Modeling volatility in Heat Rate Variability

Argentina Leite[1] and Maria Eduarda Silva[2] and Ana Paula Rocha[3], *Member, IEEE*

***Abstract*— Modeling Heart Rate Variability (HRV) data has become important for clinical applications and as a research tool. These data exhibit long memory and time-varying conditional variance (volatility). In HRV, volatility is traditionally estimated by recursive least squares combined with short memory AutoRegressive (AR) models. This work considers a parametric approach based on long memory Fractionally Integrated AutoRegressive Moving Average (ARFIMA) models with heteroscedastic errors. To model the heteroscedasticity nonlinear Generalized Autoregressive Conditionally Heteroscedastic (GARCH) and Exponential Generalized Autoregressive Conditionally Heteroscedastic (EGARCH) models are considered. The latter are necessary to model empirical characteristics of conditional volatility such as clustering and asymmetry in the response, usually called leverage in time series literature. The ARFIMA-EGARCH models are used to capture and remove long memory and characterize conditional volatility in 24 hour HRV recordings from the Noltisalis database.**

I. INTRODUCTION

Heart Rate Variability (HRV) reflects the interaction between perturbations to the cardiovascular variables and the corresponding response of the cardiovascular regulatory systems [1]. Thus modeling such variability can provide a quantitative and non-invasive method to assess the integrity of the cardiovascular system. HRV data display non stationary characteristics and exhibit long memory and time-varying conditional variance (usually designated by volatility) which may contain indicators of current disease or warnings about impending diseases.

Traditionally, HRV data can be characterized by linear AutoRegressive (AR) models, which describe only short memory in the mean. These models combined with recursive least squares have been used to estimate volatility in HRV data [2]. However, it is acknowledged that complex interactions of electrophysiological, humoral variables and autonomic and central nervous regulations induce nonlinear effects in HRV [3], [4]. Thus the analysis of HRV based on nonlinear models might elicit valuable information for its physiological interpretation as well as risk assessment. A nonlinear approach to volatility description in HRV data was proposed by Leite *et al* [5], [6] using Fractionally Integrated AutoRegressive Moving Average (ARFIMA) models with Generalized AutoRegressive Conditional Heteroscedastic (GARCH) errors, an extension of the usual AR analysis. ARFIMA-GARCH models are used to capture and remove long memory and estimate volatility in 24 hour HRV recordings [6], [7]. However, GARCH models assume that volatility depends only on the magnitude of the shocks and not on their sign, meaning that positive and negative shocks have a symmetric effect on volatility [8].

This work considers Exponential GARCH (EGARCH) [8] models which are an extension of GARCH models. EGARCH models assume that the effect of positive and negative shocks on volatility is asymmetric, an effect usually designated by leverage. A model with leverage effect is more suited to describe the complex characteristics of HRV data. The ARFIMA-EGARCH models are applied to 24 hour HRV recordings from 30 subjects of the Noltisalis database [9]: 10 healthy subjects (N), 10 patients suffering from congestive heart failure (C) and 10 heart transplanted patients (T). The database was collected by the cooperative effort of university departments and rehabilitation clinics in Italy, [9].

II. MODELS WITH CONDITIONAL VOLATILITY

In this work we consider models satisfying

$$\phi(B)(1-B)^d x_t = \theta(B)\epsilon_t \quad (1)$$

$$\epsilon_t = \sigma_t z_t \quad (2)$$

$$\sigma_t^2 = \mathrm{Var}(\epsilon_t | H_{t-1}) \quad (3)$$

where B is the backward-shift operator. Equation (1) describes the conditional mean of the process with serially uncorrelated residuals ϵ_t and is said an ARFIMA(p,d,q) with $p,q \in \mathbb{N}_0$ and $d \in \mathbb{R}$, [10]; d is the long-memory parameter, determines the long-term behaviour in mean and $(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k$ is the fractional difference operator [10]; $p,q \in \mathbb{N}$ and the polynomials $\phi(B) = 1 - \phi_1 B - ... - \phi_p B^p$ and $\theta(B) = 1 + \theta_1 B + ... + \theta_q B^q$ allow for the modeling of the short-range properties in the mean; for $-0.5 < d < 0.5$ the ARFIMA process is covariance stationary. In the range $-0.5 < d < 0.5$, the long memory parameter is related to the Hurst coefficient, H, to the fractal dimension, D, and to the slope of the (generalized) spectral density in the low frequency range, α, by $d = H - 0.5$, $H = 2 - D$ and $\alpha = 2d$, respectively. Moreover, for $0.5 \leq d < 1$ the process ARFIMA is non-stationary and mean reverting.

