Semi-Nonparametric Generalized Autoregressive Conditional Heteroscedasticity Model with Application to Bitcoin Volatility Estimation[1]

Juri Trifonov[1], **Bogdan Potanin**[2]

[1] National Research University Higher School of Economics,
11, Pokrovsky Blvd, Moscow, 109028, Russian Federation.
E-mail: yutrifonov@hse.ru

[2] National Research University Higher School of Economics
11, Pokrovsky Blvd, Moscow, 109028, Russian Federation.
E-mail: bogdanpotanin@gmail.com

This study raises the problem of modeling conditional volatility under the random shocks' normality assumption violation. To obtain more accurate estimates of GARCH process parameters and conditional volatilities, we propose two semi-nonparametric GARCH models. The implementation of the proposed methods is based on an adaptation of the [Gallant, Nychka, 1987] semi-nonparametric method to the family of GARCH models. The approach provides a flexible estimation procedure by approximating the unknown density of random shocks both using polynomials (PGN-GARCH) and splines (SPL-GARCH). To study the properties of the obtained estimators and compare them with alternatives, we conducted an analysis of simulated data considering forms of distributions other than normal. As a result, statistical evidence was found in favor of the significant advantage of the proposed methods over the classical GARCH model and some other counterparts introduced earlier in the literature. Further, the proposed PGN-GARCH and SPL-GARCH models were applied to study Bitcoin conditional volatility dynamics. During the analysis, we found statistical evidence that Bitcoin distribution of shocks in returns differs from normal. Probably due to this reason the proposed SPL-GARCH model was able to demonstrate an advantage over alternative GARCH models according to the information criteria.

[1] This work was supported by the Russian Science Foundation [21-18-00427].

Juri Trifonov – Research Associate.
Bogdan Potanin – Cand. Sci. (Econ.), Senior Lecturer.

The article was received: 20.05.2022/The article is accepted for publication: 27.10.2022.

Key words: GARCH; semi-nonparametric methods; conditional volatility; simulated data analysis; financial econometrics*.*

JEL Classification: C14, C53, C58.

DOI: 10.17323/1813-8691-2022-26-4-623-646

For citation: J. Trifonov, B. Potanin. Semi-Nonparametric Generalized Autoregressive Conditional Heteroscedasticity Model with Application to Bitcoin Volatility Estimation. *HSE Economic Journal.* 2022; 26(4): 623–646.

Introduction

The generalized Autoregressive Conditional Heteroscedasticity (GARCH) process, proposed by [Bollerslev, 1986], is one of the most popular volatility modeling approaches. The classical GARCH model is based on a normality assumption. However, various empirical studies suggest that financial times series may deviate from normality, usually in terms of skewness and kurtosis. Despite the evidence that the classical GARCH estimator may preserve consistency under normality violation [Bollerslev, Wooldridge, 1992], an incorrect distribution specification usually results in efficiency losses [Feng, Shi, 2017]. Therefore, generally, researchers estimate the GARCH model under different distributional assumptions and select the most appropriate one via some validation criteria, i.e., mean-root-squared-error (RMSE), Akaike information criteria (AIC), Bayesian information criteria (BIC), and Hannan – Quinn information criteria (HQC).

Usually, researchers consider GARCH specifications with Student and Generalized Errors (GED) distributions allowing to account for heavy tails. Other popular options are Noncentral Student and Skewed Generalized Student distributions since they may exhibit asymmetric behavior (see: [Feng, Shi, 2017] for a review). However, some studies report deviations from these patterns, i.e., [Gemmill, Saflekos, 2000] have investigated stock volatility during British elections in 1987, 1992, and 1997. Researchers have found statistical evidence in favor of the bimodal distribution of shocks. Bimodality is probably associated with the fact that the elections may result in Labor or Conservatives victory and each of these events generates different shocks' distribution. Therefore, sometimes it is possible to guess non-trivial distribution shape from the analysis of specific market conditions. However, generally, there is no guarantee that the researchers will consider all crucial facts determining shocks' distribution. It motivates the application of highly flexible distributions allowing to account for different cases without a need to make a non-trivial prior analysis of the financial market conditions.

Because of the international nature of the cryptocurrency market and its close association with the highly innovative information technologies (IT) sector, this market is subject to various influences. It is usually complicated to determine all crucial patterns of these diverse influences. Therefore, it is hardly possible to provide rather strong and reliable arguments in favor of some particular form of shock distribution in the cryptocurrency market. Consequently, it is reasonable to apply semiparametric and nonparametric statistical methods, providing a flexible approximation to this distribution. These approximations do not impose strong assumptions on the market conditions, thus decreasing the risk of obtaining inaccurate volatility estimates because of relying on some erroneous assumptions.

