqqu, 1994). In particular, since $\alpha\in(1,2]$, $\mathbb{E}(|Z_0|)<\infty$. The strict stationarity of $\{g(Z_t)\}_{t\in\mathbb{Z}}$ implies that $\mathbb{E}(|g(Z_t)|^r)=\mathbb{E}(|g(Z_0)|^r)$, for all $t\in\mathbb{Z}$. Upon applying the c_r-inequality (Roussas, 2005) with $X:=|Z_0|$ and $Y:=\mathbb{E}(|Z_0|)$, it follows that $\mathbb{E}\left(|\gamma[|Z_0|-\mathbb{E}(|Z_0|)]|^r\right)<\infty$, for all $0<r<\alpha$. The c_r-inequality applied now to $X:=\theta Z_0$ and $Y:=\gamma\left[|Z_0|-\mathbb{E}(|Z_0|)\right]$ yields

$$\mathbb{E}(|g(Z_0)|^r)=\mathbb{E}\left(|\theta Z_0+\gamma\left[|Z_0|-\mathbb{E}(|Z_0|)\right]|^r\right)=\mathbb{E}(|X+Y|^r)<c_r\left[\mathbb{E}\left(|X|^r\right)+\mathbb{E}\left(|Y|^r\right)\right]<\infty,$$

for all $0<r<\alpha$, and the result follows. ∎

Theorem 1 gives the asymptotic behavior of $\mathbb{P}(|g(Z_0)|>x)$, as x goes to infinity. This result is fundamental so we can apply the three series theorem (Roussas, 2005, theorem 11, page 380) to provide a necessary condition for the convergence of (11).

Theorem 1. *Let $\{Z_t\}_{t\in\mathbb{Z}}$ be an i.i.d. sequence of $S\alpha S$ random variables, with $\alpha\in(1,2]$, and $\{g(Z_t)\}_{t\in\mathbb{Z}}$ be the stochastic process defined by (4). Then*

$$\mathbb{P}(|g(Z_0)|>x)\sim C_{\alpha,\sigma}\left(|\theta+\gamma|^\alpha+|\theta-\gamma|^\alpha\right)x^{-\alpha},\quad \textit{as } x\to\infty,$$

where

$$C_{\alpha,\sigma}=\sigma^\alpha\left(\int_0^\infty x^{-\alpha}\sin(x)dx\right)^{-1}=\frac{\sigma^\alpha}{\Gamma(1-\alpha)\cos\left(\frac{\pi\alpha}{2}\right)}.$$

Proof. Assuming known (see, for instance, Samorodnitsky and Taqqu, 1994) that

$$\mathbb{P}(Z_0 > x) \sim C_\alpha \sigma^\alpha x^{-\alpha} \quad \text{and} \quad \mathbb{P}(Z_0 < -x) \sim C_\alpha \sigma^\alpha x^{-\alpha}, \quad \text{as } x \to \infty,$$

with $C_\alpha = \left(\int_0^\infty x^{-\alpha}\sin(x)dx\right)^{-1} = \dfrac{1-\alpha}{\Gamma(2-\alpha)\cos\left(\frac{\pi\alpha}{2}\right)}$, we proceed the proof.

Observe that $g(Z_0)$ can be rewritten as

$$g(Z_0) = \begin{cases} (\theta+\gamma)Z_0 - c, & \text{if} \quad Z_0 \geq 0, \\ (\theta-\gamma)Z_0 - c, & \text{if} \quad Z_0 < 0, \end{cases}$$

with $c := \gamma\mathbb{E}(|Z_0|)$. Then, for any $x > 0$,

$$\begin{aligned} \mathbb{P}\left(g(Z_0) > x\right) &= \mathbb{P}\left(g(Z_0) > x, Z_0 \geq 0\right) + \mathbb{P}\left(g(Z_0) > x, Z_0 < 0\right) \\ &= \mathbb{P}\left((\theta+\gamma)Z_0 > x + c, Z_0 \geq 0\right) + \mathbb{P}\left((\theta-\gamma)Z_0 > x + c, Z_0 < 0\right) \end{aligned}$$

and

$$\begin{aligned} \mathbb{P}\left(g(Z_0) < -x\right) &= \mathbb{P}\left(g(Z_0) < -x, Z_0 \geq 0\right) + \mathbb{P}\left(g(Z_0) < -x, Z_0 < 0\right) \\ &= \mathbb{P}\left((\theta+\gamma)Z_0 < -x + c, Z_0 \geq 0\right) + \mathbb{P}\left((\theta-\gamma)Z_0 < -x + c, Z_0 < 0\right). \end{aligned}$$

First, assume $\theta = \gamma$. In this case, for all $x > |c|$,

$$\mathbb{P}\left(g(Z_0) > x\right) = \mathbb{P}\left((\theta+\gamma)Z_0 > x + c, Z_0 \geq 0\right) = \begin{cases} \mathbb{P}\left(\mathbb{Z}_0 > \frac{x+c}{2\theta}\right), & \text{if} \quad \theta > 0; \\ 0, & \text{if} \quad \theta < 0; \end{cases}$$

and

$$\mathbb{P}\left(g(Z_0) < -x\right) = \mathbb{P}\left((\theta+\gamma)Z_0 < -x + c, Z_0 \geq 0\right) = \begin{cases} 0, & \text{if} \quad \theta > 0; \\ \mathbb{P}\left(\mathbb{Z}_0 > \frac{-x+c}{2\theta}\right), & \text{if} \quad \theta < 0. \end{cases}$$

Since $(x+c)^{-\alpha} \sim (x-c)^{-\alpha} \sim x^{-\alpha}$, as $x \to \infty$, one concludes that

$$\mathbb{P}(|g(Z_0)| > x) \sim C_\alpha \sigma^\alpha \frac{x^{-\alpha}}{(2|\theta|)^{-\alpha}} = C_\alpha \sigma^\alpha (|\theta+\gamma|^\alpha + |\theta-\gamma|^\alpha) x^{-\alpha}, \quad \text{as } x \to \infty.$$

The proof assuming $\theta = -\gamma$ is analogous.

Now, assume $\theta \neq \pm\gamma$. Observe that, in this case, for all $x > |c|$,

$$\mathbb{P}\left(g(Z_0) > x\right) = \begin{cases} \mathbb{P}\left(Z_0 > \frac{x+c}{\theta+\gamma}\right), & \text{if } \theta+\gamma > 0 \text{ and } \theta-\gamma > 0; \\ \mathbb{P}\left(Z_0 < \frac{x+c}{\theta-\gamma}\right), & \text{if } \theta+\gamma < 0 \text{ and } \theta-\gamma < 0; \\ \mathbb{P}\left(Z_0 > \frac{x+c}{\theta+\gamma}\right) + \mathbb{P}\left(Z_0 < \frac{x+c}{\theta-\gamma}\right), & \text{if } \theta+\gamma > 0 \text{ and } \theta-\gamma < 0; \\ 0, & \text{if } \theta+\gamma < 0 \text{ and } \theta-\gamma > 0; \end{cases}$$

and

$$\mathbb{P}\left(g(Z_0) < -x\right) = \begin{cases} \mathbb{P}\left(Z_0 < \frac{-x+c}{\theta-\gamma}\right), & \text{if } \theta+\gamma > 0 \text{ and } \theta-\gamma > 0; \\ \mathbb{P}\left(Z_0 > \frac{-x+c}{\theta+\gamma}\right), & \text{if } \theta+\gamma < 0 \text{ and } \theta-\gamma < 0; \\ 0, & \text{if } \theta+\gamma > 0 \text{ and } \theta-\gamma < 0; \\ \mathbb{P}\left(Z_0 > \frac{-x+c}{\theta+\gamma}\right) + \mathbb{P}\left(Z_0 < \frac{-x+c}{\theta-\gamma}\right), & \text{if } \theta+\gamma < 0 \text{ and } \theta-\gamma > 0. \end{cases}$$

Thus, upon applying the result $(x+c)^{-\alpha} \sim (x-c)^{-\alpha} \sim x^{-\alpha}$, as $x \to \infty$, one concludes that

$$\mathbb{P}(|g(Z_0)| > x) \sim C_\alpha \sigma^\alpha \left(|\theta+\gamma|^\alpha + |\theta-\gamma|^\alpha\right) x^{-\alpha}, \quad \text{as } x \to \infty.$$

Therefore, by letting $C_{\alpha,\sigma} := C_\alpha \sigma^\alpha$ the result holds, for any $\theta, \gamma \in \mathbb{R}$, with $|\theta| + |\gamma| > 0$. ∎

Theorem 2 provides the necessary and sufficient conditions for the convergence of the sum $\sum_{k=0}^{\infty} \lambda_{d,k} g(Z_{t-1-k})$, for all $t \in \mathbb{Z}$.

