Statistica & Applicazioni - Vol. XVIII, n. 2, 2020, pp. 177-196

FORECASTING MACROECONOMIC VOLATILITY WITH SCORE-DRIVEN MODELS

Raffaele Mattera*
Massimiliano Giacalone*

SUMMARY

Business cycle volatility has been extensively studied by means of the well-known ARCH and GARCH processes. Aim of this paper is to show that the score-driven models are instead more accurate in predicting business cycle volatility than the GARCH-type models. Motivated by fact that the empirical evidence do not support the hypothesis of Gaussianity also for business cycles, we assume the Generalized Error Distribution and its extension for skewness in estimating the volatility models within the GARCH framework. After reviewing the basic properties of the score-driven approaches, we carry out an empirical analysis with respect to the business cycles of the United States and Japan. We show that the score-driven models provide superior performances than both Gaussian and non-Gaussian GARCH processes in forecasting business cycle volatility.

Keywords: Generalized Autoregressive Score, Gross Domestic Product, GARCH, Model Confidence Set, Skewed Generalized Error Distribution.
DOI: 10.26350/999999_000037
ISSN: 18246672 (print) 2283-6659 (digital)

1. INTRODUCTION

In the last few years many authors have tried to get an accurate forecasts of Gross Domestic Product volatility because those predictions are relevant to investors, government regulators and capital markets.

Indeed, Schwert (1989) demonstrated that stock returns volatility varies over time because of the macroeconomic volatility that is also time varying.

Similarly, Lettau, Ludvigson and Wachter (2008) claim that the decline in macroeconomic volatility may lead to a decline in the equity risk-premium.

As a further evidence, Diebold and Yilmaz (2009) documented the existence of a relationship between macroeconomic volatility and stock market volatility, as well as Baele, Bekaert and Inghelrecht (2010) co-movements between bonds and stocks are driven by macroeconomic volatility, among other factors.

Therefore this overall evidence justifies the relevance of an accurate macroeconomic volatility predictions for investors and policy makers.

* Dipartimento di Economia e Statistica - Università di Napoli ''Federico II'' - via Cintia, Monte Sant'Angelo Complex, 80126, NAPOLI (e-mail: massimiliano.giacalone@unina.it; raffaele.mattera@unina.it).

Researchers examined the business cycles properties only in terms of the conditional mean (Hamori, 2000).

The GDP was studied in US, UK, and Japan, the authors tried to determine empirically the nature of macroeconomic volatility, for example whether it is asymmetric or not. By comparing different GARCH-type processes (the GARCH, the Treshold GARCH end the Exponential GARCH) in terms of in-sample goodness of fit, Hamori (2000) found that the standard GARCH process was the best in describing volatility dynamic.

On the contrary were found empirical studies excluding the dimension of conditional volatility (Ho and Tsui, 2003; Ho and Albert, 2004). In particular, revisiting the study of Hamori (2000), Ho and Tsui (2003) found statistically significant evidence in favor of asymmetric volatility in the real growth rates of the United States and Canada.

On the other side, Ho and Albert (2004) show the evidence of conditional volatility in the quarterly real GDP of China and the Exponential GARCH model among the Asymmetric Power ARCH models were employed to capture the possible existence of asymmetric conditional volatility in real GDP. The authors found that negative real GDP shocks may induce a greater impact on future volatilities compared with positive shocks of the same magnitude. Fang, Miller and Lee (2008) studied the properties of variance in quarterly real GDP growth, it rates their effects on conditional volatility for Canada, Germany, Italy, Japan, United Kingdom and the United States from 1957 to 2006; as any other effects on product growth volatility. Their initial results, based on a GARCH model of the conditional variance of residuals, find strong evidence for the persistence of volatility in the growth rate.

Later, Fang and Miller (2009) reexamined the Japanese case of Hamori (2000) with the same approach and showed the high persistence of GDP volatility.

More recently, Bodman (2009) has been provided an evidence of asymmetry in volatility across phases of the cycle for the Australian GDP by estimating various GARCH models. Similarly, Hai, Tsui and Zhang (2013) also was found the same asymmetry in volatility of GDP for East Asian tigers countries by means of an Exponential GARCH model.

As it clearly appears, the main purpose of most of these studies was to find the best GARCH model in terms of in-sample goodness of fit, discarding the forecasting ability of the underlying statistical models.

Only recently, Clark and Ravazzolo (2015) compared alternative models of time-varying volatility on the basis of the accuracy of both point and density forecasts the key macroeconomic time series for the USA.

Following Stock and Watson (2007), we claim that, accounting for the time-varying nature of the volatility could help in better forecasting the countries' GDP instead. Therefore, we aim to find the best forecasting model for GDP volatility of several countries.

The real aim is to compare the out-of-sample forecasts of the several aforementioned GARCH models, with their standard and symmetric specification as the GARCH of Bollerslev (1986), some asymmetric specification like the GJR of Glosten, Jagannathan and Runkle (1993) or the Asymmetric Power ARCH of Ding, Granger and Engle (1993) and Ding and Granger (1996), the Integrated GARCH of Engle and Bollerslev (1986) and a nonlinear and asymmetric GARCH model as in the case of the Exponential GARCH (Nelson, 1991).

Despite the relevance of higher order moments in statistics (Fiori and Zenga, 2009) they have not been only poorly considered in macroeconomic forecasting. Moreover, following some studies in stock market volatility forecasting (Wilhelmsson, 2006; Mattera and Giacalone, 2018; Köchling, Schmidtke and Posch, 2020), we also account for possible non-Gaussianity in GDP growth rates by assuming different distributional assumptions in the GARCH models. In this respect we assume, together with the classical Normal distribution, the t-student (Bollerslev, 1987) and the General Error Distribution (GED) (Nelson, 1991; Mattera and Giacalone, 2018; Giacalone, Mattera and Cozzucoli, 2019; Cerqueti, Giacalone and Panarello, 2019; Cerqueti, Giacalone and Mattera, 2020).

The last point of our innovation is to consider the predictive ability of the Generalized Autoregressive Score model of Creal, Koopman and Lucas (2013) and Harvey (2013), a very general statistical model that considers the score function of the predictive model density as the driving mechanism for time-varying parameters. A wide class of GARCH-type processes are special cases of the Generalized Autoregressive Sscore (GAS) model (Creal *et al.*, 2013).