In this article, we provide a GARCH model that is based on the flexible distribution proposed by [Gallant, Nychka, 1987]. As this approach stemmed from [Phillips, 1983], we call it the PGN (Phillips, Gallant, and Nychka) distribution. The advantages of the PGN-GARCH model are as follows. First, in many respects, PGN is more flexible than most popular alternatives, including GED and α-tempered stable distribution. Particularly PGN may approximate multimodal distributions. Second, in contrast to some non-parametric alternatives, including the GMM-GARCH approach [Skoglund, 2001], PGN-GARCH allows for approximating and visualizing the distribution of shocks. The visualization is important since it may provide the researchers with intuition concerning financial market conditions. Third, the PGN density and cumulative distribution functions, and moments have closed form expressions. It highly simplifies software implementation and combination of PGN distribution with copulas that may be useful for multivariate times series modeling. Fourth, researchers may select an arbitrary level of shock distribution approximation accuracy by varying the number of PGN parameters. Fifth, unlike kernel, polynomial, and other non-parametric density estimators, the PGN distributions generalize normal distribution, so it is possible to select between classical GARCH and PGN-GARCH models using the likelihood ratio test, Wald test, or Lagrange multiplier test.

The density function of the PGN distribution is a product of a squared polynomial and standard normal density function. High-order polynomials provide greater approximation accuracy but usually at the cost of numeric instability risks. Therefore, we have proposed PGN distribution generalization by substituting the polynomial with spline, i.e., spline distribution (SPL). The model that is based on this distribution we call SPL-GARCH.

To demonstrate PGN-GARCH and SPL-GARCH comparative advantages, we have conducted simulated data analysis under different shocks' distribution assumptions. We compare the classical GARCH model, GARCH with some flexible distributions, PGN-GARCH, SPL-GARCH, and kernel-based KDE-GARCH models. The results of the analysis suggest that PGN-GARCH and SPL-GARCH models usually notably outperform other approaches in terms of parameters and conditional variances estimation accuracy, especially if deviations from normality are rather strong. We also apply these models to the analysis of Bitcoin cryptocurrency volatility. According to AIC and BIC information criteria, SPL-GARCH provides notably better results than the classical GARCH model, being approximately equally accurate as models based on skewed generalized error distribution and skew Student distribution.

PGN Distribution

Let's say that random variable X has K-order PGN distribution with parameters vector $\tau = \left(\tau_0, \tau_1, \ldots, \tau_K\right)$ if the probability density function has the form:

$$f_X(x;\tau) = f_X(x) = \frac{\left(\tau_0 + \tau_1 x + \tau_2 x^2 + \ldots + \tau_K x^K\right)^2}{\sum_{i=0}^{K}\sum_{j=0}^{K}\tau_i\tau_j M(i+j)}\phi(x), \quad (1)$$

where $\phi(x)$ and $M(i+j)$ are a density function and the $(i+j)$-th moment of a standard normal distribution, correspondingly. Denominator ensures that integral of $f_X(x)$ over R converges to one. Nominator contains K-th order polynomial responsible for flexibility. Note that multiplying τ by nonzero constant yields the same density. Therefore, we impose a normalization condition $\tau_0 = 1$. Note that if $K=0$, then PGN distribution becomes a standard normal.

It is straightforward to derive the representations for the moments, and they always exist:

$$E\left(X^k\right) = \frac{\sum_{i=0}^{K}\sum_{j=0}^{K}\tau_i\tau_j M\left(i+j+k\right)}{\sum_{i=0}^{K}\sum_{j=0}^{K}\tau_i\tau_j M\left(i+j\right)}, \quad k \in N. \tag{2}$$

According to [Gallant, Nychka, 1987], it is possible to provide an arbitrarily close approximation to a broad family of distributions by increasing the polynomial degree K. Therefore, if one estimates the parameters of some model, it may be reasonable to substitute unknown density with $f_X(x)$ and increase K along with the sample size to ensure consistency [Gallant, Nychka, 1987]. We apply this idea to estimate GARCH model parameters relaxing the assumption that random shocks follow the particular distribution.

SPL Distribution

The density of the SPL random variable Y is similar to the density of the PGN random variable but is based on a spline instead of a polynomial:

$$f_Y(x;\tau,\kappa) = f_Y(x) = \frac{\left(\tau_0 B_{0,K}(x) + \tau_1 B_{1,K}(x) + \ldots + \tau_{n_\kappa - K - 2} B_{n_\kappa - K - 2,K}(x)\right)^2}{\sum_{t=1}^{n_\kappa - 1}\sum_{i=0}^{K}\sum_{j=0}^{K}\eta_{i,t}\eta_{j,t} M_{TR}\left(i+j;\kappa_t,\kappa_{t+1}\right)}\phi(x), \tag{3}$$

where $B_{i,K}$ are B-splines, $\kappa = \left(\kappa_1,\ldots,\kappa_{n_\kappa}\right)$ is a knots vector, and $M_{TR}\left(i+j;\kappa_t,\kappa_{t+1}\right)$ is a $(i+j)$-th moment of a standard normal distribution truncated from below and from above by κ_t and κ_{t+1} correspondingly, where $\kappa_{t+1} \geq \kappa_t$. In addition, $\eta_{i,t}$ is a coefficient in front of x^i for a spline (calculated as the linear combination of B-splines in the nominator) interval $\left[\kappa_t,\kappa_{t+1}\right)$. Therefore, $\eta_{i,t}$ are not parameters but constants that are uniquely determined by τ and κ.

Since the support of SPL distribution is $\left[\kappa_1,\kappa_{n_\kappa}\right)$, it is evident that standard normal distribution does not belong to this class of distributions. However, it is straightforward to show that all truncated (with some finite numbers from below and from above) standard normal distributions belong to SPL with $K=0$ and $n_\kappa = 2$.