Theorem 2. *Let $\{Z_t\}_{t\in\mathbb{Z}}$ be an i.i.d. sequence of $S\alpha S$ random variables, with $\alpha \in (1,2]$, $\{g(Z_t)\}_{t\in\mathbb{Z}}$ be the stochastic process defined by (4) and $\lambda(\cdot)$ be the operator defined by (9). Then, the series*

$$\ln(h_t^2) - \omega = \sum_{k=0}^{\infty} \lambda_{d,k} g(Z_{t-1-k}), \quad \textit{for all } t \in \mathbb{Z}, \tag{13}$$

is well defined and converges almost surely (that is, with probability 1) if and only if $(1-d)\alpha > 1$. Moreover, if $(1-d)\alpha > 1$ then the series (13) converges in $\mathcal{L}^r$ norm, for all $0 < r < \alpha$. Furthermore, if $d \leq 0$, the series (13) converges absolutely with probability 1.

Proof. First assume $(1-d)\alpha > 1$. Observe that, since $1 < \alpha \leq 2$, there exists $1 < \nu_0 < \alpha$ such that $(1-d)\alpha > (1-d)\nu > 1$, for all $\nu_0 \leq \nu < \alpha$. From, theorem 1.2 in Lopes and Prass (2013a), for each $r \in \{0, \cdots, s-1\}$,

$$\lambda_{d,sk+r} = \frac{1}{\Gamma(d) k^{1-d}} \frac{a(1)}{b(1)} + O(k^{d-1}) \mathbb{I}_{\mathbb{N}\setminus\{0,1\}}(s) + O(k^{d-2}), \quad \text{as } k \to \infty,$$

and thus $\sum_{k=0}^{\infty} |\lambda_{d,k}|^\nu < \infty$, for any $\nu_0 \leq \nu < \alpha$. Moreover, from Lemma 1, $\mathbb{E}\left(|g(Z_t)|^\nu\right) < \infty$, for all $t \in \mathbb{Z}$ and $0 < \nu < \alpha$. Therefore, the series (13) converges with probability 1 and in $\mathcal{L}^\nu$ norm, for all $1 < \nu_0 \leq \nu < \alpha$. In particular, the following inequality holds (see Avram and Taqqu, 1986)

$$\mathbb{E}\left(\left|\sum_{k=0}^{\infty} \lambda_{d,k} g(Z_{t-1-k})\right|^\nu\right) \leq 2\mathbb{E}\left(|g(Z_{t-1-k}|^\nu\right) \sum_{k=0}^{\infty} |\lambda_{d,k}|^\nu, \quad \text{for any } 1 < \nu_0 \leq \nu < \alpha.$$

The convergence for $0 < r < \nu_0$ follows immediately from the inequality $\mathbb{E}(|X|^r)^{1/r} \leq \mathbb{E}(|X|^\nu)^{1/\nu}$, for all $0 < r \leq \nu$. It is easy to see that $d \leq 0$ implies $\sum_{k=0}^{\infty} |\lambda_{d,k}| < \infty$ and, consequently, $\mathbb{E}\left(\sum_{k=0}^{\infty} |\lambda_{d,k} g(Z_{t-1-k})|\right) = \mathbb{E}\left(|g(Z_0)|\right) \sum_{k=0}^{\infty} |\lambda_{d,k}| < \infty$. Therefore, for $d \leq 0$, (13) converges absolutely with probability 1.

Now, assume that (13) converges almost surely. Notice that, from the three series theorem (Roussas, 2005, theorem 11, page 380), the almost sure convergence of (13) implies that the series $\sum_{k=0}^{\infty} \mathbb{P}(|\lambda_{d,k} g(Z_{t-1-k})| \geq x)$ necessarily converges, for all $x > 0$. The i.i.d. property of $\{g(Z_t)\}_{t\in\mathbb{Z}}$ implies that

$$\sum_{k=0}^{\infty} \mathbb{P}(|\lambda_{d,k} g(Z_{t-1-k})| \geq x) = \sum_{k=0}^{\infty} \mathbb{P}(|g(Z_0)| \geq x|\lambda_{d,k}|^{-1}), \quad \text{for all } x > 0. \tag{14}$$

Since $\lambda_{d,k} \to 0$, as $k \to \infty$, Theorem 1 implies that

$$\mathbb{P}(|g(Z_0)| \geq x|\lambda_{d,k}|^{-1}) \sim C_{\alpha,\sigma} \left(|\theta+\gamma|^\alpha + |\theta-\gamma|^\alpha\right) x^{-\alpha} |\lambda_{d,k}|^\alpha, \quad \text{as } k \to \infty.$$

From theorem 1.2 in Lopes and Prass (2013a), $\lambda_{d,k} = O(k^{d-1})$, as $k \to \infty$. Therefore, for the convergence of (14) to hold, necessarily, $(1-d)\alpha > 1$. ∎

Theorem 3 gives the conditions for the strict stationarity and ergodicity properties of $\{X_t\}_{t\in\mathbb{Z}}$, $\{h_t\}_{t\in\mathbb{Z}}$ and $\{\ln(h_t^2)-\omega\}_{t\in\mathbb{Z}}$. This theorem also states the condition for the invertibility property of $\{\ln(h_t^2)-\omega\}_{t\in\mathbb{Z}}$.

Theorem 3. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be an α-SFIEGARCH$(p,d,q)_s$, given in Definition 1.*

(i) *If $d<1-\frac{1}{\alpha}$, then $\{X_t\}_{t\in\mathbb{Z}}$, $\{h_t\}_{t\in\mathbb{Z}}$ and $\{\ln(h_t^2)-\omega\}_{t\in\mathbb{Z}}$ are strictly stationary and ergodic.*

(ii) *If $|d|<1-\frac{1}{\alpha}$ and $a(z)\neq 0$ in the closed disk $\{z:|z|\leq 1\}$, then*

$$\lim_{m\to\infty}\mathbb{E}\left(\left|\sum_{k=0}^{m}\tilde{\lambda}_{d,k}(\ln(h_t^2)-\omega)-g(Z_{t-1})\right|^r\right)=0,\quad \textit{for all}\quad 0<r<\alpha,$$

where $\tilde{\lambda}_{d,k}$, for $k\in\mathbb{N}$, are defined through

$$\sum_{k=0}^{\infty}\tilde{\lambda}_{d,k}z^k:=\frac{b(z)}{a(z)}(1-z^s)^d=\lambda^{-1}(z),\quad |z|<1.$$

Proof. (i) Assume $d<1-\frac{1}{\alpha}$. By hypothesis, Z_t is finite with probability 1, for all $t\in\mathbb{Z}$. From Theorem 2, $\ln(h_t^2)-\omega$ is finite with probability 1, and so is h_t, for all $t\in\mathbb{Z}$. Consequently, X_t is finite with probability 1, for all $t\in\mathbb{Z}$. Moreover, $\{Z_t\}_{t\in\mathbb{Z}}$ is an i.i.d. sequence and, from theorem 1.2 in Lopes and Prass (2013a), $\sum_{k=0}^{\infty}|\lambda_{d,k}|^\alpha<\infty$. Therefore, from theorem 3.5.8 in Stout (1974), $\{\ln(h_t^2)-\omega\}_{t\in\mathbb{Z}}$, $\{h_t\}_{t\in\mathbb{Z}}$ and $\{X_t\}_{t\in\mathbb{Z}}$ are strictly stationary and ergodic processes.

(ii) Assume $|d|<1-\frac{1}{\alpha}$ and $a(z)\neq 0$ in the closed disk $\{z:|z|\leq 1\}$. From theorem 1.2 in Lopes and Prass (2013a), if $-1+\frac{1}{\alpha}<d<0$, then $\sum_{k=0}^{\infty}|\lambda_{d,k}|<\infty$ and $\sum_{k=0}^{\infty}|\tilde{\lambda}_{d,k}|^\alpha<\infty$ and the result follows from theorem 2.2 in Kokoszka and Taqqu (1995) with $\Psi(\cdot):=\lambda(\cdot)$ and $C(\cdot):=\lambda^{-1}(\cdot)$. If $0<d<1-\frac{1}{\alpha}$, then theorem 1.2 in Lopes and Prass (2013a) implies $\sum_{k=0}^{\infty}|\lambda_{d,k}|^\alpha<\infty$ and $\sum_{k=0}^{\infty}|\tilde{\lambda}_{d,k}|<\infty$ and the result follows from theorem 2.2 in Kokoszka and Taqqu (1995) with $\Psi(\cdot):=\lambda^{-1}(\cdot)$ and $C(\cdot):=\lambda(\cdot)$. The case $d=0$ is trivial. Therefore, the result follows for all $|d|<1-\frac{1}{\alpha}$. ∎

3 Parameter Estimation: Bayesian Inference using MCMC

Let $\boldsymbol{\eta}$ be the vector of unknown parameters in (8), namely,

$$\boldsymbol{\eta}=(\alpha,d,\theta,\gamma,\omega,a_1,\cdots,a_p,b_1,\cdots,b_q)':=(\eta_1,\eta_2,\cdots,\eta_{5+p+q})'.$$

and $\boldsymbol{\eta}_{(-i)}$ be the vector containing all parameters in $\boldsymbol{\eta}$ except η_i, for each $i\in\{1,\cdots,5+p+q\}$.