Since the GAS model is based on the score, it exploits the data's complete density structure rather than just a few moments.

To establish the superiority of the score-driven approach with respect the classical GARCH processes, we provide an empirical analysis about the business cycles of the United States and Japan.

What we found is that the score-driven models provide better out-of-sample performances with respect both Gaussian and non-Gaussian GARCH processes in forecasting business cycle volatility.

The structure of the paper is as it follows. In Section 2 are going to be discussed the statistical models implemented in the paper. Then, in Section 3 the data and the methodology for out-of-sample comparisons is presented, as well as the results. In last paragraph, some final remarks are analyzed and discussed.

2. STATISTICAL MODELS

In what follows we present the statistical models implemented in the paper. Briefly, we compare both Gaussian and non-Gaussian GARCH processes, presented in the Section 2.1, with the score-driven model (the GAS) presented in the Section 2.2.

2.1 *GARCH processes*

The first class of stochastic processes that we use for modeling business cycle volatility are the usual ARCH and GARCH processes of Engle (1982) and Bollerslev (1986).

First of all, following most of previous studies on business cycle volatility modeling briefly presented in the Introduction, we assume general framework in which the Gaussian distribution assumption is not violated. Then, we relax this assumption by considering alternative extensions of the GARCH models in the context of non-Normal distributions (Wilhelmsson, 2006; Mattera and Giacalone, 2018; Cerqueti *et al.*, 2020).

Given a time series y_t, we can formalize the standard $GARCH(p,q)$ model for the volatility of y_t $(\sigma_t^2 : t \geq 0)$ as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i z_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{1}$$

with $\omega > 0$ and $\alpha_i > 0, \beta_j > 0$, for each $i = 1, \ldots, p$ and $j = 1, \ldots, q$. The positivity condition on ω, the α'a and β's ensures the positivity of the variance. The term $(z_t : t \geq 0)$ is given by the product of squared observations of y_t and an usually Gaussian stochastic process with i.i.d. time-realizations.

However, as many studies have claimed, business cycles are proved to be asymmetric. Therefore, to account for the asymmetry, a common extended GARCH model that introduces an additional term for the asymmetry is the GJR-GARCH, proposed by Glosten *et al.*, (1993):

$$\sigma_t^2 = \omega + \alpha z_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma z_{t-1}^2 I(z_{t-1} < 0) \tag{2}$$

where I is an indicator function that, with $\gamma = 0$, assumes a symmetric variance response to a past shock (z_{t-1}), whether the shock is positive or negative.

Other GARCH models that account for asymmetry are the Exponential GARCH model of Nelson (1991) and the Treshold GARCH model of Zakoian (1994). However, the Exponential GARCH assumes non linearity in the volatility process and this is the reason why it is widely used in business cycle volatility modeling.

Moreover, the GARCH process as in (1) can be also extended to account for highly persistence in conditional variances, a well documented fact for business cycles.

Indeed, in order to ensure the existence of the unconditional variance, in the standard GARCH setting we know that one needs $\alpha + \beta < 1$.

However, by imposing the restriction that $\alpha + \beta = 1$ in the Equation (1), we obtain the Integrated GARCH (I-GARCH), by analogy with the unit root literature:

$$\sigma_t^2 = \omega + \alpha(z_{t-1}^2 - \sigma_{t-1}^2) + \sigma_{t-1}^2. \tag{3}$$

Finally, another important extension – which is mainly related to non linearity – is the Asymmetric Power Autoregressive Conditional Heteroskedasticity (AP-ARCH) proposed by Ding *et al.*, (1993) and Ding and Granger (1996):

$$\sigma_t^\delta = \omega + \alpha(|z_{t-1}| - \lambda z_{t-1})^{2\delta} + \beta \sigma_{t-1}^\delta \tag{4}$$

where $\delta > 0,\ \omega > 0,\ \alpha > 0,\ \beta \geq 0$ and $|\lambda| \leq 1$, hence being nonlinear in parameters. This is a very general model and includes, most of the other that we have mentioned.

As we have said, all these basic model specifications assume a Gaussian distribution for the innovation term included in z_t^2. However, due to the complexity of the real economic system, in most of the empirical applications it more reasonable to assume non-Gaussian innovations (Bollerslev, 1987).

A very general and flexible family of distribution is represented by the Generalized Error Distribution (also called Exponential Power Function).

The GED random variable X has the following probability density function (Giacalone, Panarello and Mattera, 2018a):

$$f(z;\mu,\sigma,p) = \frac{p\exp(\frac{1}{2}\mid\frac{z-\mu_p}{\sigma_p}\mid^p)}{2p^{(1+\frac{1}{p})}\sigma\Gamma(\frac{1}{p})} \tag{5}$$

where $\mu_p \in (-\infty, +\infty)$ is called location parameter, $\sigma_p > 0$ is called scale parameter, $p > 0$ is a measure of fatness of tails and is called shape parameter (Zhu and Zinde-Walsh, 2009) and:

$$\Gamma(a) = \int_0^\infty x^{a-1}\exp(-x)dx. \tag{6}$$

is the Gamma function. Since the distribution is symmetric and unimodal, the location parameter is also the mode, median and mean of the distribution (see Figure 1).

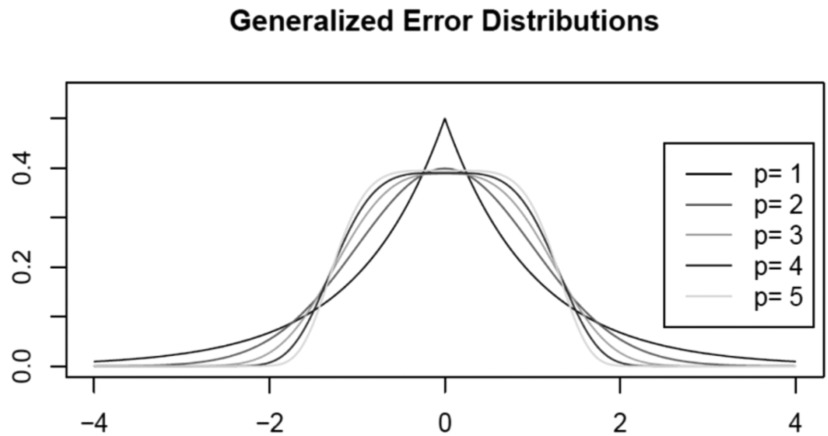

FIGURE 1. - *Generalized Error Distribution for different values of skewness*

The variance and kurtosis are given by (Giacalone, 2020):

$$Var(X) = \sigma^2 2^{\frac{2}{p}} \frac{\Gamma(3p^{-1})}{\Gamma(p^{-1})} \tag{7}$$

$$Ku(X) = \frac{\Gamma(5p^{-1})}{\Gamma(3p^{-1})}\frac{\Gamma(p^{-1})}{\Gamma(3p^{-1})} \tag{8}$$

The Gaussian distribution is a special case of the GED when $p = 2$, and when $p < 2$ the GED distribution has fatter tails than a Gaussian one (Cerqueti *et al.*, 2019).