*This work is financed by the ERDF - European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme within project *POCI-01-0145-FEDER-006961*, and by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia as part of projects UID/EEA/50014/2013, CIDMA UID/MAT/04106/2013 and CMUP UID/MAT/00144/2013, funded by FCT (Portugal) with national (MEC) and European structural funds through the programs FEDER, under the partnership agreement PT2020.

[1]Argentina Leite is with Escola de Ciências e Tecnologia, Universidade de Trás-os-Montes e Alto Douro & C-BER & INESC TEC, Portugal tinucha@utad.pt

[2]Maria Eduarda Silva is with Faculdade de Economia, Universidade do Porto & CIDMA, Portugal mesilva@fep.up.pt

[3]Ana Paula Rocha is with Faculdade de Ciências, Universidade do Porto & CMUP, Portugal aprocha@fc.up.pt

Equations (2) and (3) describe the conditional variance of the process which varies over time as in time-varying AR models. In (2) ϵ_t are called shocks and z_t, independent and identically distributed random variables with zero mean and unit variance, are the standardized shocks. The conditional variance σ_t^2 in (2) depends on H_{t-1} which is the history of process and includes past variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_{t-1}^2$, and past shocks $\epsilon_1, \epsilon_2, \ldots, \epsilon_{t-1}$.

There are several models to govern the evolution of σ_t^2. The most common is the GARCH(P,Q), $P,Q \in \mathbb{N}_0$ model [8] under which

$$\sigma_t^2 = u_0 + \sum_{i=1}^{P} v_i \, \sigma_{t-i}^2 + \sum_{i=1}^{Q} u_i \, \epsilon_{t-i}^2 \qquad (4)$$

This model requires non-negativity constraints on the parameters to ensure positive conditional variances, $u_0 > 0$, $v_1, ..., v_P$, $u_1, ..., u_Q \geq 0$, $\sum_{i=1}^{P} v_i + \sum_{j=1}^{Q} u_j < 1$. The parameters, u_i and v_i characterize the volatility clustering phenomena observed in many data sets. It is noteworthy that GARCH models assume that conditional volatility depends only on magnitude of the shocks and not on their sign.

Several extensions of GARCH models have been proposed so that empirical characteristics of the data such as asymmetric response to positive and negative shocks, are adequately modeled. One such model is the EGARCH(P,Q) [8] defined as follows:

$$\log\sigma_t^2 = u_0^* + \sum_{i=1}^{P} v_i \log\sigma_{t-i}^2 + \sum_{i=1}^{Q} u_i \, |z_{t-i}| + \sum_{i=1}^{Q} \xi_i z_{t-i} \quad (5)$$

where $u_0^* = u_0 - \sum_{i=1}^{Q} u_i \sqrt{\frac{2}{\pi}}$ and $z_t = \epsilon_t/\sigma_t$.

This process does not require constraints on the parameters for ensuring the positivity of the variance. The parameters u_i and v_i characterise the volatility clustering phenomena and the parameters ξ_i describe the leverage effects. The impact of positive shocks, $\epsilon_{t-i} > 0$ is $(u_i + \xi_i)\frac{\epsilon_{t-i}}{\sigma_{t-i}}$, while for negative shocks it is $(u_i - \xi_i)\frac{\epsilon_{t-i}}{\sigma_{t-i}}$. If $\xi_i = 0$, $\log\sigma_t^2$ responds symmetrically to ϵ_{t-i}.

In this work, we consider ARFIMA$(p,d,0)$-GARCH$(1,1)$ and ARFIMA$(p,d,0)$-EGARCH$(1,1)$ models, since they are a natural extension of the classic AR(p) models usual in the analysis of HRV. Additionally to the parameters d, which characterises the long memory in the mean, u_1 and v_1, which characterize the volatility clustering, we give special attention to ξ which describes the leverage effect in the conditional variance.

Given a time series $x_1, \ldots, x_n$, to estimate the parameters of the above mentioned models proceed as follows [8], [10]: (i) estimate d using the semi-parametric local Whittle estimator; (ii) define the filtered data $y_t = (1-B)^d x_t$; (iii) estimate the AR(p)-GARCH$(1,1)$ or AR(p)-EGARCH$(1,1)$ parameters in the filtered data y_t by maximum likelihood (Econometrics Toolbox of MATLAB [11]), with the order p determined by the Akaike Information Criterion (AIC).