Denote by $p_Z(\cdot|\alpha)$ the probability density function of Z_0 given α, for any $\alpha\in(1,2]$. Notice that, by using Zolotarev's integral representation (Zolotarev, 1986), $p_Z(\cdot|\alpha)$ can be written as

$$p_Z(x|\alpha)=\begin{cases}\frac{1}{\pi}\Gamma\left(1+\frac{1}{\alpha}\right), & \text{if}\quad x=0,\\ \frac{\alpha(|x|-\zeta)^{\frac{1}{\alpha-1}}}{\pi|\alpha-1|}\int_0^{\pi/2}H(t,\alpha)\exp\left\{-|x|^{\frac{\alpha}{\alpha-1}}H(t,\alpha)\right\}dt, & \text{if}\quad x\neq 0,\end{cases}\tag{15}$$

where $H(t,\alpha)=\left[\frac{\cos(t)}{\sin(\alpha t)}\right]^{\frac{\alpha}{\alpha-1}}\frac{\cos(\alpha t-t)}{\cos(t)}$, for any $\alpha\in(1,2]$ and $t\in[0,\pi/2]$.

Let $\{X_t\}_{t\in\mathbb{Z}}$ be an α-SFIEGARCH process, defined by (7) and (8). Denote by $p_{X_t}(\cdot|\boldsymbol{\eta},\mathcal{F}_{t-1})$ the conditional probability density function of X_t, given $\boldsymbol{\eta}$ and $\mathcal{F}_{t-1}$, for all $t\in\mathbb{Z}$. Then, from the $\mathcal{F}_{t-1}$ measurability of h_t, for all $t\in\mathbb{Z}$, and the i.i.d. property of $\{Z_t\}_{t\in\mathbb{Z}}$, it follows that

$$p_{X_t}(x_t|\boldsymbol{\eta},\mathcal{F}_{t-1}) = \frac{1}{h_t} p_Z\left(x_t h_t^{-1}|\alpha\right), \quad \text{with} \quad h_t = \exp\left\{\frac{1}{2}\left[\omega + \sum_{k=0}^{\infty}\lambda_{d,k} g(z_{t-1-k})\right]\right\}, \tag{16}$$

for all $x_t\in\mathbb{R}$ and $t\in\mathbb{Z}$, with $p_Z(\cdot|\nu)$ given in (15). Furthermore, from (16), the conditional probability of $\boldsymbol{X} := (X_1,\cdots,X_n)'$ given $\boldsymbol{\eta}$ and $I_0\in\mathcal{F}_0$ can be written as

$$p_{\boldsymbol{X}}(x_1,\cdots,x_n|\boldsymbol{\eta},I_0) = p_{X_n}(x_n|\boldsymbol{\eta},I_{n-1})\times\cdots\times p_{X_1}(x_1|\boldsymbol{\eta},I_0) = \prod_{t=1}^{n}\frac{1}{h_t}p_Z\left(x_t h_t^{-1}|\alpha\right), \tag{17}$$

for all $(x_1,\cdots,x_n)'\in\mathbb{R}^n$, where $I_k\in\mathcal{F}_k$ is obtained recursively through $I_k = \{z_k\}\cup I_{k-1}$, with $z_k = x_k h_k^{-1}$, for all $k>0$.

Hence, upon applying the Bayes' rule, the conditional probability density function of η_i given $\boldsymbol{X}$, $\boldsymbol{\eta}_{(-i)}$ and any $I_0\in\mathcal{F}_0$, can be written as

$$p\left(\eta_i|\boldsymbol{X},\boldsymbol{\eta}_{(-i)},I_0\right) = \frac{p_{\boldsymbol{X}}(\boldsymbol{X}|\boldsymbol{\eta},I_0)\times p_{I_0}(I_0|\boldsymbol{\eta})\times p_{(-i)}(\boldsymbol{\eta}_{(-i)}|\eta_i)\times\pi_i(\eta_i)}{p_{(-i)}\left(\boldsymbol{X},\boldsymbol{\eta}_{(-i)},I_0\right)}, \tag{18}$$

for each $i\in\{1,\cdots,5+p+q\}$, where $p_{\boldsymbol{X}}(\cdot|\boldsymbol{\eta},\mathcal{F}_0)$ is given in (17); $p_{I_0}(\cdot|\boldsymbol{\eta})$ is the prior conditional probability density function for I_0 given $\boldsymbol{\eta}$; $p_{(-i)}(\cdot|\eta_i)$ is the prior conditional probability density function for $\boldsymbol{\eta}_{(-i)}$ given η_i; $\pi_i(\cdot)$ is the prior[1] density function for η_i and $p_{(-i)}(\cdot,\cdot,\cdot)$ is the joint probability density function of $\boldsymbol{X}$, $\boldsymbol{\eta}_{(-i)}$ and I_0, which does not depend on η_i.

The parameter estimation is then carried out by using the MCMC method as described below.

3.1 Gibbs Sampling with Metropolis Steps

Gibbs sampling (Geman and Geman, 1984; Gelfand and Smith, 1990) is one of the most popular MCMC algorithm used to sample from multivariate probability distribution when direct sampling is difficult or even not possible. The algorithm was developed based on the assumption that each random variable has a known conditional probability distribution which can be easily sampled from. The steps of the sampling procedure are the following.

Step 1. Let $\boldsymbol{X} = (X_1,\cdots,X_n)'$ be the vector of the observed time series values and I_0 be the set of initial conditions. Set $m=0$ and select an arbitrary initial value for the vector of parameters $\boldsymbol{\eta}$, namely, $\boldsymbol{\eta}^{(0)} = (\eta_1^{(0)},\cdots,\eta_{5+p+q}^{(0)})'$.

Step 2. Given the sample $\boldsymbol{\eta}^{(m)} = (\eta_1^{(m)},\cdots,\eta_{5+p+q}^{(m)})'$,

- generate $\eta_1^{(m+1)}$ from $p\left(\eta_1|\boldsymbol{X},\eta_2^{(m)},\eta_3^{(m)},\cdots,\eta_{5+p+q}^{(m)},I_0\right)$;
- generate $\eta_2^{(m+1)}$ from $p\left(\eta_2|\boldsymbol{X},\eta_1^{(m+1)},\eta_3^{(m)},\cdots,\eta_{5+p+q}^{(m)},I_0\right)$;

$$\vdots \qquad \vdots \qquad \vdots$$

- generate $\eta_{5+p+q}^{(m+1)}$ from $p\left(\eta_{5+p+q}|\boldsymbol{X},\eta_1^{(m+1)},\cdots\eta_{4+p+q}^{(m+1)},I_0\right)$;

[1] The priors $\pi_i(\cdot)$ are not necessarily probability density functions. For instance, $\pi(x)=1$ and $\pi(x)=1/x$, are examples of improper priors (i.e., they do not integrate to 1) used in practice.

Step 3. Once the vector $\boldsymbol{\eta}^{(m+1)} = (\eta_1^{(m+1)}, \cdots, \eta_{5+p+q}^{(m+1)})'$ is obtained, return to **Step 2**, with $m = m + 1$, until $m = N$, where N is the desired sample size.

It can be shown that, under some regularity conditions (Gelfand and Smith, 1990; Tierney, 1994), the Markov chain obtained by following the steps just described converges to a stationary distribution, as N goes to infinity. Moreover, the limit distribution is the posterior joint distribution of $\boldsymbol{\eta}$ given $\boldsymbol{X}$ and I_0.

Unfortunately, in practice, sampling directly from $p\left(\eta_i|\boldsymbol{X}, \boldsymbol{\eta}_{(-i)}, I_0\right)$ is not always possible, for one or more $i \in \{1, \cdots, 5 + p + q\}$. Under this scenario, a common approach is to consider a combination of Gibbs sampler and Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970) algorithms. In the literature, this method is referred to as *Gibbs sampler with Metropolis steps*. The procedure is very similar to Gibbs sampler itself. The difference is that in **Steps 1-3** described above, instead of sampling directly from $p\left(\eta_i|\boldsymbol{X}, \boldsymbol{\eta}_{(-i)}, \mathcal{F}_0\right)$, one shall consider the Metropolis-Hastings algorithm with $p\left(\eta_i|\boldsymbol{X}, \boldsymbol{\eta}_{(-i)}, \mathcal{F}_0\right)$ as the invariant (target) distribution.

The steps for the Metropolis-Hastings sampling procedure are described below. Since the algorithm does not require knowing the normalization constant $p_{(-i)}\left(\boldsymbol{X}, \boldsymbol{\eta}_{(-i)}, I_0\right)$, given in (18), we shall consider instead $p_*(\cdot)$, which denotes any non-normalized probability density function corresponding to $p\left(\eta_i|\boldsymbol{X}, \boldsymbol{\eta}_{(-i)}, \mathcal{F}_0\right)$, for $i \in \{1, \cdots, 5 + p + q\}$.