A very important feature of this family of distributions, that has been proved to be useful in modeling stock market volatility (Giacalone *et al.*, 2019; Trucíos, 2019; Cerqueti *et al.*, 2020), is that it includes also other common distributions, for different values of shape parameter p. In particular, when $p = 1$ we have a Laplace distribution, and for $p = +\infty$ we have the Uniform distribution (Giacalone *et al.*, 2018a).

However, empirical evidence suggests that financial returns exhibit a negative symmetry in distribution; thus, we here propose to use skewed distribution in GARCH modeling (Zhu and Zinde-Walsh, 2009).

In this respect, we can hypothetically use either the Skew Normal or the Skew t distributions. Nevertheless, despite there are several alternatives of estimating and modeling skewness and kurtosis (Theodossiou, 1998; Zenga, Polisicchio and Greselin, 2004), according to the discussion above a very interesting extension for skewness is the Skewed-GED distribution, which can be derived in order to take into account for the skewness and the leptokurtosis (see Figure 2).

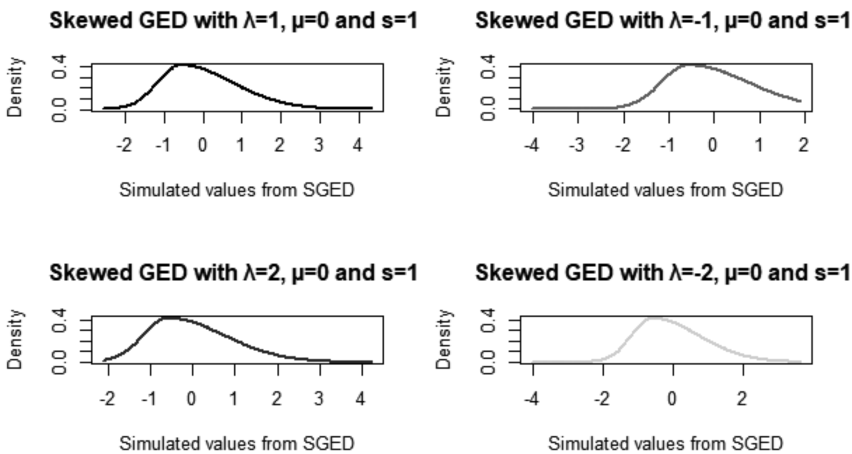

FIGURE 2. - *Skewed Generalized Error Distribution for different values of skewness*

The probability density function for non-centered Skewed GED can be defined as follows (Cerqueti *et al.*, 2020):

$$f(z;\mu,\sigma,\lambda,p) = \frac{p\text{ex p}\left(-\frac{1}{p}\left|\frac{z-\mu_p+m}{\nu\sigma_p(1+\lambda\text{sign}(z-\mu+m))}\right|^p\right)}{2\nu\sigma\Gamma(\frac{1}{p})} \tag{9}$$

where $z \in \mathbf{R}$, μ_p is the location parameter, σ_p the scale parameter, λ the skewness parameter and p the shape parameter, while Γ is as in (6). Function sign is the sign function which assumes value of -1 for negative values and 1 for positive ones. Moreover, m is defined as follow:

$$m = \frac{2^{\frac{2}{p}}\nu\sigma\lambda\Gamma(\frac{1}{2}+\frac{1}{p})}{\sqrt{\pi}},$$

while ν:

$$\nu = \frac{\pi(1+3\lambda^2)\Gamma(\frac{3}{p}) - 16^{\frac{1}{p}}\lambda^2\Gamma(\frac{1}{2}+\frac{1}{p})\Gamma(\frac{1}{p})}{\pi\Gamma(\frac{1}{p})}.$$

The shape parameter p controls the tails and the peak of the distribution; a small value of p means that the tails of the distribution become flat, with the center becoming largely peaked.

The skewness parameter λ ranges in $[-1, 1]$; in the case of negative skewness ($\lambda < 0$) the density function is skewed to the left and vice versa for $\lambda > 0$.

Also the Skewed GED is a very special case of other distributions. For example, supposing $\lambda = 0$ (allowing p to change) we can obtain a wide family of non-skewed distributions (Zhu and Zinde-Walsh, 2009; Theodossiou and Savva, 2016).

In particular, when $\lambda = 0$ we have the GED, while $\lambda = 0$ and $p = 2$ means Normal distribution, $\lambda = 0$ and $p = \infty$ is the Uniform distribution and $\lambda = 2$ and $p = 2$ is the skewed Normal (Zhu and Zinde-Walsh, 2009; Cerqueti *et al.*, 2020).

Also for the SGED-GARCH model (based on Skewed GED), the specification is the same as (1) but in this case we suppose that z_t follows a Skewed GED. The parameter estimation for the GED-type GARCH processes is based on Maximum Likelihood method (Wiśniewska and Wyłomańska, 2017).

Therefore, because of the generality of these two distributions, we will explore below the empirical effectiveness of the GED and its extension for skewness when predicting volatility of business cycles by means of GARCH models.

2.2 *Generalized Autoregressive Score*

Another interesting approach for modeling business cycle volatility can be found in the GAS model of Creal *et al.* (2013) and Harvey (2013). It is a very general statistical model that considers the score function of the predictive model density as the driving mechanism for time-varying parameters.

We have to highlight that a wide class of GARCH-type processes are special cases of the GAS model (Creal *et al.*, 2013).

Since the GAS model is based on the score, it exploits the data's complete density structure rather than just a few moments.

The GAS model can be formalized as follows. Let be y_t a time series generated by the following observation conditional density $p(\cdot)$:

$$y_t \sim p(y_t|f_t, F_t; \theta), \tag{10}$$

where f_t is a vector of time-varying parameters at time t, F_t is the available information at time t and θ a vector of static parameters.