The flexibility of this density form is ensured by the fact that any spline of degree K with knots vector κ may be represented as a linear combination of $\eta_\kappa - K - 1$ B-splines [Boor de, 1978, p. 99]. It is straightforward to calculate these B-splines using a well-known recurrence relation [Boor de, 1978, p. 89]:

(4) $$B_{i,0}(x)=\begin{cases}1, \text{ if } \kappa_i \le x < \kappa_{i+1}\\ 0, \text{ otherwise,}\end{cases}$$

(5) $$B_{i,K}(x)=g(x;i,K-1)B_{i,K-1}+\left(1-g(x;i+1,K-1)\right)B_{i+1,K-1}, \text{ for } K\ge 1,$$

(6) $$g(x;i,K)=\begin{cases}\dfrac{x-\kappa_i}{\kappa_{i+K}-\kappa_i}, \text{ if } \kappa_i \ne \kappa_{i+K}\\ 0, \text{ otherwise.}\end{cases}$$

Therefore, the function $S(x;\kappa,K)$ is a spline of degree K with a knots vector κ if and only if there are such $\tau_0,\ldots,\tau_{n_\kappa-K-2}\in R$ that:

(7) $$S(x;\kappa,K)=\tau_0 B_{0,K}(x)+\ldots+\tau_{n_\kappa-K-2}B_{n_\kappa-K-2,K}(x), \text{ for any } x\in R.$$

Moments of the SPL distribution have the following form:

(8) $$E\left(X^k\right)=\frac{\sum_{t=1}^{n_\kappa-1}\sum_{i=0}^{K}\sum_{j=0}^{K}\tau_{i,t}\tau_{j,t}M_{TR}\left(i+j+k;\kappa_t,\kappa_{t+1}\right)}{\sum_{t=1}^{n_\kappa-1}\sum_{i=0}^{K}\sum_{j=0}^{K}\tau_{i,t}\tau_{j,t}M_{TR}\left(i+j;\kappa_t,\kappa_{t+1}\right)},\ k\in N.$$

Similarly to the PGN distribution, in order to increase the approximation accuracy of SPL distribution, one needs to increase the number of knots while the degree of the spline may remain the same. Note that the number of distribution parameters decreases along with the degree of the spline. Therefore, if the degree is great, then to achieve appropriate approximation accuracy, there should be many knots as well. In practice, splines of 2-nd or 3-rd degrees are usually applied.

PGN-GARCH and SPL-GARCH Models

Without loss of generality, let's consider the most popular specification of the GARCH process, which is GARCH(1,1):

(9) $$y_t=\mu+\varepsilon_t,$$

(10) $$\varepsilon_t=\sigma_t\xi_t,$$

(11) $$\xi_t\sim\Theta,\ i.i.d.,\ E(\xi_t)=0,\ Var(\xi_t)=1,$$

(12) $$\sigma_t^2=\omega+\alpha\varepsilon_{t-1}^2+\beta\sigma_{t-1}^2,$$

where ω, α and β are the main parameters of interest since they determine dynamics of conditional variance σ_t^2. Standardized shocks ξ_t follow some distribution Θ that is usually unknown. Incorrect specification of Θ may result in inaccurate estimates of conditional variances. Therefore, we approximate Θ with a flexible family of standardized PGN distributions.

Consider i.i.d. random variables ξ_t^* following K order $\mathrm{PGN}(\tau)$ distribution. To ensure that standardized shocks have zero mean and identity variance let's normalize them:

$$\xi_t = \frac{\xi_t^* - E\left(\xi_t^*\right)}{\sqrt{Var\left(\xi_t^*\right)}}. \tag{13}$$

Then shocks will have the form $\varepsilon_t = \sigma_t \xi_t$ with a density function:

$$f_{\sigma_t \xi_t}\left(\varepsilon_t\right) = f_{\sigma_t \frac{\xi_t^* - E(\xi_t^*)}{\sqrt{Var(\xi_t^*)}}}\left(\varepsilon_t\right). \tag{14}$$

Then the cumulative distribution function of shocks takes the form:

$$\begin{aligned} F_{\sigma_t \xi_t}\left(\varepsilon_t\right) &= P\left(\sigma_t \frac{\xi_t^* - E\left(\xi_t^*\right)}{\sqrt{Var\left(\xi_t^*\right)}} \le \varepsilon_t\right) = P\left(\xi_t^* \le \frac{\sqrt{Var\left(\xi_t^*\right)}}{\sigma_t}\varepsilon_t + E\left(\xi_t^*\right)\right) = \\ &= F_{\xi_t^*}\left(\frac{\sqrt{Var\left(\xi_t^*\right)}}{\sigma_t}\varepsilon_t + E\left(\xi_t^*\right)\right), \end{aligned} \tag{15}$$

and the differentiation yields:

$$\begin{aligned} f_{\sigma_t \xi_t}\left(\varepsilon_t\right) &= \frac{f_{\sigma_t \frac{\xi_t^* - E(\xi_t^*)}{\sqrt{Var(\xi_t^*)}}}\left(\varepsilon_t\right) = d\left(F_{\xi^*}\left(\frac{\sqrt{Var\left(\xi_t^*\right)}}{\sigma_t}\varepsilon_t + E\left(\xi_t^*\right)\right)\right)}{d\varepsilon_t} = \\ &= \frac{\sqrt{Var\left(\xi_t^*\right)}}{\sigma_t} f_{\xi^*}\left(\frac{\sqrt{Var\left(\xi_t^*\right)}}{\sigma_t}\varepsilon_t + E\left(\xi_t^*\right)\right). \end{aligned} \tag{16}$$