Step 1. Select a transition kernel[2] (also called the proposal distribution) $q(\cdot|\cdot)$ for which the sampling procedure is known.

Step 2. Let $m = 0$ and set an arbitrary initial value y_0 for the chain.

Step 3. Generate a draw ξ from $q(\cdot|y_m)$.

Step 4. Calculate $\tau(y_m, \xi) = \min\left\{1, \dfrac{p_*(\xi)q(y_m|\xi)}{p_*(y_m)q(\xi|y_m)}\right\}$.

Step 5. Draw $u \sim \mathcal{U}[0, 1]$.

Step 6. Define $y_{m+1} = \begin{cases} \xi, & \text{if } u < \tau(y_m, \xi); \\ y_m, & \text{otherwise.} \end{cases}$

Step 7. If $m + 1 < N$ (where N is the desired sample size), let $m = m + 1$ and go to **Step 3**.

To remove the influence of the initial point $\boldsymbol{\eta}^{(0)}$ and to assure the chain convergence, it is advised to discard the first B (for some $B < N$) observations (that is, the burn-in sample). As it should be expected, the value of B varies for each problem.

The sample obtained by considering Gibbs sampler with Metropolis steps is not independent. On the other hand, due to the ergodicity property of the Markov chain, parameter estimation is not necessarily affected by the correlation in the sample. If necessary, the autocorrelation can be reduced by thinning the Markov chain, that is, by keeping only every k-th simulated draw from each sequence, or by running parallel chains. However, in the literature there is no common sense whether it is better to run one long chain instead of several shorter ones (Gelman and Rubin, 1992; Geyer, 1992). Also, MacEachern and Berliner (1994) show that using the entire chain lead to more precise posterior estimates than using the thinned one.

[2] A transition kernel is a function $q(x|y)$ which is a probability measure with respect to x, so $\int_{\mathbb{R}} q(x|y)dx = 1$.

4 Simulation Study

For this simulation study we assume $p = q = 0$. Under this scenario, the vector of unknown parameters is $\boldsymbol{\eta} = (\alpha, d, \theta, \gamma, \omega)' := (\eta_1, \cdots, \eta_5)'$. The Bayesian inference approach, using MCMC to obtain posterior density functions, is used to estimate the model parameters.

4.1 Data Generating Process

The samples from α-SFIEGARCH$(0, d, 0)_s$ processes, for $s \in \{1, 7\}$, are obtained by setting the following.

- We consider all combinations of $d \in \{0.10, 0.25, 0.35, 0.45\}$ and $\alpha \in \{1.25, 1.45, 1.65, 1.85, 2.00\}$ satisfying $(1 - d)\alpha > 1$.
- For any d and α we set $\omega = -5.40$, $\theta = -0.15$ and $\gamma = 0.24$. These values are close to the ones already observed in practical applications considering FIEGARCH/SFIEGARCH models (see, for instance, Bollerslev and Mikkelsen, 1996; Ruiz and Veiga, 2008; Lopes and Prass, 2013a,b).
- For all models the infinite sum in (11) is truncated at $m^* \in \{50000, 100000\}$, respectively, when $s \in \{1, 7\}$. The coefficients $\{\lambda_{d,k}\}_{k=1}^{m^*}$ are calculated through the recurrence formula given in Lopes and Prass (2013b), proposition 1.1.

For each combination of s, d and α, a sample $\{z_t\}_{t=-m^*}^{n}$, of size $m^* + n + 1$, with $n = 10000$, is drawn from the $S\alpha S$ distribution, with scaling parameter $\sigma = 1$. Then the sample $\{x_t\}_{t=1}^{n}$, from the α-SFIEGARCH$(0, d, 0)_s$ process, is obtained through the relation

$$\ln(h_t^2) = \omega + \sum_{k=0}^{m^*} \lambda_{d,k} g(z_{t-1-k}) \quad \text{and} \quad x_t = h_t z_t, \quad \text{for all } t \in \{1, \cdots, n\}.$$

4.2 Parameter Estimation Setting

The samples from the posterior distributions are obtained by considering the Gibbs sampler algorithm with Metropolis steps as described in Section 3. The transition kernel $q(\cdot|\cdot)$ considered in the Metropolis-Hastings algorithm is the function defined as

$$q(x|y) = f(x; y, \mathfrak{s}, \mathfrak{a}, \mathfrak{b}),$$

where $f(\cdot; \cdot, \cdot, \cdot, \cdot)$ is the truncated normal density function, given by

$$f(x; \mathfrak{m}, \mathfrak{s}, \mathfrak{a}, \mathfrak{b}) = \begin{cases} \frac{1}{\mathfrak{s}} \frac{\phi\left(\frac{x-\mathfrak{m}}{\mathfrak{s}}\right)}{\Phi\left(\frac{\mathfrak{b}-\mathfrak{m}}{\mathfrak{s}}\right) - \Phi\left(\frac{\mathfrak{a}-\mathfrak{m}}{\mathfrak{s}}\right)}, & \text{if } \mathfrak{a} \leq x \leq \mathfrak{b}, \\ 0, & \text{otherwise,} \end{cases}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are, respectively, the probability density and cumulative distribution functions of the standard normal distribution; $\mathfrak{a}, \mathfrak{b} \in \mathbb{R}$ are, respectively, the lower and upper limits of the distribution's support; $\mathfrak{m}$ and $\mathfrak{s}$ denote, respectively, the distribution's (non-truncated version) mean and standard deviation.

Table 1 presents parameters for the transition kernel used in this simulation study to obtain the sample from the posterior distribution of the parameter η_i, for each $i \in \{1, \cdots, 5\}$. In this

Table 1: Parameters of the truncated normal distribution (transition kernel) associated to parameter η_i in $\boldsymbol{\eta} = (\alpha, d, \theta, \gamma, \omega)' := (\eta_1, \cdots, \eta_5)'$, for each $i \in \{1, \cdots, 5\}$, at m-th iteration of the Gibbs sampler.

Parameter	α	d	θ	γ	ω
Mean (y)	$\alpha^{(m-1)}$	$d^{(m-1)}$	$\theta^{(m-1)}$	$\gamma^{(m-1)}$	$\omega^{(m-1)}$
Standard Deviation ($\mathfrak{s}$)	0.05	0.05	0.05	0.05	1.50
Lower Limit ($\mathfrak{a}$)	1.00	−0.50	−1.00	0.00	−15.00
Upper Limit ($\mathfrak{b}$)	2.00	0.50	0.00	1.00	15.00

Note: $\eta_i^{(m-1)}$, for any $i \in \{1, \cdots, 5\}$, denotes the parameter value obtained in the $(m-1)$-th iteration.

table, m denotes the iteration number of the Gibbs sampler with Metropolis steps and $\eta_i^{(m-1)}$, for any $i \in \{1, \cdots, 5\}$, denotes the parameter value obtained in the $(m-1)$-th iteration.

To select $\boldsymbol{\eta}^{(0)}$, we calculate $p_{\boldsymbol{X}}(\boldsymbol{X}|\boldsymbol{\eta}, \mathcal{F}_0)$ for different combinations of $\alpha, d, \theta, \gamma$ and ω, always assuming s known. Then $\boldsymbol{\eta}^{(0)}$ is defined as the vector $\boldsymbol{\eta} = (\alpha, d, \theta, \gamma, \omega)'$ with higher likelihood function value. In all cases a single chain is obtained, thinning is not applied and a burn-in of size 1000 is considered. An example showing the influence of using a large chain (with size 200801), the thinned sub-sample or only the first 1000 observations of the entire chain (after burn-in) for parameter estimation in FIEGARCH$(0, d, 0)$ models can be found in Prass et al. (2013).

As initial conditions I_0, we set $g(Z_k) = 0$, for all $k \leq 0$ and $p_{I_0}(\cdot|\eta) \propto 1$. The prior distributions for $\alpha, d, \theta, \gamma$ and ω are selected by considering only the basic set of information usually available in practice as follows.

- $\alpha \sim \mathcal{U}(1, 2)$. This choice is based on the fact that, when $0 < \alpha \leq 1$, the stable distribution does not have finite first moment and α-SFIEGARCH processes are not well defined.
- $d \sim \mathcal{U}(-0.5, 0.5)$. From Section 2 it is known that, if $|d| < 1 - \frac{1}{\alpha}$, $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$ is a strictly stationary and invertible process. The wider range for d is obtained when $\alpha = 2$.
- $\theta \sim \mathcal{U}(-1, 0)$ and $\gamma \sim \mathcal{U}(0, 1)$. To the best of our knowledge, an SFIEGARCH, a FIEGARCH or an EGARCH model for which θ or γ are not in the intervals, respectively, $[-1, 0]$ and $[0, 1]$ has never been reported in the literature.
- $\omega \sim \mathcal{U}(-15, 15)$. The choice is based on the fact that $\omega = \mathbb{E}(\ln(h_t^2)) = \mathbb{E}(\ln(X_t^2)) + \mathbb{E}(\ln(Z_t^2))$. In practice, the interval $[-15, 15]$ should be adjusted according to the magnitude of the data.