The length of the vector f_t crucially depends by the assumption we make about the density (10). As an example, if we specify a Gaussian density, such that $p \sim N(\mu, \sigma^2)$, we have that $f_t = (\mu_t, \sigma_t^2)$ but with different densities we could obtain more time varying parameters.

The available information F_t, is defined as a collection of the past realizations of the time series y_t and of its time varying parameters f_t.

Given two integers $0 \leq p, q \leq T-1$, it is possible to express the GAS of order p and q, $GAS(p, q)$ can be written as:

$$f_t = \omega + \sum_{i=1}^{n} \mathbf{A}_i s_{t-i} + \sum_{j=1}^{m} \mathbf{B}_j f_{t-j} \tag{11}$$

where ω is a real vector and the **A**'s and the **B**'s are real matrices with an appropriate dimension.

The variable s_t is the *scaled* score of the conditional distribution (10), and it is a function of the data and the parameters and it is equal to:

$$s_t = S_t \cdot \nabla_t \tag{12}$$

where $S_t = S_t(f_t, \mathcal{F}_t; \theta)$ is a positive definite scaling matrix known at time t and $\nabla_t(y_t, f_t, F_t; \theta)$ is the score of y_t evaluated with respect to f_t, i.e.:

$$\nabla_t = \frac{\partial \log p(y_t | f_t, F_t; \theta)}{\partial f_t} \tag{13}$$

A common approach of scaling (Cerqueti, Giacalone and Mattera, 2021) is to consider the inverse of the information matrix of f_t to a power $\gamma \geq 0$ as the scaling matrix:

$$S_t = E_{t-1}\left[\nabla_t \nabla_t'\right]^{-\gamma} \tag{14}$$

where the conditional score ∇_t has been defined in (13).

The parameter γ usually takes value in the set $\{0, \frac{1}{2}, 1\}$. When $\gamma = 0$, then S_t is the identity matrix and there is no scaling. Differently, if $\gamma = 1$, then the conditional score ∇_t is premultiplied by the inverse to obtain (14) while, if $\gamma = \frac{1}{2}$, ∇_t is scaled to its square-root.

As in the case of z_t in the previous Section, the most simple assumption about the predictive density (10) is the Gaussian distribution.

Therefore, assuming $y_t \sim N(\mu_t, \sigma_t^2)$:

$$p(y_t | f_t, \mathcal{F}_t; \theta) = \frac{1}{\sigma_t \sqrt{2\pi}} e^{-(y_t - \mu_t)^2 / 2\sigma_t^2}$$

we have that $f_t = (\mu_t, \sigma_t^2)$. In the case of a Gaussian-GAS(1,1) process, the updating mechanism for time varying parameters μ_t and σ_t^2 could be specified as follows:

$$f_t = \omega + \mathbf{A} s_{t-1} + \mathbf{B} f_{t-1}$$

where θ is the vector containing all the scalar parameters in $\omega, \mathbf{A}, \mathbf{B}$ and s_t is scaled by (14) setting $\gamma = 1$. In particular, the conditional score vectors are in this case given by:

$$\nabla_t^{(\mu)} = \frac{(y_t - \mu_t)}{\sigma_t^2}$$

$$\nabla_t^{(\sigma)} = \frac{(y_t - \mu_t)^2}{2\sigma_t^4} - \frac{T}{2\sigma_t^2}$$

However is common to assume a density different from the Gaussian one. In order to account for non-normality, one can assume that y_t follows a t-student distribution with location μ_t, scale ϕ_t and degrees of freedom $v_t > 2$ with its density given by:

$$p(y_t|f_t, \mathcal{F}_t; \theta) = \frac{\Gamma\left(\frac{v_t+1}{2}\right)}{\Gamma\left(\frac{v_t}{2}\right)\phi_t\sqrt{\pi v_t}}\left(1 + \frac{(y_t - \mu_t)^2}{v_t\phi_t}\right)^{\frac{v_t+1}{2}}$$

Such that we have $f_t = (\mu_t, \phi_t, v_t)$.

In the case of a t-GAS(1,1), the updating mechanism for f_t is the same as the Gaussian density but with the conditional score vectors that now are equal to:

$$\nabla_t^{(\mu)} = \frac{(v_t + 1)(y_t - \mu_t)}{v_t\phi_t + (y_t - \mu_t)^2}$$

$$\nabla_t^{(\phi)} = \frac{1}{2\phi_t}\left[\frac{(v_t + 1)(y_t - \mu_t)^2}{v_t\phi_t + (y_t - \mu_t)^2} - 1\right]$$

$$\nabla_t^{(v)} = \frac{1}{2}\left\{\psi\left(\frac{v_t + 1}{2}\right) - \psi\left(\frac{v_t}{2}\right) - \frac{1}{v_t} - \log\left(1 + \frac{(y_t - \mu_t)^2}{v_t\phi_t}\right) + \frac{(v_t + 1)(y_t - \mu_t)^2}{v_t\left[v_t\phi_t + (y_t - \mu_t)^2\right]}\right\}$$

where $\psi(\cdot)$ is the Digamma function. This model has also been defined as Beta-t-EGARCH model by Harvey (2013) and Harvey and Sucarrat (2014).

In the end, we also specify a Skew-t distribution in order to evaluate the performances of a skew non Gaussian assumption within the class of the *score-driven* models.

3. EMPIRICAL ANALYSIS

3.1 *Data*

In this paper we use the data on the quarterly real GDP in the US and Japan. The source is the Federal Reserve Economic Database (FRED)[1], that is an online data-

[1] The data are available at the following link: //https://fred.stlouisfed.org/tags/series

base consisting of economic data time series from scores of national, international, public, and private sources. It contains many macroeconomic data of different types.

The sample period for the United Stated is the first quarter of 1947 through the first quarter of 2020. In the case of Japan, instead, we collected data from the first quarter of 1994 up to the first one of the 2020 (see Figure 3).