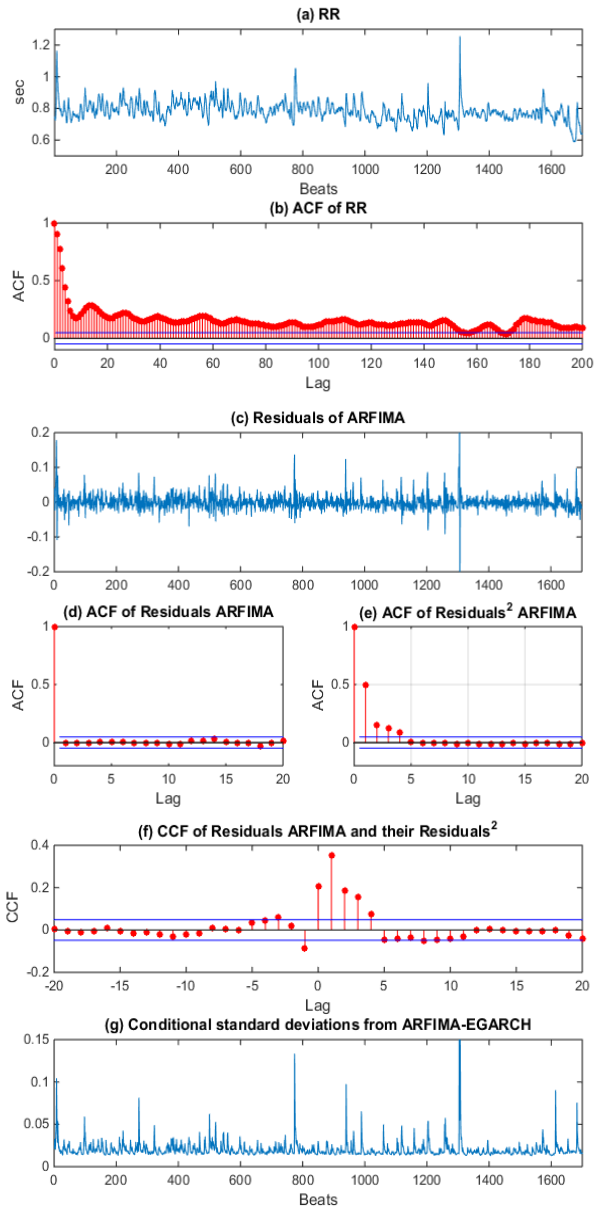

Fig. 1. Short-term HRV data: (a) tachogram of a normal subject (segment with 1700 beats of RR series for subject-N6 from the Noltisalis database), (b) ACF of the data, (c) residuals of the fitted ARFIMA(9,0.34,0) model, (d) ACF of the residuals, (e) ACF of the squared residuals and (f) CCF of the residuals and the squared residuals. The horizontal lines (- -) show the 95% confidence limits, (g) estimated volatility, $\hat{\sigma}_t$.

III. MODELING HRV

To motivate the use of ARFIMA(p,d,0)-EGARCH(1,1) models in HRV data, consider the tacogram of the healthy subject N6 (segment of RR series with 1700 beats) represented in Fig. 1(a). The ACF (autocorrelation function) of the data, represented in (b), shows a typical very slow decay indicating the presence of long memory which can be adequately modelled with an ARFIMA, equation (1). The residuals ($\hat{\epsilon}_t$) from the fitted ARFIMA$(p = 9, d = 0.34, 0)$ (chosen by AIC) are represented in Fig. 1(c). The corresponding ACF in (d), exhibits small correlation indi-

cating that the ARFIMA model is adequate to explain the dynamics of the conditional mean of the data. However, the squared residuals exhibit significant autocorrelation in Fig. 1(e), indicating time-varying conditional variance which can be modeled by a GARCH model, (4), hereafter denoted as *Model I*: ARFIMA(9,0.34,0)-GARCH(1,1). Additionally, the cross-correlation between the residuals $\hat{\epsilon}_t$ and their squares $\hat{\epsilon}_t^2$, Fig. 1(f), indicates that the conditional variance depends also on the lagged shocks, an effect that can be modeled by an EGARCH model, (5), hereafter denoted as *Model II*: ARFIMA(9,0.34,0)-EGARCH(1,1). The results for two models are summarized in Table I.