The corresponding conditional priors for $\boldsymbol{\eta}_{(-i)}$ given η_i, for $i \in \{1, \cdots, 5\}$, are chosen as follows

- $p_{(-1)}(d, \theta, \gamma, \omega|\alpha) \propto p(d|\alpha) \prod_{j\notin\{1,2\}} \pi_j(\eta_j)$, with $p(d|\alpha) := \begin{cases} 1, & \text{if } \ |d| < 1 - \frac{1}{\alpha}; \\ 0, & \text{otherwise.} \end{cases}$
- $p_{(-2)}(\alpha, \theta, \gamma, \omega|d) \propto p(\alpha|d) \prod_{j\notin\{1,2\}} \pi_j(\eta_j)$, with $p(\alpha|d) := \begin{cases} 1, & \text{if } \ \frac{1}{(1-|d|)} < \alpha \leq 2; \\ 0, & \text{otherwise.} \end{cases}$
- $p_{(-i)}(\boldsymbol{\eta}_{(-i)}|\eta_i) \propto \prod_{j\neq i} \pi_j(\eta_j), \quad$ for any $i \in \{3, 4, 5\}$.

The estimation procedure is performed by considering the entire sample $\{x_t\}_{t=1}^{10000}$ and sub-samples of size $n = 2000$ and 4000 from this time series. The sub-samples are taken from the end of the original time series. In practice this would correspond to more up to date information. The aim of this analysis is to determine whether or not increasing the sample size of the observed time series improves parameter estimation.

4.3 Estimates and Performance Measures

Let $\{\eta_i^{(k)}\}_{k=1}^M$ be a sample of size M from the posterior distribution of η_i in $\boldsymbol{\eta} = (\alpha, d, \theta, \gamma, \omega)' := (\eta_1, \cdots, \eta_5)'$, for any $i \in \{1, \cdots, 5\}$. Denote by $\bar{\eta}_i$ and sd_{η_i}, respectively, the sample mean and standard deviation of $\{\eta_i^{(k)}\}_{k=1}^M$, namely,

$$\bar{\eta}_i = \frac{1}{M}\sum_{k=1}^{M}\eta_i^{(k)} \quad \text{and} \quad \text{sd}_{\eta_i} = \sqrt{\frac{1}{M}\sum_{k=1}^{M}(\eta_i^{(k)} - \bar{\eta}_i)^2}, \quad \text{for any} \quad i \in \{1, \cdots, 5\}.$$

Then the estimate $\hat{\eta}_i$ of η_i is defined as $\hat{\eta}_i := \bar{\eta}_i$.

Moreover, let $\hat{q}_i(\mathfrak{p})$ denote the quantile of order $\mathfrak{p}$ [3] for the posterior sample distribution of η_i, for any $\mathfrak{p} \in [0, 1]$ and $i \in \{1, \cdots, 5\}$. Then a $100(1-\mathfrak{p})\%$ credibility interval for η_i is given by

$$CI_{1-\mathfrak{p}}(\eta_i) = \left[\hat{q}_i\left(\frac{\mathfrak{p}}{2}\right), \hat{q}_i\left(1 - \frac{\mathfrak{p}}{2}\right)\right], \quad \text{for any} \quad i \in \{1, \cdots, 5\}.$$

Furthermore, the estimation bias and the absolute percentage error (ape) of estimation are given, respectively, by

$$\text{bias}_{\eta_i} = \bar{\eta}_i - \eta_i \quad \text{and} \quad \text{ape}_{\eta_i} = \left|\frac{\text{bias}_{\eta_i}}{\eta_i}\right|, \quad \text{for any} \quad i \in \{1, \cdots, 5\}.$$

4.4 Results

Table 2: TABLE FOR $s = 1$ WILL BE INCLUDED HERE.

Table 3: TABLE FOR $s = 7$ WILL BE INCLUDED HERE.

5 Empirical Applications

In this section we consider Bayesian inference using MCMC to estimate the parameters of α-SFIEGARCH models for two observed time series. The first data set corresponds to the São Paulo stock exchange index (also known as Bovespa index or IBovespa) log-return time series, observed in the period from January 02, 1995 to December 10, 2001 (1717 observations). The second data set is the one-hour log-return time series corresponding to the S&P500 stock exchange index observed in the period from March 21, 2007 to August 13, 2009 (4232 observations). The complete analysis of these time series, including model identification, parameter estimation through QMLE procedure, forecasting and comparison to other ARCH-type models can be found, respectively, in Lopes and Prass (2013a) and Lopes and Prass (2013b).

5.1 São Paulo Stock Exchange Index

Figure 1 shows the IBovespa log-return time series. In this work, this time series shall be denoted by $\{r_t^I\}_{t=1}^{1717}$. The study conducted in Lopes and Prass (2013a) show that the time series $\{r_t^I\}_{t=1}^{1717}$

[3] In this work, the following definition is adopted (Brockwell and Davis, 1991). Given any $0 \leq \mathfrak{p} \leq 1$, the number $q(\mathfrak{p})$ satisfying $\mathbb{P}(X \leq q(\mathfrak{p})) \geq \mathfrak{p}$ and $\mathbb{P}(X \geq q(\mathfrak{p})) \geq 1 - \mathfrak{p}$, is called a quantile of order $\mathfrak{p}$ for the random variable X (or for the distribution function of X).

presents small correlation, which can be easily removed by considering an ARMA$(0,1)$ model. The authors also show that the volatility for this time series presents non-seasonal long-range dependence.

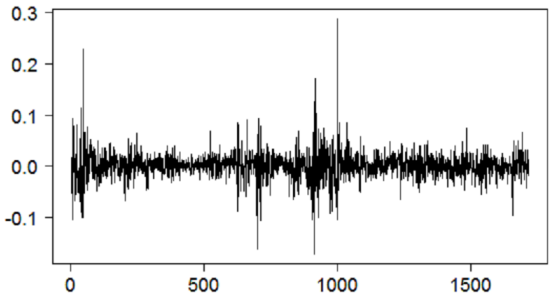


Figure 1: São Paulo stock exchange index log-return time series, observed in the period from January 02, 1995 to December 10, 2001 (1717 observations).

In Lopes and Prass (2013a), the ARMA$(0,1)$ model fitted to the IBovespa log-return time series is given by (the numbers in parenthesis correspond to the standard errors)

$$r_t^I = X_t^I + \underset{(0.0257)}{0.0776} X_{t-1}^I, \quad \text{for all } t \in \{1, \cdots, 1717\},$$

with $X_t^I = 0$, if $t \leq 0$. In this work these values shall be assumed fixed and only the parameters for the α-SFIEGARCH models corresponding to $\{X_t^I\}_{t=1}^{1717}$ are estimated.

Since Section 4 only considers the case $p = q = 0$, we conduct a brief simulation study to select the priors and kernels parameters for the observed time series. In this study we consider α-SFIEGARCH$(0, d, 1)_1$ models, assuming as true parameter values the QMLE estimates obtained in Lopes and Prass (2013a), for different values of α. The priors for $\alpha, d, \theta, \gamma, \omega$ and b_1 are taken to be uniform. The lower and upper limits for the uniform distributions corresponding to the model's parameters are given in Table 4.

Table 4: Lower and upper limits for the prior uniform distribution associated to each parameter in the α-SFIEGARCH$(0, d, 1)_1$ model.

Parameter	α	d	θ	γ	ω	b_1
Lower Limit	1.89	0.30	-0.20	0.25	-8.00	0.60
Upper Limit	2.00	0.40	-0.10	0.35	-6.50	0.75

Moreover, the conditional priors $p_{(-i)}(\boldsymbol{\eta}_{(-i)}|\eta_i)$ were defined as follows

- given $\eta_1 = \alpha$: analogous to Section 4, that is,

$$p_{(-1)}(d, \theta, \gamma, \omega, b_1|\alpha) \propto \begin{cases} \prod_{j \notin \{1,2\}} \pi_j(\eta_j), & \text{if } \ |d| < 1 - \frac{1}{\alpha}; \\ 0, & \text{otherwise;} \end{cases}$$

- given $\eta_2 = d$: analogous to Section 4, that is,

$$p_{(-2)}(\alpha, \theta, \gamma, \omega, b_1|d) \propto \begin{cases} \prod_{j \notin \{1,2\}} \pi_j(\eta_j), & \text{if } \ \frac{1}{(1-|d|)} < \alpha \leq 2; \\ 0, & \text{otherwise;} \end{cases}$$

- given η_i, with $\eta_i \in \{\theta, \gamma\}$: analogous to Section 4, that is,

$$p_{(-i)}(\boldsymbol{\eta}_{(-i)}|\eta_i) \propto \prod_{j \neq i} \pi_j(\eta_j);$$

- given $\eta_i = b_1$: we set

$$p_{(-i)}(\boldsymbol{\eta}_{(-i)}|b_1) \propto \begin{cases} 1, & \text{if} \quad |b_1| < 1 \text{ when } |z| \leq 1; \\ 0, & \text{otherwise.} \end{cases}$$

For each parameter in the α-SFIEGARCH$(0, d, 1)_1$ models, the lower ($\mathfrak{a}$) and upper ($\mathfrak{b}$) limits for the kernels were assumed to be the same as the corresponding priors. The kernels' means (y) were defined analogously to Table 1. In all cases the kernels' standard deviations ($\mathfrak{s}$) were set as 0.005. The sample size of the simulated α-SFIEGARCH time series is $n = 2000$. The posterior means are obtained upon considering a sample of size $N = 1000$, from the posterior distribution, after a burn-in of size 1000.