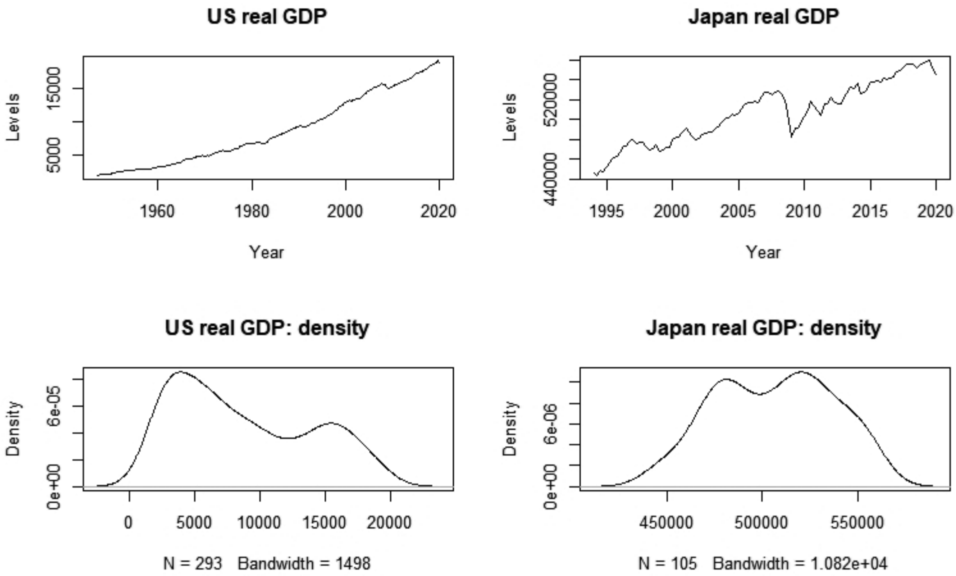

FIGURE 3. - *Real Gross Domestic Product: time series and densities*

Both time series are seasonally adjusted, since it is known that seasonal adjustment improves forecasts accuracy (Giacalone, Mattera and Nissi, 2020). The data's main descriptive statistics are reported in the Table 1. The values showed in the Table 1 are in terms of billions US dollars for United States and in Japanese yen billions for the Japan.

TABLE 1. - *Descriptive statistics: real GDPs*

Statistics	N	Mean	St. Dev.	Min	Max
Real GDP US	296	8817.810	5246.427	2023.452	19253.960
Real GDP Japan	108	506688.500	30312.770	443805.100	559066.900

Then the business cycle has to be estimated for both time series.

The decomposition of macroeconomic time series into trend and cyclical components is crucial for macroeconomic modeling. The Hodrick and Prescott (1997) filter (HP) is widely used for detrending macroeconomic time series.

Like other trend removal techniques, such as trend regression, moving average detrending, and band-pass filtering, the HP filter is often used to produce new time series that reflects cyclical fluctuations.

The HP filter is based on the following minimization problem:

$$\min_{y_t^T} \sum_{t=1}^{T} \left(y_t - y_t^T\right)^2 + \lambda \sum_{t=2}^{T} \left(\left(y_t^T - y_{t-1}^T\right) - \left(y_{t-1}^T - y_{t-2}^T\right)\right)^2 \tag{15}$$

where y_t^T is the trend. Therefore the cyclical component y_t^C is equal to:

$$y_t^C = y_t - y_t^T \tag{16}$$

The business cycles of US and Japan estimated by means of the HP filter are shown in the Figure 4.

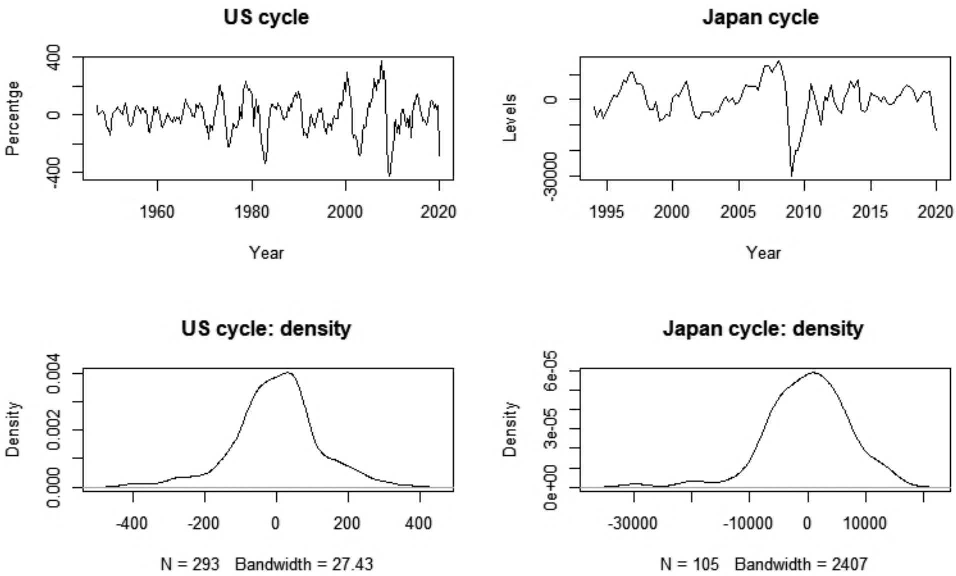

FIGURE 4. - *Business cycles: time series and densities*

The descriptive statistics associated to the estimated business cycles are showed in the Table 2.

TABLE 2. - *Descriptive statistics: estimated business cycles*

Statistics	N	Mean	St. Dev.	Min	Max
US cycle (HP)	296	0.000	159.014	−1563.944	476.295
JPN cycle (HP)	108	0.000	7103.394	−29795.080	15093.220

Exactly like previous Table 1, also the values showed in the Table 2 are in terms of billions US dollars for United States and in Japanese yen billions for the Japan.

As the Figure 4 shows, the empirical densities of the business cycles are not normally distributed. This evidence justifies the non-Gaussian modeling approaches, within the GARCH framework and the score-driven one, discussed above.

3.2 *Out-of-sample analysis*

The statistical methods evaluation is conducted in out-of-sample and at this aim we take advantage of a rolling-window approach.

More in details, we first consider an out-of-sample window for accuracy testing equal to the last 20% of the observations within the initial sample.

Then, we set an initial window equal to w such that we estimate the statistical model parameters with an estimation window equal to $t = 1, \ldots, w$. Hence we use these estimates to get the point forecast at time $t + 1$.

Then, we repeat the operation by eliminating from the sample the first observation $t = 1$ and including the newest one $t = w + 1$. In other words, in the second step the estimation window is equal to $t = 2, \ldots, w + 1$. The rolling window operation is repeated until we reach the last point of the sample $w = T$.