McLeod-Li testing [10] of the residuals of Model II gives a test statistic $Q(20) = 41.40$, indicating no significant conditional heteroscedasticity in the residual series. Moreover, the cross-correlation between the residuals and their squares in lag 1, $CCF(1)$, is 0.04 indicating no significant correlation and the AIC criterion favours Model II.

These results indicate that the ARFIMA(9,d,0)-EGARCH(1,1) (Model II) model leads to further characterisation of the HRV data. In fact, the estimate $\hat{\xi}_1 = 0.24$ indicates asymmetric response to positive and negative shocks. The conditional standard deviation estimate $\hat{\sigma}_t$ represented in Fig. 1(g), captures very well the heteroscedasticity in the original data, plotted in (a). Similar results were obtained in other short HRV recordings.

TABLE I

ESTIMATES (STANDARD DEVIATION) FOR MODEL I AND MODEL II FITTED TO DATA IN FIG. 1. $Q(20)$ STANDS FOR MCLEOD-LI' S TEST STATISTIC FOR THE RESIDUALS.

Parameter	Model I	Model II
$\hat{d}$	0.34	0.34
$Q(20)$ of $\hat{\epsilon}_t$	494.32	494.32
CCF(1) of $\hat{\epsilon}_t$,$\hat{\epsilon}_t^2$	0.35	0.35
$\hat{u}_0$	$0.13e^{-03}(0.16e^{-04})$	-2.87 (0.20)
$\hat{u}_1$	0.33 (0.02)	0.47 (0.05)
$\hat{v}_1$	0.40 (0.05)	0.63 (0.03)
$\hat{\xi}_1$	—	0.24 (0.02)
$Q(20)$ of $\hat{z}_t$	28.84	41.40
CCF(1) of $\hat{z}_t$,$\hat{z}_t^2$	0.12	0.04
AIC of $\hat{z}_t$	-8578.00	-8623.00

In the case of long recordings, such as ambulatory 24 hour HRV data (approximately 100000 beats), exhibiting several non stationary characteristics, ARFIMA-EGARCH modeling is combined with adaptive segmentation [6]: long records are decomposed into short records of variable length and the break points, which mark the end of consecutive short records, are identified by AIC criterion. The short records thus obtained have a minimum length 512 and are subsequently modeled using ARFIMA-EGARCH models.

IV. RESULTS AND DISCUSSION

In this section, the above methodology is applied to long-term HRV series of subjects from the Noltisalis database [9]. The results are first illustrated for the healthy subject N6 and the patient C10, in Figs. 2 and 3, respectively.

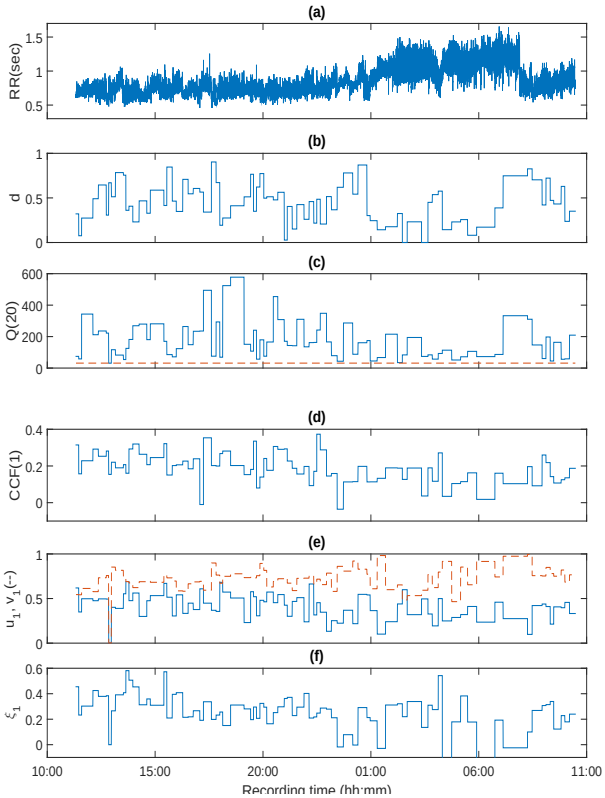

Fig. 2. (a) Tachogram of healthy subject N6, 24 hour recordings Noltisalis database, Evolution over 24 hours of $\hat{d}$ in (b), $\hat{u}_1$(-) and $\hat{v}_1$(- -) in (e) and $\hat{\xi}_1$ in (f) estimated using ARFIMA-EGARCH models and segmentation; statistic $Q(20)$ of the McLeod-Li test for conditional heteroscedasticity in (c) and cross-correlation between the residuals ϵ_t and their squares of lag 1 in (d).