The simulation results are reported in Table 5. The values reported in these tables are the posterior mean, standard deviation and the 95% credibility interval, for each parameter in the model. Table 5 indicates that, for the considered α-SFIEGARCH$(0, d, 1)_1$ models, all parameters, except ω, are well estimated. In all cases the parameter ω is always underestimated. Moreover, as α increases the estimation bias for ω decreases.

Table 5: Summary statistics for the sample from the posterior distribution of η_i, for each η_i in $\boldsymbol{\eta} = (\nu, d, \theta, \gamma, \omega, b_1)'$. The reported values are the true parameter value (η), the sample mean ($\bar{\eta}$), the sample standard deviation (sd_η) and the lower ($q_{(0.025)}$) and upper ($q_{(0.975)}$) limits for the 95% credibility interval.

Parameter	α	d	θ	γ	ω	b_1
η	1.9000	0.3578	-0.1661	0.2972	-7.2247	0.6860
$\bar{\eta}$	1.9168	0.3477	-0.1807	0.3003	-6.9497	0.6775
sd_η	0.0167	0.0286	0.0120	0.0211	0.0519	0.0356
$q_{(0.025)}$	1.8910	0.3040	-0.1982	0.2595	-7.0543	0.6112
$q_{(0.975)}$	1.9482	0.3986	-0.1557	0.3429	-6.8326	0.7373
η	1.9500	0.3578	-0.1661	0.2972	-7.2247	0.6860
$\bar{\eta}$	1.9646	0.3389	-0.1812	0.3034	-6.9706	0.6895
sd_η	0.0154	0.0267	0.0122	0.0215	0.0545	0.0325
$q_{(0.025)}$	1.9307	0.3025	-0.1989	0.2655	-7.0789	0.6244
$q_{(0.975)}$	1.9873	0.3948	-0.1558	0.3426	-6.8520	0.7408
η	2.0000	0.3578	-0.1661	0.2972	-7.2247	0.6860
$\bar{\eta}$	1.9918	0.3435	-0.1756	0.3138	-7.0315	0.6875
sd_η	0.0073	0.0276	0.0144	0.0221	0.0551	0.0305
$q_{(0.025)}$	1.9718	0.3039	-0.1978	0.2726	-7.1426	0.6297
$q_{(0.975)}$	1.9997	0.3967	-0.1456	0.3477	-6.9225	0.7350

The results on the parameter estimation for the IBovespa log-returns time series are reported in Table 6

5.2 S&P500 Stock Exchange Index

Figure 2 shows the S&P500 log-return time series, denoted by $\{r_t^S\}_{t=1}^{4232}$. The study conducted in Lopes and Prass (2013b) show that the time series $\{r_t^S\}_{t=1}^{4232}$ presents small correlation and a seasonal (short-memory) behavior, which can be easily removed by considering a constrained

Table 6: TABLE WILL BE INCLUDED HERE.

ARMA$(0, 263)$ model. Regarding the volatility behavior, the authors shows that the S&P500 one-hour log-return time series shows seasonal long-range dependence with seasonal parameter $s = 7$ (one day cycle).

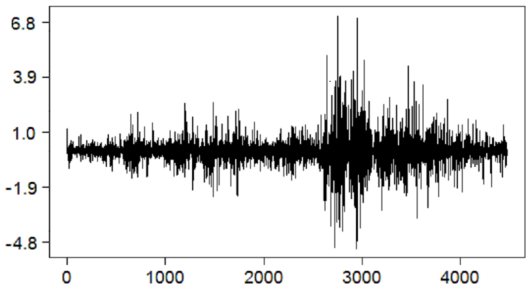


Figure 2: One-hour log-return time series corresponding to the S&P500 stock exchange index observed in the period from March 21, 2007 to August 13, 2009 (4232 observations).

In Lopes and Prass (2013b), the constrained ARMA model fitted to the S&P500 one-hour log-return time series is given by

$$r_t^S = X_t^S \underset{(0.0128)}{- 0.0391} X_{t-7}^S \underset{(0.0169)}{- 0.0834} X_{t-13}^S \underset{(0.0204)}{+ 0.0803} X_{t-53}^S \underset{(0.0215)}{+ 0.0748} X_{t-109}^S \underset{(0.0093)}{- 0.0433} X_{t-190}^S \\ \underset{(0.0221)}{+ 0.0742} X_{t-203}^S \underset{(0.0167)}{+ 0.0680} X_{t-218}^S \underset{(0.0111)}{- 0.0421} X_{t-263}^S,$$

for all $t \in \{1, \cdots, 4232\}$, with $X_t^S = 0$, if $t \leq 0$. In this work these values shall be assumed fixed and only the parameters for the α-SFIEGARCH models corresponding to $\{X_t^S\}_{t=1}^{4232}$ are estimated.

Analogous to Section 5.1, we conduct a brief simulation study to select the priors and kernels parameters for the observed time series. In this study we consider α-SFIEGARCH$(6, d, 6)_7$ models, assuming as true parameter values the QMLE estimates obtained in Lopes and Prass (2013b), for different values of α. The priors for $\alpha, d, \theta, \gamma$, ω, a_j and b_j, for any $j \in \{1, \cdots, 6\}$, are taken to be uniform. The lower and upper limits for the uniform distribution associated to η_i are set, respectively, as $\eta_i - 0.10$ and $\eta_i + 0.10$, for each η_i in $\boldsymbol{\eta} = (\nu, d, \theta, \gamma, \omega, a_1, \cdots, a_p, b_1, \cdots, b_q)'$.

Moreover, the conditional priors $p_{(-i)}(\boldsymbol{\eta}_{(-i)}|\eta_i)$ were defined as follows

- given $\eta_1 \in \{\alpha, d, \theta, \gamma, \omega\}$: analogous to Section 5.1;
- given η_i, with $\eta_i \notin \{a_1, \cdots, a_p\}$: analogous to Section 4, that is,

$$p_{(-i)}(\boldsymbol{\eta}_{(-i)}|\eta_i) \propto \prod_{j \neq i} \pi_j(\eta_j);$$

- given η_i, with $\eta_i \in \{b_1, \cdots, b_q\}$: we set

$$p_{(-i)}(\boldsymbol{\eta}_{(-i)}|\eta_i) \propto \begin{cases} 1, & \text{if} \quad 1 - \sum_{k=1}^{q} b_k z^k \neq 0 \text{ when } |z| \leq 1; \\ 0, & \text{otherwise.} \end{cases}$$

For each parameter in the α-SFIEGARCH$(6, d, 6)_7$ models, the lower ($\mathfrak{a}$) and upper ($\mathfrak{b}$) limits for the kernels were assumed to be the same as the corresponding priors. The kernels' means (y) were defined analogously to Table 1. In all cases the kernels' standard deviations ($\mathfrak{s}$) were set as 0.005. The sample size of the simulated α-SFIEGARCH time series is $n = 4000$. The posterior means are obtained upon considering a sample of size $N = 1000$, from the posterior distribution, after a burn-in of size 1000.

Table 7: Summary statistics for the sample from the posterior distribution of η_i, for each η_i in $\boldsymbol{\eta} = (\nu, d, \theta, \gamma, \omega, a_1, \cdots, a_6, b_1, \cdots, b_6)'$. The reported values are the true parameter value (η), the sample mean ($\bar{\eta}$), the sample standard deviation (sd_η) and the lower ($q_{(0.025)}$) and upper ($q_{(0.975)}$) limits for the 95% credibility interval.

RESULTS WILL BE INCLUDED HERE.

Table 8: TABLE WILL BE INCLUDED HERE.

The simulation results are reported in Table 7. The values reported in these tables are the posterior mean, standard deviation and the 95% credibility interval, for each parameter in the model.

The results on the parameter estimation for the S&P500 log-returns time series are reported in Table 8

6 Conclusions

Acknowledgments

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

References

Avram F. and M.S. Taqqu (1986). "Weak Convergence of Moving Averages with Infinite Variance". *In: Dependence in Probability and Statistics*, E. Eberlein and M.S. Taqqu (Editors.), 399-415. Boston: Birkhäuser.