The result of this operation is a time series of forecasts for the j-th model, namely $\hat{y}_{j,t}$ of length $T - w$. In order to evaluate the out-of-sample accuracy of each j-th statistical model, we compute the following loss function:

$$L_j\left(\hat{y}_{j,t}, y_t\right) = g\left(\hat{y}_{j,t} - y_t\right) \tag{17}$$

where $\hat{y}_{j,t}$ is the forecast of the j-th statistical model for the time series y_t and $g(\cdot)$ can by any loss function. In this paper, following previous studies (Wilhelmsson, 2006; Giacalone, Cusatelli, Romano, Buondonno and Santarcangelo, 2018b), we compute the Mean Square Forecast Error (MSFE):

$$\text{MSFE}_j = \frac{\sum_n \left(\hat{y}_{j,t} - y_t\right)^2}{n} \tag{18}$$

and the Mean Absolute Forecast Error (MAFE):

$$\text{MAFE}_j = \frac{\sum_n \left|\hat{y}_{j,t} - y_t\right|}{n} \tag{19}$$

that is more robust to outliers. However, in many practical applications it is not easy to understand if there is a single model that significantly dominates all the others. This happen because often the data are not sufficiently informative to give an unequivocal solution.

To overcome this issue, we take advantage of the Model Confidence Set (MCS) procedure developed by Hansen, Lunde and Nason (2003, 2011).

The MCS procedure is based on a set of subsequent tests. More in details, it is based on an equivalence test, δ_M, and an elimination rule, e_M. The equivalence test is applied to the the initial set of statistical models under comparison. If δ_M is rejected, there is evidence that the all objects in the set are not equally ‘‘good’’ and the elimination rule e_M is used to delete from the initial set of models the one with the poorest performances.

This procedure is repeated until the equivalence test δ_M is not rejected and, therefore, all the models ‘‘are equally good’’. This equivalence ensures that a *superior set* of models has been found.

As the equivalence test we take advantage of the Diebold and Mariano (2002) test. Considering a generic loss function as $L_j(\hat{y}_{j,t}, y_t)$, we can compute the loss difference of two competing models i and j:

$$d_{ij,t} = L_j(\hat{y}_{j,t}, y_t) - L_i(\hat{y}_{i,t}, y_t) \tag{20}$$

Testing the equal predictive accuracy (EPA) assumption for the two competing models i and j means to consider a test with the following null hypothesis:

$$H_0 : E(d_{ij,t}) = 0$$

and with the following alternative one:

$$H_1 : E(d_{ij,t}) \neq 0$$

The test statistics is (Diebold and Mariano, 2002; Hansen *et al.*, 2011):

$$z_{ij} = \frac{\sqrt{T}\bar{d}_{ij}}{\sqrt{\hat{var}(\bar{d}_{ij})}} \tag{21}$$

where $\bar{d}_{ij} = T^{-1}\sum_{t=1}^{T} d_{ij,t}$ and $\hat{var}(\bar{d}_{ij})$ is the estimated variance. Since we employ a Diebold Mariano-like test for δ_M, following Hansen *et al.* (2011), we choose the following appropriate elimination rule e_M:

$$e_M = \arg\max_i \sup_j z_{ij} = \arg\max_i \left\{ \sup_j \frac{\bar{d}_{ij}}{\sqrt{\hat{var}(\bar{d}_{ij})}} \right\} \tag{22}$$

In the following a summarizing table (Table 3) with the list of the estimated statistical models is provided.

TABLE 3. - *List of the 15 estimated statistical models*

Model	Authors	Distributions
GARCH	Bollerslev (1986)	Gaussian, GED, Skew GED
I-GARCH	Engle and Bollerslev (1986)	Gaussian, GED, Skew GED
GJR-GARCH	Glosten et al. (1993)	Gaussian, GED, Skew GED
E-GARCH	Nelson (1991)	Gaussian, GED, Skew GED
GAS	Creal et al. (2013)	Gaussian, Skew-t
Beta-t-E-GARCH	Harvey and Sucarrat (2014)	t-student

3.3 *Model Confidence Set: results*

In what follows we show the out-of-sample comparisons of the different 15 statistical models (see Table 3). Because we are comparing a lot of models, we take advantage of the Model Confidence Set procedure of Hansen *et al.*, (2003, 2011) as previously mentioned. The Table 4 shows the Model Confidence Set results for the US business cycle volatility by assuming the robust MAFE loss function.

TABLE 4. - *Model Confidence Set: superior set of models for US business cycle volatility - MAFE loss*

Model	e_M	MCS p-value	MAFE	Rank
Skew-t-GAS	-1.0945	1.000	0.2740	1
Beta-t-E-GARCH	1.0945	0.620	0.2806	2
Gaussian-GAS	1.3075	**0.243**	0.3539	3

Over 15 models only 3 of them are placed in the *superior set* and all of them are *score-driven models*. Among the score-driven models, the most accurate forecasts are reached with the Skew-t-GAS models while the worst with the Gaussian-GAS that, however, still is in the superior set. All the GARCH-processes are excluded from the superior set due to their poor forecasting performances. The Table 5 shows, instead, the results assuming the MSFE as loss function.

TABLE 5. - *Model Confidence Set: superior set of models for US business cycle volatility - MSFE loss*

Model	e_M	MCS p-value	MSFE	Rank
Skew-t-GAS	-1.105476	1.0000	0.00247	1
Beta-t-E-GARCH	1.105476	0.6066	0.00464	2
Gaussian-GAS	1.321406	**0.2312**	0.01283	3

Clearly, the resulting models' ranking is the same. In other words, we can argue that the superior performances of the score-driven models are in this case robust to the selection of the loss function.

Therefore, we can also conclude that the score-driven approach, due to its flexibility, is more able to forecasts the US business cycle volatility.

Let now analyze the case of Japanese business cycle (also shown in Figure 4). The results assuming the more robust MAFE error function are shown in the Table 6. In the case of Japan we have several GARCH-type models that are placed within the superior set *with* the score-driven models.

Indeed, Table 6 shows interesting results. First of all, almost all of the Gaussian

GARCH processes have been excluded for the superior set of models, while we have that assuming an underlying Skew Generalized Error Distribution (Skew GED) distributional assumption ensures the highest out-of-sample performances.