The long memory estimates $\hat{d}$ in Fig. 2(b) and Fig. 3(b), change over time showing circadian variation, with lowest values during the night periods, $0 < \hat{d} < 0.5$ in contrast with $0.5 < \hat{d} < 1$ for the day period. These results are in concordance with Leite *et al.*[6].

The residuals from the ARFIMA modeling are tested for conditional heteroscedasticity with McLeod-Li test. The corresponding test statistics $Q(20)$ are represented in plots (c) in Figs. 2 and 3. It is found that the percentage of segments with heteroscedasticity are 98.8% and 94.5 % for subjects N6 and C10, respectively. Moreover, the volatility parameters estimates $\hat{u}_1$ and $\hat{v}_1$, Fig. 2(e) and Fig. 3(e), change over time with some circadian variation for the healthy subject. For the sick subject, the estimate $\hat{v}_1$ increases and $\hat{u}_1$ decreases. Note that the estimated values for parameter v_1 are over 0.5 indicating some persistence in conditional variance. These results are in agreement with Leite *et al.*[6].

The residuals from the ARFIMA models are also checked for the asymmetric effect (leverage). Plot (d) in Figs. 2 and 3 represent the values of CCF(1) between the residuals and their squares. The correlations indicate that the conditional variance depends also on the lagged shocks, leverage effect. In fact, the percentage of segments with leverage effect are 93.8% and 43.8 % for subjects N6 and C10, respectively.

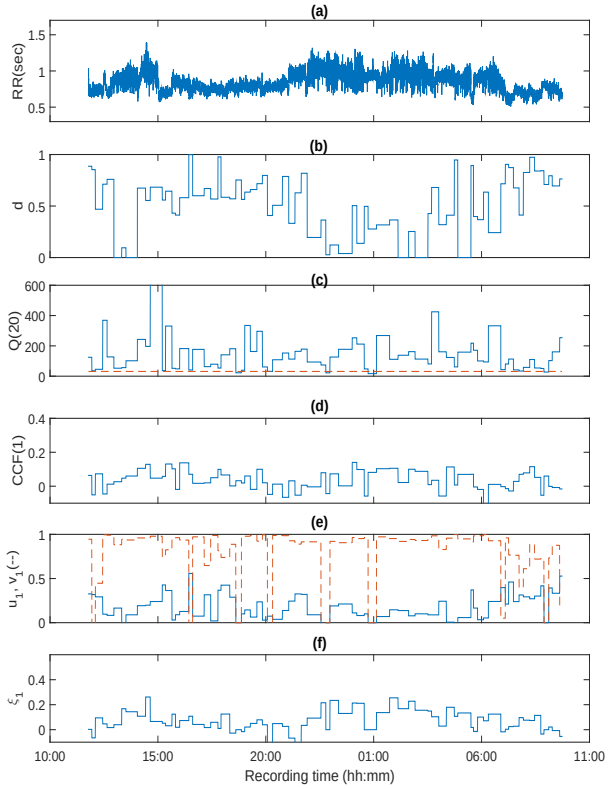

Fig. 3. (a) Tachogram of patient C10 affected by congestive heart failure, 24 hour recordings Noltisalis database. Evolution over 24 hours of $\hat{d}$ in (b), $\hat{u}_1$(-) and $\hat{v}_1$(- -) in (e) and $\hat{\xi}_1$ in (f) estimated using ARFIMA-EGARCH models and segmentation; statistic $Q(20)$ of the McLeod-Li test for conditional heteroscedasticity in (c) and cross-correlation between the residuals ϵ_t and their squares of lag 1 in (d).