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

Bordignon, S.; M. Caporin and F. Lisi (2007). "Generalized long memory GARCH for intradaily volatility modelling". *Computational Statistics & Data Analysis*, vol. 51, 5900-5912.

Bordignon, S.; M. Caporin and F. Lisi (2009). "Periodic Long-Memory GARCH models". *Econometric Reviews*, vol. 28(1), 60-82.

Brockwell, P.J. and R.A. Davis (1991). *Time Series: Theory and Methods*. Second Edition. New York: Springer-Verlag.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, vol. 2. New York: Wiley.

Gelfand, A.E. and A.F.M Smith (1990). "Sampling-Based Approaches to Calculating Marginal Densities". *Journal of the American Statistical Association*, vol. 85(410), 398-409.

Gelman, A. and D. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, vol. 7, 457-511.

Geman, S. and D. Geman (1984). "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, 609-628.

Geyer, C. J. (1992). Practical Markov Chain Monte Carlo. *Statistical Science*, vol. 7, 473-511.

Hastings, W.K. (1970). "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". *Biometrika*, vol. 57(1), 97-109.

Kokoszka, P.S. and M.S. Taqqu (1995). "Fractional ARIMA with stable Innovations". *Stochastic Processes and their Applications*, vol. **60**, 19-47.

Kozubowski, T.J.; A.K. Panorska and S.T. Rachev (2003). "Statistical Issues in Modeling Multivariate Stable Portfolios". In Rachev, S.T. (editor), *Handbook of Heavy Tails Distributions in Finance*. Amsterdam: Elsevier, 131-162.

Laha, R.G. and V.K Rohatgi (1979). *Probability Theory*. New York: Wiley.

Lévy, P. (1925). *Calcul des Probabilités*. Paris: Gauthier-Villars.

Lopes, S.R.C and T.S. Prass (2013a). "Theoretical Results on Fractionally Integrated Exponential Generalized Autoregressive Conditional Heteroskedastic Processes". Under Review.

Lopes, S.R.C and T.S. Prass (2013b). "Seasonal FIEGARCH Processes". To Appear in: *Computational Statistics and Data Analysis*.

MacEachern, S.N. and L.M. Berliner (1994). "Subsampling the Gibbs Sampler". *The American Statistician*, vol. 48, 188-190.

Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. (1953). "Equations of State Calculations by Fast Computing Machines". *Journal of Chemical Physics*, vol. 21(6), 1087-1092.

Nelson, D.B. (1991). "Conditional Heteroskedasticity in Asset Returns: A New Approach". *Econometrica*, vol. 59, 347-370.

Panorska, A.K.; S. Mittnik and A. Rachev (1995). "Stable GARCH models for financial time series". *Applied Mathematics Letters*, vol. 8(5), 33-37.

Prass, T.S; S.R.C. Lopes and J.A. Achcar (2013). "MCMC Bayesian Estimation in FIEGARCH Models". Under Review.

Roussas, G.G. (2005). *Measure-Theoretic Probability*. San Diego: Elsevier.

Ruiz, E. and H. Veiga (2008). "Modelling long-memory volatilities with leverage effect: A-LMSV versus FIEGARCH". *Computational Statistics and Data Analysis*, vol. 52(6), 2846-2862.

Samorodnitsky, G. and M.S. Taqqu (1994). *Stable Non-Gaussian Random Processes. Stochastic Models with Infinite Variance*. New York: Chapman & Hall.

Stout, W.F. (1974). *Almost Sure Convergence*. New York: Academic Press.

Tierney, L. (1994). "Markov Chains for Exploring Posterior Distributions". *Annals of Statistics*, vol. 22(4), 1701-1786.

Zolotarev, V.M. (1986). *One Dimensional Stable Distribution*, vol 65 of "Translations of Mathematical Monographs". American Mathematical Society.

Apêndice J

Estimador Whittle Modificado: Processos α-SFIEGARCH$(0, d, 0)_s$, com $s = 1$

Neste apêndice apresentamos os resultados de uma simulação de Monte Carlo realizada com o objetivo de analisar o comportamento do estimador de Whittle para os processos FIEGARCH com distribuição α-estável. Os resultados aqui apresentados são parte de um trabalho apresentado no 19º SINAPE, realizado em São Pedro, em julho de 2010.

Nesta simulação consideramos:

- $p = q = 0$, $\omega = 0$ e $\theta = \gamma = 0.10$ (o sinal de θ não é identificável através do método),
- $\alpha \in \{1.25, 1.50, 1.75, 2.00\}$, $d \in \{0.1, 0.2, 0.3, 0.4\}$, sob a restrição $d < 1 - \frac{1}{\alpha}$,
- em todas as simulações realizadas, fixamos o parâmetro σ associado à distribuição Gaussiana como sendo 1 (para detalhes, veja a Seção 5.4.3),
- devido ao custo computacional apenas 100 replicações foram consideradas (na época em que realizamos o trabalho o CESUP ainda não era utilizado),
- para cada replicação, a série temporal gerada possui tamanho amostral $n = 5000$.

O vetor de parâmetros $\boldsymbol{\eta} = (d, \theta, \gamma)'$ do modelo são estimados utilizando-se o estimador Whittle modificado, descrito na Seção 5.4.3. Lembramos que, dado um processo FIEGARCH qualquer, as variáveis aletórias $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$ seguem um modelo ARFIMA. Em particular, quando $\gamma = 0$ e $Z_0 \sim S\alpha S$, então $\ln(h_t^2)$ possui distribuição α-estável. Tendo em vista esse resultado e observando que Kokoszka e Taqqu (1999) provam a consistência e normalidade assintótica do estimador em questão para os processos ARFIMA, utilizamos ambas as séries temporais $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$ e $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$ na estimação dos parâmetros. Ainda a título de comparação, para a sére temporal $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$, utilizamos além da função periodograma normalizado, a função periodograma tradicional.

Os resultados da simulação considerando $\{\ln(h_t^2)\}_{t\in\mathbb{Z}}$ e as funções periodograma normalizado e periodograma tradicional são dadas nas Tabelas J.1 e J.2, respectivamente. A Tabela J.3 mostra os resultados da simulação considerando-se $\{\ln(X_t^2)\}_{t\in\mathbb{Z}}$. Os valores reportados a seguir são o valor médio ($\bar{\eta}$) estimado, o desvio padrão do estimador (sd), o vício ($bias$), o erro médio absoluto (mae) e o erro quadrático médio (mse).

Tabela J.1: Estatísticas para os valores estimados de d, θ e γ considerando-se a série temporal $\{\ln(h_t^2)\}_{t=1}^n$ e a função periodograma tradicional, para $d_1 = 0.1$, $d_2 = 0.2$, $d_3 = 0.3$ e $d_4 = 0.4$.

η_0	d_1	θ	γ	d_2	θ	γ	d_3	θ	γ	d_4	θ	γ
						$\alpha = 1.25, \quad \theta = \gamma = 0.1$						
$\bar{\eta}$	0.1016	2.3855	-0.0656									
$bias$	0.0016	2.2855	-0.1656									
sd	0.0001	3.9189	0.1183	$d \geq 1 - \frac{1}{\alpha}$			$d \geq 1 - \frac{1}{\alpha}$			$d \geq 1 - \frac{1}{\alpha}$		
mae	0.0059	2.2855	0.2245									
mse	0.0001	9.1423	0.1458									
						$\alpha = 1.50, \quad \theta = \gamma = 0.1$						
$\bar{\eta}$	0.0993	0.7795	0.1239	0.2009	0.7805	0.0992	0.3031	0.7749	-0.1961			
$bias$	-0.0007	0.6795	0.0239	0.0009	0.6805	-0.0008	0.0031	0.6749	-0.2961			
sd	0.0001	0.3617	0.0066	0.0002	0.3596	0.0131	0.0005	0.3380	0.0727	$d \geq 1 - \frac{1}{\alpha}$		
mae	0.0066	0.6795	0.0661	0.0080	0.6805	0.0767	0.0098	0.6749	0.2993			
mse	0.0001	0.8234	0.0072	0.0002	0.8228	0.0131	0.0005	0.7935	0.1603			
						$\alpha = 1.75, \quad \theta = \gamma = 0.1$						
$\bar{\eta}$	0.0989	0.4063	0.1582	0.1990	0.4154	0.1176	0.2989	0.4166	-0.0381	0.3979	0.4182	-0.0798
$bias$	-0.0011	0.3063	0.0582	-0.0010	0.3154	0.0176	-0.0011	0.3166	-0.1381	-0.0021	0.3182	-0.1798
sd	0.0001	0.2084	0.0056	0.0001	0.2052	0.0041	0.0001	0.1898	0.0574	0.0001	0.1973	0.0261
mae	0.0071	0.3063	0.0746	0.0072	0.3154	0.0440	0.0072	0.3166	0.1572	0.0076	0.3182	0.1876
mse	0.0001	0.3022	0.0090	0.0001	0.3047	0.0044	0.0001	0.2901	0.0765	0.0001	0.2985	0.0584
						$\alpha = 2.00, \quad \theta = \gamma = 0.1$						
$\bar{\eta}$	0.0989	0.1261	0.1725	0.1991	0.1424	0.1357	0.2993	0.1597	0.0636	0.3979	0.1606	-0.0011
$bias$	-0.0011	0.0261	0.0725	-0.0009	0.0424	0.0357	-0.0007	0.0597	-0.0364	-0.0021	0.0606	-0.1011
sd	0.0001	0.0002	0.0005	0.0001	0.0000	0.0003	0.0001	0.0000	0.0004	0.0001	0.0001	0.0033
mae	0.0094	0.0285	0.0726	0.0095	0.0424	0.0360	0.0095	0.0597	0.0373	0.0095	0.0606	0.1090
mse	0.0001	0.0009	0.0058	0.0001	0.0018	0.0016	0.0001	0.0036	0.0017	0.0001	0.0038	0.0136

Tabela J.2: Estatísticas para os valores estimados de d, θ e γ considerando-se a série temporal $\{\ln(h_t^2)\}_{t=1}^n$ e a função periodograma normalizado, para $d_1 = 0.1$, $d_2 = 0.2$, $d_3 = 0.3$ e $d_4 = 0.4$.