In contrast to the classical GARCH model, $f_{\sigma_t \xi_t}\left(\varepsilon_t\right)$ depends on parameters τ responsible for the flexibility of the shocks' distribution. If all parameters τ (except the fixed one

$\tau_0 = 1$) are zero, the shocks follow the normal distribution, so PGN-GARCH coincides with the classical GARCH model.

The quasi log-likelihood function takes the form:

$$\ln L(\tau,\mu,\omega,\alpha,\beta;\varepsilon) = \sum_{t=1}^{n} \ln\left(f_{\sigma_t \xi_t}(\varepsilon_t) \right) = \frac{n}{2}\ln\left(Var\left(\xi_1^*\right)\right) + \sum_{t=1}^{n} \ln\left(f_{\xi_t^*}\left(\frac{\sqrt{Var\left(\xi_1^*\right)}}{\sigma_t}\varepsilon_t + E\left(\xi_1^*\right) \right)\right) - \ln(\sigma_t). \quad (17)$$

Maximization of quasi log-likelihood function provides semi-nonparametric estimators for μ, ω, α, β and τ. Substitution of $\hat{\tau}$ into $f_{\xi}(.)$ yields estimator for the density of standardized shocks. Therefore, it is straightforward to visualize the approximation of the unknown distribution of shocks and to calculate its parameters, i.e., modes, median, moments, and so on. The same estimator of conditional variance as in the classical GARCH model depends only on $\hat{\omega}$, $\hat{\alpha}$ and $\hat{\beta}$.

The SPL-GARCH model is similar. The only difference is the assumption that ξ_t^* has SPL distribution with parameters vector (τ,κ). Therefore, internal knots $\kappa_2,\dots,\kappa_{n_\kappa-1}$, coefficients $\tau_0,\dots,\tau_{n_\kappa-K-2}$, and the GARCH process parameters μ, ω, α, β are estimated via maximization of the quasi log-likelihood function. During the quasi log-likelihood function maximization routine, boundary knots $\kappa_1,\kappa_{n_\kappa}$ are set to the minimum and maximum of standardized shocks correspondingly.

KDE-GARCH Model

We have adopted the Bayesian KDE-GARCH model of [Zhang, King, 2013] to the non-Bayesian paradigm. The reasons to withdraw from the Bayesian paradigm were mainly due to the need of decreasing computational costs. In addition, we have tried to adopt the KDE-GARCH specification of [Wang et al., 2020] that was initially developed for realized GARCH model. The main idea is to maximize the quasi log-likelihood function with respect to parameters, where the true density of standardized shocks should be substituted with a kernel density estimator. During the numeric optimization process each iteration of a kernel density estimator is constructed from the standardized shocks, i.e., setting bandwidth via some mean integrated squared error (MISE) minimization criteria. However, we have found that this estimator seems to be poorly suited for the classical GARCH models since it usually provides notably less accurate estimates than the classical GARCH model. In addition, the need to select a bandwidth during each iteration requires many computational resources. Therefore, we omit this approach from the analysis.

Following [Zhang, King, 2013] suggestions, to avoid corner solutions associated with a nearly zero bandwidth, we have used a leave-one-out kernel density estimator:

(18) $$\hat{f}_{\xi_t^*}(\varepsilon_t/\sigma_t) = \frac{1}{(n-1)h} \sum_{i\in\{1,\dots,n\},\, i\neq t}^{n} K\left(\frac{\varepsilon_t/\sigma_t - \varepsilon_i/\sigma_i}{h}\right),$$

where h is bandwidth and K is a kernel function. Moreover, similarly to [Zhang, King, 2013], we use Gaussian kernel $K(.) = \phi(.)$, and fix $\hat{\omega} = \hat{\sigma}^2\left(1 - \hat{\alpha} - \hat{\beta}\right)$ where $\hat{\sigma}^2 = \sum_{t=1}^{n} \frac{(y_t - \overline{y})^2}{n-1}$ and $\overline{y} = \sum_{t=1}^{n} \frac{y_t}{n}$. During the preliminary analysis, we have also tried to apply the Epanechnikov and cosine kernels but have found that the results seem to be very similar, independently of the kernel type. Finally, in contrast to [Zhang, King, 2013], we have fixed $\hat{\mu} = \overline{y}$ since it seems to noticeably improve the accuracy of conditional volatilities estimates. Therefore, the estimates are obtained via maximization of a quasi-loglikelihood function:

(19) $$\ln L(\tau,\hat{\mu},\hat{\omega},\alpha,\beta,h;\varepsilon) = \sum_{t=1}^{n} \ln\left(\frac{1}{(n-1)h} \sum_{i\in\{1,\dots,n\},\, i\neq t}^{n} \phi\left(\frac{\varepsilon_t/\sigma_t - \varepsilon_i/\sigma_i}{h}\right)\right) - \ln(\sigma_t),$$

where $\hat{\mu} = \overline{y}$ and $\hat{\omega} = \hat{\sigma}^2\left(1 - \hat{\alpha} - \hat{\beta}\right)$. Our preliminary analysis has shown that if only one or none of these constraints are imposed, then KDE-GARCH conditional volatilities estimates are usually less accurate than estimates of the classical GARCH model. We leave a comprehensive analysis of this phenomenon for future research.