TABLE 6. - *Model Confidence Set: superior set of models for Japan business cycle volatility - MAFE loss*

Model	e_M	MCS p-value	MAFE	Rank
Beta-t-E-GARCH	-0.7476642	1.0000	0.090737	1
Skew-t-GAS	0.7476642	0.9814	0.097248	2
Gaussian-GAS	1.3862454	0.7132	0.105817	3
GED-E-GARCH	1.5341423	0.5722	0.105824	4
Gaussian-E-GARCH	1.5833124	0.5368	0.105825	5
Skew GED-E-GARCH	1.7096317	0.4430	0.105828	6
Skew GED-I-GARCH	1.8480747	0.3382	0.105829	7
Skew GED-GARCH	1.9536345	**0.2790**	0.105829	8

More in details, in the Table 6 we have that the 60% of the GARCH process within the superior set assuming a Skew GED underlying distribution for the innovation term. Moreover, the only standard GARCH that is placed within the superior set is also based on the same distribution. This evidence is in line with the findings for different assets in the stock market (Wilhelmsson, 2006; Mattera and Giacalone, 2018; Giacalone *et al.* 2019; Trucíos, 2019; Cerqueti *et al.* 2020).

Regardless the distributional assumption, we have that all the Exponential GARCH processes are placed within the superior set. This result, indeed, is also in line within previous findings with respect the Japanese business cycle (Hamori, 2000; Ho and Tsui, 2003; Hai *et al.* 2013).

Moreover, the Table 6 confirms that the score-driven models are superior to the GARCH ones, exactly as in the empirical evidence for the United Stated.

However, differently from the US, the best score-driven model for the Japanese business cycle volatility is the Beta-t-E-GARCH model of Harvey and Sucarrat (2014), highly nonlinear, asymmetric and non-Gaussian process. On the other hand, the Gaussian-GAS is again the score-driven model with the poorest performances. These results clearly reflect the non-Gaussianity of the Japanese business cycle.

Then, we also evaluate the performances of the models assuming a MSFE function for the prediction error (see Table 7).

TABLE 7. - *Model Confidence Set: superior set of models for Japan business cycle volatility - MSFE loss*

Model	e_M	MCS p-value	MSFE	Rank
Beta-t-E-GARCH	-1.377183	**1.000**	0.00213	1

In terms of MSFE the Model Confidence Set procedure highlights a single model within the *superior set*, meaning that this statistical model dominates all the others. More in details, once again the best identified model is a *score-driven* one, particularly the beta-t-E-GARCH of Harvey and Sucarrat (2014).

Indeed, in both US and Japanese cases we get enough evidence of score-driven models superiority in forecasting business cycles volatility.

4. CONCLUSIONS

What has been trying to demonstrate through this paper is a contribution to the existing literature showing some innovative results in the business cycle volatility forecasting topic.

In the first part of this manuscript we examined how the previous literature studied the volatility of the business cycles. Some of the contributions have tried to test the hypothesis that volatility is asymmetric, time-varying and would be lower when GDP grows and higher when it contracts (Hamori, 2000). Other studies have used different ARCH and GARCH models to predict business cycle volatility. In this first analysis it emerged that, if forecasting business cycles is a difficult task, the nature of its volatility makes the task even more difficult.

Indeed, in the economic context of business cycle volatility, the forecasting activity is very important for a variety of audience such as investors, government regulators and capital markets.

Our contribution has been to predict volatility with different score-driven models, including those of the GARCH type, in order to establish the best statistical model to describe dynamic volatility in business cycles.

The paper starts with analyzing the literature on the volatility models of asset returns considering the ARCH model (Engle, 1982) and the GARCH (Bollerslev, 1986).

In the introduction and in the subsequent paragraphs, the various forms of GARCH models are specified, which have constituted the key of solving most of the problems in the macroeconomic volatility literature.

Moreover, particular attention has been given to the comparison between Gaussian and non-Gaussian GARCH processes and in this perspective the importance of the GED-GARCH model (based on the Generalized Error Distribution) and of the SGED-GARCH one (based on the Skewed GED) has been emphasized as valid tools for forecasting the volatility of business cycles. For the last models, the estimation of the parameters with the maximum likelihood method is the procedure suggested in literature.

The latter methods offer a wide range of models to choose from and naturally the question arises: What is the best volatility model?

Unfortunately, this question is not easy to answer because asset returns do not provide anyway enough information to identify a single volatility pattern as ‘‘the best one’’.

Our paper tries to answer this question. In the Section two of the third paragraph we characterize the volatility patterns of business cycles that significantly dominate the others, in an out-of-the-sample context.

In the paragraph three the score-driven approach for evaluating the predictions of volatility models, i.e. the model confidence set (MCS) method developed in Hansen *et al.* (2003, 2011), has been introduced.

Finally we applied the Model Confidence Set to the 15 volatility models (considering both the Gaussian and the non-Gaussian approaches for the GARCH and the score-driven models) to the real GDP quartery data in the United States and in Japan.

We considered all available data up to the first quarter of 2020 to avoid data disruption due to the spread of the COVID-19 pandemic.

The results obtained from the application of the Model Confidence Sets are very interesting and extremely positive and narrow our attention from the first 15 models considered to the 3 models that presumably represent a better way to capture the volatility forecasting on the analyzed data.

We can observe that, among the GARCH models proposed in this paper, on the basis of distributional approach we emphasized the important role of the Generalized Error Distribution in the symmetric and in the asymmetric form in the case of Non-Gaussian volatility from the point of view of the loss functions.

However, we have shown that the score-driven approaches are even more accurate than non-Gaussian GARCH processes in forecasting business cycles volatility.

Indeed, the Model Confident Set results of the empirical analysis conducted on the business cycles of the United States and Japan provided enough evidence in favor of the superiority of the score-driven approaches.

Moreover, in the case of Japan, it has been possible to highlight a single model within the superior set that dominates all the other 14 models justifying the score-driven approach followed in the paper.

Finally, the Beta-t-E-GARCH, that is a highly nonlinear and non-Gaussian volatility process, allows us to conclude the evidence of the score-driven models best performance in forecast volatility with regard to business cycles, respect to the other models considered by the literature and studied in the paper.

REFERENCES

Baele L., Bekaert G., Inghelbrecht K. (2010). The determinants of stock and bond return comovements. *The Review of Financial Studies*, **23**(6), 2374-2428.

Bodman P. (2009). Output volatility in Australia. *Applied Economics*, **41**(24), 3117-3129.

Bollerslev T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**(3), 307-327.