.

η_0	d_1	θ	γ	d_2	θ	γ	d_3	θ	γ	d_4	θ	γ
						$\alpha = 1.25, \quad \theta = 0.1 = \gamma$						
$\bar{\eta}$	0.1017	2.4727	-0.0515									
$bias$	0.0017	2.3727	-0.1515									
sd	0.0001	0.0002	0.0112	$d \geq 1 - \frac{1}{\alpha}$			$d \geq 1 - \frac{1}{\alpha}$			$d \geq 1 - \frac{1}{\alpha}$		
mae	0.0058	2.3727	0.1752									
mse	0.0001	5.6299	0.0341									
						$\alpha = 1.50, \quad \theta = 0.1 = \gamma$						
$\bar{\eta}$	0.0994	2.4802	-0.0533	0.2008	2.3732	0.0525	0.3031	2.0615	-0.7159			
$bias$	-0.0006	2.3802	-0.1533	0.0008	2.2732	-0.0475	0.0031	1.9615	-0.8159			
sd	0.0001	0.0001	0.0116	0.0003	0.0035	0.0849	0.0005	0.0352	0.1586	$d \geq 1 - \frac{1}{\alpha}$		
mae	0.0066	2.3802	0.1719	0.0082	2.2732	0.2136	0.0099	1.9615	0.8226			
mse	0.0001	5.6656	0.0351	0.0003	5.1711	0.0871	0.0005	3.8827	0.8243			
						$\alpha = 1.75, \quad \theta = 0.1 = \gamma$						
$\bar{\eta}$	0.0988	2.4826	-0.0610	0.1990	2.3879	0.0688	0.2989	2.1350	-0.7280	0.3979	1.7378	-0.5101
$bias$	-0.0012	2.3826	-0.1610	-0.0010	2.2879	-0.0312	-0.0011	2.0350	-0.8280	-0.0021	1.6378	-0.6101
sd	0.0001	0.0000	0.0063	0.0001	0.0003	0.0761	0.0001	0.0077	0.1235	0.0001	0.0419	0.0107
mae	0.0070	2.3826	0.1705	0.0073	2.2879	0.1997	0.0074	2.0350	0.8280	0.0076	1.6378	0.6101
mse	0.0001	5.6770	0.0322	0.0001	5.2350	0.0771	0.0001	4.1490	0.8090	0.0001	2.7243	0.3829
						$\alpha = 2.00, \quad \theta = 0.1 = \gamma$						
$\bar{\eta}$	0.0989	2.4823	-0.0377	0.1989	2.3882	0.0625	0.2992	2.1264	-0.8072	0.3981	1.7928	-0.5411
$bias$	-0.0011	2.3823	-0.1377	-0.0011	2.2882	-0.0375	-0.0008	2.0264	-0.9072	-0.0019	1.6928	-0.6411
sd	0.0001	0.0000	0.0132	0.0001	0.0003	0.0774	0.0001	0.0094	0.1715	0.0001	0.0200	0.0081
mae	0.0095	2.3823	0.1642	0.0094	2.2882	0.2006	0.0094	2.0264	0.9072	0.0095	1.6928	0.6411
mse	0.0001	5.6754	0.0322	0.0001	5.2361	0.0789	0.0001	4.1158	0.9946	0.0001	2.8855	0.4191

Tabela J.3: Estatísticas para os valores estimados de d, θ e γ considerando-se a série temporal $\{\ln(X_t^2)\}_{t=1}^n$ e a função periodograma normalizado, para $d_1 = 0.1$, $d_2 = 0.2$, $d_3 = 0.3$ e $d_4 = 0.4$.

η_0	d_1	θ	γ	d_2	θ	γ	d_3	θ	γ	d_4	θ	γ
					$\alpha = 1.25, \quad \theta = \gamma = 0.1$							
$\bar{\eta}$	0.1222	0.9916	0.5918									
$bias$	0.0222	0.8916	0.4918									
sd	0.0032	0.1030	0.0163	$d \geq 1 - \frac{1}{\alpha}$			$d \geq 1 - \frac{1}{\alpha}$			$d \geq 1 - \frac{1}{\alpha}$		
mae	0.0465	0.9375	0.4918									
mse	0.0037	0.8979	0.2581									
					$\alpha = 1.50, \quad \theta = \gamma = 0.1$							
$\bar{\eta}$	0.0609	1.0660	0.3825	0.1396	1.0407	0.3559	0.2363	0.9357	0.3304			
$bias$	-0.0391	0.9660	0.2825	0.0396	0.9407	0.1559	0.1363	0.8357	0.0304			
sd	0.0031	0.0002	0.0154	0.0048	0.0013	0.0248	0.0076	0.0471	0.0500	$d \geq 1 - \frac{1}{\alpha}$		
mae	0.0517	0.9660	0.2825	0.0622	0.9407	0.1715	0.1382	0.8569	0.1513			
mse	0.0046	0.9333	0.0952	0.0063	0.8861	0.0490	0.0261	0.7456	0.0509			
					$\alpha = 1.75, \quad \theta = \gamma = 0.1$							
$\bar{\eta}$	0.0319	1.0150	0.2395	0.0882	1.0071	0.2103	0.1567	0.9680	0.1726	0.2417	0.8702	0.1311
$bias$	-0.0681	0.9150	0.1395	-0.0118	0.9071	0.0103	0.0567	0.8680	-0.1274	0.1417	0.7702	-0.2689
sd	0.0063	0.0089	0.0167	0.0061	0.0093	0.0213	0.0069	0.0307	0.0349	0.0086	0.0717	0.0568
mae	0.0853	0.9150	0.1593	0.0602	0.9071	0.1063	0.0708	0.8781	0.1834	0.1423	0.7952	0.3283
mse	0.0109	0.8460	0.0362	0.0063	0.8322	0.0214	0.0101	0.7840	0.0511	0.0286	0.6649	0.1291
					$\alpha = 2.00, \quad \theta = \gamma = 0.1$							
$\bar{\eta}$	0.0154	0.9570	0.1395	0.0482	0.9558	0.1147	0.0923	0.9507	0.0803	0.1505	0.9325	0.0323
$bias$	-0.0846	0.8570	0.0395	-0.0518	0.8558	-0.0853	-0.0077	0.8507	-0.2197	0.0505	0.8325	-0.3677
sd	0.0116	0.0007	0.0104	0.0101	0.0006	0.0105	0.0070	0.0009	0.0107	0.0049	0.0043	0.0107
mae	0.0932	0.8570	0.0864	0.0710	0.8558	0.1089	0.0588	0.8507	0.2233	0.0757	0.8325	0.3677
mse	0.0188	0.7351	0.0120	0.0128	0.7330	0.0178	0.0070	0.7247	0.0590	0.0074	0.6974	0.1459

Os resultados apresentados nas Tabelas J.1 - J.3 mostram que, quando consideramos a série temporal $\{\ln(h_t^2)\}_{t=1}^n$, o parâmetro d sempre fica bem estimado. O mesmo não ocorre quando consideramos $\{\ln(X_t^2)\}_{t=1}^n$. A estimação de θ e γ não é satisfatória em nenhum dos casos estudados. Esse resultado não nos surpreende pois, como mencionamos anteriormente, assumimos $\sigma = 1$. Basta observar que, por exemplo, fixando-se $\sigma = 1$ no caso Gaussiano ($\alpha = 2.00$), o modelo não está corretamente especificado pois $Z_0 \sim S_\alpha(0, \sigma, 0)$ implica $Z_0 \sim \mathcal{N}(0, 2\sigma^2)$, para qualquer $\sigma > 0$. Espera-se que, assumindo σ desconhecido e estimando-se esse valor juntamente com os demais parâmetros, a estimação dos coeficientes do modelo FIEGARCH melhore.