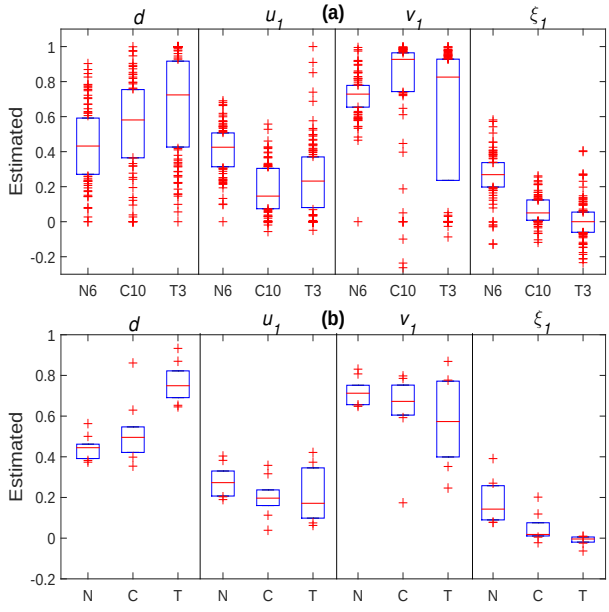

Fig. 4. Boxplots for $\hat{d}$, $\hat{u}_1$, $\hat{v}_1$ and $\hat{\xi}_1$ using ARFIMA-EGARCH models and segmentation: (a) an healthy subject-N6, a patient suffering from congestive heart failure-C10 and a heart transplanted patient-T3, (b) subjects mean values for each group N, C, T.

Moreover, the leverage parameter estimates $\hat{\xi}_1$, Fig. 2(f) and Fig. 3(f), change over time and present higher values for the healthy subject. The same analysis is performed for patient T3. The results for subjects N6, C10 and T3 are summarized in Fig. 4(a).

Finally, the mean estimates $\hat{d}$, $\hat{u}_1$, $\hat{v}_1$ and $\hat{\xi}_1$ for each of the 30 subjects in the database are summarized in Fig. 4(b). These overall results indicate that the long memory parameter d is lower for healthy subjects N while the volatility and the leverage parameters u_1, ξ_1 respectively, are lower for sick subjects C and T. The parameter v_1 presents high variability for the sick subjects T. The results are promising in differentiating health and disfunction situations deserving further study.

V. CONCLUSIONS

This paper contributes for further characterization of the complex dynamics of HRV with a leverage parameter. In particular, the model under discussion indicates that values of HRV under the mean lead to less variability in HRV than values over the mean- asymmetric effect captured by the leverage parameter. Furthermore this study indicates that this effect is stronger in healthy subjects. However, further studies are necessary to assess the importance of this parameter in risk stratification and autonomic nervous system dysfunction characterization.

REFERENCES

[1] Task Force of the European Society of Cardiology and North American Society of Pacing Electrophysiology, Heart rate variability: standards of measurement, physiological interpretation and clinical use, Circulation, vol. 93, pp. 1043-1065, 1996.

[2] L. T. Mainardi, On the quantification of heart rate variability spectral parameters using time-frequency and time-varying methods, Phil. Trans. R. Soc. A, vol. 367, pp. 2555-275, 2009.

[3] R. T. Baillie, A. A. Cecen, and C. Erkal, Normal heartbeat series are nonchaotic, nonlinear, and multifractal: New evidence from semiparametric and parametric tests, Chaos, vol. 19, pp. 028503, 2009.

[4] A. Voss, S. Schulz, R. Schroeder, M. Baumert, and P. Caminal, Methods derived from nonlinear dynamics for analysing heart rate variability. Phil. Trans. R. Soc. A, vol. 367, pp. 277-296, 2009.

[5] A. Leite, A. P. Rocha, and M. E. Silva, Long memory and volatility in HRV: an ARFIMA-GARCH approach, in Computing in Cardiology, vol. 36, pp. 165-168, 2009.

[6] A. Leite, A. P. Rocha, and M. E. Silva, Beyond long memory in heart rate variability: an approach based on fractionally integrated autoregressive moving average time series models with conditional heteroscedasticity, Chaos, vol. 23, pp. 023103, 2013.

[7] M. Y. Hu, and C. Tsoukalas, Conditional volatility properties of sleep disordered breathing, Comput. Biol. Med., vol. 36, pp. 303-312, 2006.

[8] R. S. Tsay, Analysis of Financial Time Series (2nd ed.). Wiley-Interscience, 2005.

[9] M. G. Signorini, R. Sassi, S. Cerutti, Working on the NOLTISALIS Database: Measurement of nonlinear properties in heart rate variability signals. In: Proceedings of IEEE-EMBS International Conference, Istanbul, Turkey (IEEE, Piscataway), pp. 547-550, 2001.

[10] P. Boukhan, G. Oppenheim, and M. S. Taqqu, Theory and applications of long-range dependence, Boston: Birkhäuser, 2003.

[11] MATLAB, *version 8.5 (R2015a)*, The MathWorks Inc., Natick, Massachusetts, 2015.