Simulated Data Analysis Design

We have considered Student distribution and its modifications, summarized in Table 1 and depicted in Fig. 1. We have considered modifications of Student distribution since heavy tails are supposed to be the most frequently observable (in literature) deviation from normality in time series. Therefore, by considering modifications of Student distribution, we combine this type of deviation with skewness and bimodality.

The motivation is to test the PGN-GARCH and SPL-GARCH models' ability to account for departures from normality in terms of tails' thickness, asymmetry, and bimodality. All distributions under consideration are particular cases of the mix of noncentral Student distributions (let's denote the random variable with this distribution as Z) with the following density function:

(20) $$f_Z(z) = p\psi(z - m/2; d, c) + (1-p)\psi(z + m/2; d, c),$$

where ψ denotes a density function of a noncentral Student distribution with d degrees of freedom and non-centrality parameter c. The parameter p determines the mix proportion while the parameter m is responsible for the distance between modes. We have set $d = 5$ to ensure that the tails are rather heavy. Similarly, $c = 10$ provides a notable asymmetry. The parameter p has been set to 0.7 in order to provide a moderate dissimilarity in probabilities concentrations around modes. Finally, we have set m to 5 and 20 for a Student and a noncentral Student distributions correspondingly since it seems to clearly distinguish the modes of distribution but do not make a gap with low probability masses between them (such a gap seems to be a shocks' pattern that has not been described in empirical studies).

Table 1.

The parameters of the distributions

Distribution	Parameters				Departure from normality
	d	*c*	*m*	*p*	
Student	5	0	0	1	Heavy tails
Noncentral Student	5	10	0	1	Heavy tails and asymmetry
Mix of Student	5	0	5	0.7	Heavy tails and bimodality
Mix of noncentral Student	5	10	20	0.7	Heavy tails, bimodality, and asymmetry

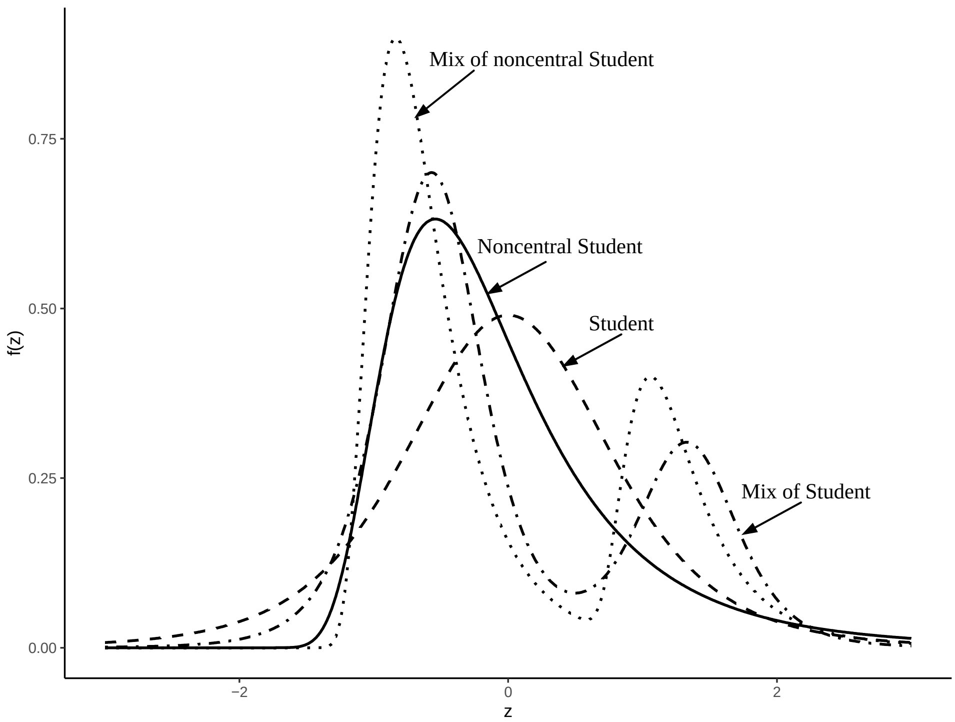

Fig. 1. Simulated distributions

We have simulated $s=100$ GARCH processes per distribution setting parameters to some common values, i.e., $\mu=1$, $\omega=0.1$, $\alpha=0.2$, and $\beta=0.7$. We have limited the study to 100 simulations because the first and the last 50 simulations have provided similar results and the sample standard deviation of the mean accuracy metrics (RMSE, AIC, and so forth) is rather small. Moreover, it takes a rather long time to perform these simulations, therefore, a smaller number of simulations facilitates reproducibility. Each simulation includes $n=1000$ observations (time periods) to compromise the need to satisfy the asymptotic properties of the QMLE estimator and the sample sizes that usually arise in practice. Samples of size above 1000 observations are probably subject to structural breaks. Therefore, researchers generally split such samples according to breakpoints and apply a separate model to each of the subsamples. At the same time application of the GARCH type models to the samples of noticeably smaller sizes may result in estimators' inefficiency.