Bollerslev T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 542-547.

Cerqueti R., Giacalone M., Panarello D. (2019). A generalized error distribution copula-based method for portfolios risk assessment. *Physica A: Statistical Mechanics and its Applications*, **524**, 687-695.

Cerqueti R., Giacalone M., Mattera R. (2020). Skewed non-gaussian GARCH models for cryptocurrencies volatility modelling. *Information Sciences*, **527**, 1-26.

Cerqueti R., Giacalone M., Mattera R. (2021). Model-based fuzzy time series clustering of conditional higher moments. *International Journal of Approximate Reasoning*, **134**, 34-52.

Clark T.E., Ravazzolo F. (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, **30**(4), 551-575.

Creal D., Koopman S.J. Lucas A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, **28**(5), 777-795.

Diebold F.X., Mariano R.S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics*, **20**(1), 134-144.

Diebold F.X., Yilmaz K. (2009). Measuring financial asset return and volatility spillovers, with application to global equity markets. *The Economic Journal1,* **119**(534), 158-171.

Ding Z., Granger C.W., Engle R.F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, **1**(1), 83-106.

Ding Z., Granger C.W. (1996). Modeling volatility persistence of speculative returns: a new approach. *Journal of Econometrics*, **73**(1), 185-215.

Engle R.F. (1982.) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987-1007.

Engle R.F., Bollerslev T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, **5**(1), 1-50.

Fang W., Miller S.M., Lee C. (2008). Cross-country evidence on output growth volatility: Nonstationary variance and GARCH models. *Scottish Journal of Political Economy*, **55**(4), 509-541.

Fang W., Miller S.M. (2009). Modeling the volatility of real GDP growth: The case of Japan revisited. *Japan and the World Economy*, **23**(3), 312-324.

Fiori A.M., Zenga M. (2009). Karl Pearson and the origin of kurtosis. *International Statistical Review*, **77**(1), 40-50.

Giacalone M. (2020). A combined method based on kurtosis indexes for estimating ‘‘p’’ in non-linear Lp-norm regression. *Sustainable Futures*, **2**, 100008.

Giacalone M., Cusatelli C., Romano A., Buondonno A., Santarcangelo V. (2018b). Big Data and forensics: An innovative approach for a predictable jurisprudence. *Information Sciences*, **426**, 160-170.

Giacalone M., Mattera R., Cozzucoli P.C. (2019). Improving volatility forecasts with GED-GARCH model: Evidence from us stock market. *The Empirical Economics Letters*, **18**(7), 785-791.

Giacalone M., Mattera R., Nissi E. (2020). Economic indicators forecasting in presence of seasonal patterns: time series revision and prediction accuracy. *Quality & Quantity*, **54**(1), 67-84.

Giacalone M., Panarello D., Mattera R. (2018a). Multicollinearity in regression: an efficiency comparison between Lp-norm and least squares estimators. *Quality & Quantity*, **52**(4), 1831-1859.

Glosten L.R., Jagannathan R., Runkle D.E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, **48**(5), 1779-1801.

Hai V.T., Tsui A.K., Zhang Z. (2013). Measuring asymmetry and persistence in conditional volatility in real output: evidence from three east Asian tigers using a multivariate GARCH approach. *Applied Economics*, **45**(20), 2909-2914.

Hamori S. (2000). Volatility of real GDP: Some evidence from the United States, the United Kingdom and Japan. *Japan and the World Economy*, **12**(2), 143-152.

Hansen P.R., Lunde A., Nason J.M. (2003). Choosing the best volatility models: the model confidence set approach. *Oxford Bulletin of Economics and Statistics*, **65**, 839-861.

Hansen P.R., Lunde A., Nason J.M. (2011). The model confidence set. *Econometrica*, **79**(2), 453-497.

Harvey A. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*, Cambridge University Press.

Harvey A., Sucarrat G. (2014). EGARCH models with fat tails, skewness and leverage. *Computational Statistics & Data Analysis*, **76**, 320-338.

Ho K.Y., Tsui A.K. (2003). Asymmetric volatility of real GDP: Some evidence from Canada, Japan, the United Kingdom and the United States. *Japan and the World Economy*, **15**(4), 437-445.

Ho K.Y., Albert K. (2004). Analysis of real GDP growth rates of greater China: an asymmetric conditional volatility approach. *China Economic Review*, **15**(4), 424-442.

Hodrick R.J., Prescott E.C. (1997). Postwar us business cycles: an empirical investigation. *Journal of Money, Credit, and Banking*, 1-16.

Köchling G., Schmidtke P., Posch P.N. (2020). Volatility forecasting accuracy for Bitcoin. *Economics Letters*, 191, 108836.

Lettau M., Ludvigson S.C., Watcher J.A. (2008). The declining equity premium: What role does macroeconomic risk play? *The Review of Financial Studies*, **21**(4), 1653-1687.

Mattera R., Giacalone M. (2018). Alternative distribution based GARCH models for Bitcoin volatility estimation. *The Empirical Economics Letters*, **17**(11), 1283-1288.

Nelson D.B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347-370.

Schwert G.W. (1989). Why does stock market volatility change over time? *The Journal of Finance*, **44**(5), 1115-1153.

Stock J.H., Watson. M.W. (2007). Why has us inflation become harder to forecast? *Journal of Money, Credit and Banking*, **39**, 3-33.

Theodossiou P. (1998). Financial data and the Skewed Generalized t Distribution. *Management Science*, **44**(12-part-1), 1650-1661.

Theodossiou P., Savva C.S. (2016). Skewness and the relation between risk and return. *Management Science*, **62**(6), 1598-1609.

Trucíos C. (2019). Forecasting Bitcoin risk measures: A robust approach. *International Journal of Forecasting*, **35**(3), 836-847.

Wilhelmsson A. (2006). GARCH forecasting performance under different distribution assumptions. *Journal of Forecasting*, **25**(8), 561-578.

Wiśniewska M., Wyłomańska A. (2017). GARCH process with GED distribution. *Cyclostationarity: Theory and Methods III*, 83-103. Springer.

Zakoian J.M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, **18**(5), 931-955.

Zenga M., Polisicchio M., Greselin F. (2004). The variance of Gini's mean difference and its estimators. *Statistica*, **64**(3), 455-475.

Zhu D., Zinde-Walsh V. (2009). Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics*, **148**(1), 86-99.