For each simulation, we have used six types of models. First, the classical GARCH model, i.e., assuming a normal distribution of shocks. Second, popular flexible models assume the skewed generalized errors distribution SGED-GARCH and the skew Student distribution SSTD-GARCH. Third, the model with the true distribution TRUE-GARCH, i.e., with distributional parameters being fixed to true values according to Table 1. Fourth, the KDE-GARCH model with a Gaussian kernel. Fifth, the group of PGN-GARCH models with polynomial degrees K varying from 1 to 8. Six, the group of SPL-GARCH models with the number of knots varying from 5 to 10. To simplify the analysis, we consider cubic splines only by setting $K=3$. Also, we restrict knots vector elements to avoid multiplicities, i.e., $\kappa_i \neq \kappa_j$ for all $i, j \in \{\kappa_1, \ldots, \kappa_{n_\kappa}\}$. Both restrictions ensure high smoothness of the approximating function and simplify the optimization routine by the cost of flexibility penalty.

In each simulation, we choose the best of the PGN-GARCH and SPL-GARCH models following AIC and BIC minimization criteria. We call these models PGN-AIC-GARCH, PGN-BIC-GARCH, SPL-AIC-GARCH, and SPL-BIC-GARCH. The goal is to determine whether these information criteria are efficient tools for the specification of PGN-GARCH and SPL-GARCH models, i.e., for the selection of the polynomial degree and the number of knots n_κ.

We are using root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) accuracy metrics to compare estimates of conditional volatility among models:

$$\text{RMSE}(\hat{\sigma}) = \sum_{i=1}^{s}\left(\sqrt{\frac{1}{n}\sum_{t=1}^{n}(\sigma_t - \hat{\sigma}_t)^2}\right), \tag{21}$$

$$\text{MAE}(\hat{\sigma}) = \frac{1}{s}\sum_{i=1}^{s}\left(\frac{1}{n}\sum_{t=1}^{n}\left|\sigma_t - \hat{\sigma}_t\right|\right), \tag{22}$$

$$\text{MAPE}(\hat{\sigma}) = \frac{1}{s}\sum_{i=1}^{s}\left(\frac{100}{n}\sum_{t=1}^{n}\left|\frac{\sigma_t - \hat{\sigma}_t}{\sigma_t}\right|\right). \tag{23}$$

We also estimate the share of victories of each model over the classical GARCH model, i.e., the number of simulations in which the model has provided a lower RMSE value than the classical GARCH model. The goal is to investigate whether these models not only on average but usually outperform the classical GARCH model.

To maximize the quasi log-likelihood function for the models of both types, we have used the popular R-package "rugarch" implemented by [Ghalanos, 2020]. We also have slightly improved GARCH estimates obtained via "rugarch" by performing additional 1000 iterations of the elitist hybrid genetic algorithm (with analytical gradient and BFGS local optimizer) starting from the point returned from the "rugarch". For other (manually implemented in R) models, we have used 200 iterations of an elitist hybrid genetic algorithm with a BFGS local optimizer for the PGN-GARCH model (since we have implemented an analytical gradient for this model) and Nelder-Mead – for other models. Since KDE-GARCH, PGN-GARCH, and SPL-GARCH quasi log-likelihood functions are not necessarily concave, there is always a chance that the local maximum will be obtained instead of the global one. Therefore, the results presented below demonstrate some sort of a "lower bound efficiency" for the KDE-GARCH, PGN-GARCH, and SPL-GARCH models.

Simulated Data Analysis Results

The results of the simulated data analysis are presented in Table 2. It is slightly surprising that RMSE, MAE, and MAPE of classical GARCH model estimates are higher for unimodal than for bimodal distributions. We have checked that this result is not driven by the outliers since after removing of 10 most inaccurate (largest RMSE) simulations for each of these distributions, estimates for bimodal distributions remain more precise. Also, we have found that this result holds for the classical GARCH model even if we moderately vary the parameters of the distributions from Table 1.

Table 2.

Accuracy metrics for conditional volatilities estimates

Metric	GARCH model				
	Classic	SGED	SSTD	TRUE	KDE
	Student distribution				
$100 \cdot \text{RMSE}(\hat{\sigma})$	6.738	5.996	5.958	5.396	8.485
$\text{MAPE}(\hat{\sigma})$	4.870	4.519	4.446	3.874	6.441
$100 \cdot \text{MAE}(\hat{\sigma})$	4.749	4.339	4.314	3.782	6.070
Victories %		0.68	0.65	0.76	0.35
AIC (average)	2.567	2.494	2.483	2.482	2.505
BIC (average)	2.587	2.524	2.513	2.502	2.521
HQC (average)	2.574	2.505	2.495	2.490	2.506

Continues

Metric	GARCH model				
	Classic	SGED	SSTD	TRUE	KDE
	Noncentral Student distribution				
$100 \cdot \text{RMSE}(\hat{\sigma})$	12.316	6.879	5.947	3.403	8.817
$\text{MAPE}(\hat{\sigma})$	7.892	6.449	5.379	2.655	7.879
$100 \cdot \text{MAE}(\hat{\sigma})$	7.604	5.601	4.717	2.402	6.780
Victories %		0.77	0.86	0.98	0.74
AIC (average)	2.436	1.947	1.929	1.915	1.956
BIC (average)	2.456	1.977	1.959	1.935	1.971
HQC (average)	2.444	1.958	1.940	1.922	1.956
	Mix of Student distributions (bimodal)				
$100 \cdot \text{RMSE}(\hat{\sigma})$	3.496	5.451	4.034	2.426	3.557
$\text{MAPE}(\hat{\sigma})$	2.692	4.802	3.315	1.884	2.838
$100 \cdot \text{MAE}(\hat{\sigma})$	2.661	4.675	3.226	1.867	2.760
Victories %		0.25	0.37	0.81	0.42
AIC (average)	2.713	2.534	2.558	2.259	2.290
BIC (average)	2.733	2.563	2.587	2.279	2.305
HQC (average)	2.720	2.545	2.569	2.267	2.290
	Mix of noncentral Student distribution (bimodal)				
$100 \cdot \text{RMSE}(\hat{\sigma})$	4.187	4.724	7.526	1.552	3.446
$\text{MAPE}(\hat{\sigma})$	2.932	4.331	7.231	1.209	2.839
$100 \cdot \text{MAE}(\hat{\sigma})$	2.962	4.199	6.974	1.188	2.723
Victories %		0.41	0.15	0.91	0.61
AIC (average)	2.703	2.189	2.227	1.822	1.873
BIC (average)	2.722	2.218	2.256	1.842	1.888
HQC (average)	2.710	2.200	2.238	1.830	1.873

Continues

Metric	GARCH model					
	PGN-AIC	PGN-BIC	PGN-HQC	SPL-AIC	SPL-BIC	SPL-HQC
Student distribution						
$100 \cdot \mathrm{RMSE}(\hat{\sigma})$	6.408	6.099	6.147	5.740	5.837	5.786
$\mathrm{MAPE}(\hat{\sigma})$	4.706	4.555	4.520	4.394	4.445	4.418
$100 \cdot \mathrm{MAE}(\hat{\sigma})$	4.565	4.380	4.378	4.195	4.242	4.219
Victories %	0.55	0.62	0.58	0.69	0.68	0.69
AIC (average)	2.500	2.504	2.500	2.491	2.496	2.493
BIC (average)	2.548	2.544	2.545	2.550	2.544	2.546
HQC (average)	2.518	2.519	2.518	2.514	2.514	2.513
Noncentral Student distribution						
$100 \cdot \mathrm{RMSE}(\hat{\sigma})$	8.397	8.432	8.362	5.976	6.120	5.997
$\mathrm{MAPE}(\hat{\sigma})$	6.484	6.453	6.455	5.507	5.585	5.492
$100 \cdot \mathrm{MAE}(\hat{\sigma})$	5.880	5.870	5.854	4.834	4.915	4.826
Victories %	0.80	0.81	0.81	0.88	0.86	0.87
AIC (average)	1.987	1.988	1.987	1.926	1.927	1.927
BIC (average)	2.045	2.044	2.045	1.980	1.977	1.977
HQC (average)	2.009	2.009	2.009	1.947	1.946	1.946
Mix of Student distributions (bimodal)						
$100 \cdot \mathrm{RMSE}(\hat{\sigma})$	2.918	3.004	2.930	2.837	2.762	2.820
$\mathrm{MAPE}(\hat{\sigma})$	2.295	2.342	2.301	2.232	2.178	2.218
$100 \cdot \mathrm{MAE}(\hat{\sigma})$	2.265	2.322	2.272	2.206	2.146	2.189
Victories %	0.67	0.65	0.66	0.75	0.75	0.74
AIC (average)	2.282	2.287	2.283	2.276	2.278	2.276
BIC (average)	2.334	2.330	2.332	2.345	2.341	2.342
HQC (average)	2.302	2.303	2.301	2.302	2.302	2.301
Mix of noncentral Student distribution (bimodal)						
$100 \cdot \mathrm{RMSE}(\hat{\sigma})$	2.882	2.908	2.886	2.385	2.353	2.375
$\mathrm{MAPE}(\hat{\sigma})$	2.255	2.253	2.256	2.116	2.094	2.107
$100 \cdot \mathrm{MAE}(\hat{\sigma})$	2.212	2.220	2.213	2.014	1.992	2.005
Victories %	0.76	0.76	0.76	0.77	0.78	0.78
AIC (average)	1.928	1.929	1.928	1.867	1.869	1.867
BIC (average)	1.987	1.986	1.986	1.943	1.939	1.940
HQC (average)	1.950	1.951	1.950	1.896	1.896	1.895

According to the RMSE, MAE, and MAPE accuracy metrics TRUE-GARCH model seems to noticeably outperform other alternatives in terms of conditional volatility estimation precision. This is to be expected since the TRUE-GARCH model is based on the true distribution of random shocks. Moreover, no semi-parametric or non-parametric GARCH model is likely to provide more efficient estimates than the TRUE-GARCH model. For this reason no misspecification in the TRUE-GARCH model provide asymptotically efficient estimators, since they have been obtained via the maximum likelihood method. Therefore, TRUE-GARCH model provides some kind of upper bound efficiency for other methods. So, we will say that some models are the first-best (according to their accuracy metrics), meaning that they are the first-best except for the TRUE-GARCH model.

The SSTD-GARCH model is the most accurate for noncentral Student distribution and the third-best for Student distribution being just slightly less accurate than SPL-AIC-GARCH and SGED-GARCH models. The SGED-GARCH model seems to nicely approximate heavy tails since it is the second-best for Student distribution. However, probably, it deals poorly with skewness because the SGED-GARCH model is almost the worst for noncentral Student distribution. Both SSTD-GARCH and SGED-GARCH models provide inaccurate results for bimodal distributions being noticeably less accurate than the classical GARCH model

The KDE-GARCH model outperforms classical GARCH for unimodal and bimodal versions of noncentral Student distribution. Unfortunately, it provides much less accurate estimates than GARCH models based on flexible distributions. Wherein the KDE-GARCH model usually provides the lowest information criteria values. Therefore, these information criteria seem to be non-informative when comparing KDE-GARCH with the alternatives. However, for other models, there is a clear positive correlation between AIC criteria and accuracy metrics. The relationship of accuracy metrics with BIC and HQC is sometimes ambiguous. For example, for Student and noncentral Student distributions SPL-GARCH model slightly outperforms SSTD-GARCH and GED-GARCH models. But the latter provides much lower BIC and HQC values. Therefore, we conclude that AIC seems to be a more informative measure when comparing the quality of the models under consideration.

The PGN-GARCH model demonstrates the advantage over the classical GARCH model for all distributions, and over SGED-GARCH and SSTD-GARCH models for bimodal distributions. However, PGN-GARH seems to be less accurate than SPL-GARCH.

Also, we have compared the accuracy of parameter estimates across the methods. The results are presented in Table 3. They are predominantly consistent with the aforementioned findings. Note that KDE-GARCH seems to provide the least accurate estimates of the parameter μ that has been fixed to be equal to the sample mean. Probably, it is possible to improve the accuracy of the KDE-GARCH model by substituting the sample mean estimate of μ with an estimate of other GARCH model. The same is true for a ω parameter.

To check the robustness of the results to changes in sample size, we have replicated the analysis for n = 300 observations. In general, the results (see Tables 4–5) are similar to those mentioned above. Particularly, PGN-GARCH and SPL-GARCH demonstrate an advantage over other alternatives, especially for the mixes of distributions. In addition, AIC, BIC, and HQC seem to be equally accurate specification selection criteria for these models.

Table 3.

Accuracy metrics for parameters estimates, 1000 observations

Metric	GARCH model				
	Classic	SGED	SSTD	TRUE	KDE
Student distribution					
MAPE($\hat{\mu}$)	2.356	2.224	2.122	1.719	2.594
MAPE($\hat{\omega}$)	35.663	29.793	30.553	30.210	54.263
MAPE($\hat{\alpha}$)	20.672	17.605	16.995	16.403	27.243
MAPE($\hat{\beta}$)	9.418	8.127	7.956	7.883	13.171
Noncentral Student distribution					
MAPE($\hat{\mu}$)	2.089	1.803	1.810	1.716	2.173
MAPE($\hat{\omega}$)	47.665	16.409	16.246	12.228	30.922
MAPE($\hat{\alpha}$)	43.764	16.188	14.620	10.928	22.634
MAPE($\hat{\beta}$)	14.573	4.702	4.273	3.683	8.877
Mix of Student distributions (bimodal)					
MAPE($\hat{\mu}$)	2.142	2.300	3.341	1.183	2.427
MAPE($\hat{\omega}$)	28.317	28.587	25.975	15.521	27.926
MAPE($\hat{\alpha}$)	15.731	17.469	14.832	10.142	14.643
MAPE($\hat{\beta}$)	7.186	6.111	6.864	3.824	6.456
Mix of noncentral Student distribution (bimodal)					
MAPE($\hat{\mu}$)	2.104	1.947	3.565	0.748	2.494
MAPE($\hat{\omega}$)	31.548	19.927	16.276	11.113	25.197
MAPE($\hat{\alpha}$)	19.285	13.326	12.643	6.687	14.721
MAPE($\hat{\beta}$)	7.602	3.852	4.351	2.791	6.673

Continues

Metric	GARCH model					
	PGN-AIC	PGN-BIC	PGN-HQC	SPL-AIC	SPL-BIC	SPL-HQC
Student distribution						
MAPE($\hat{\mu}$)	2.113	2.173	2.161	2.132	2.189	2.155
MAPE($\hat{\omega}$)	35.883	34.449	35.103	30.970	34.485	32.143
MAPE($\hat{\alpha}$)	18.835	17.120	17.690	15.747	15.435	15.950
MAPE($\hat{\beta}$)	8.931	8.549	8.658	7.987	8.841	8.321
Noncentral Student distribution						
MAPE($\hat{\mu}$)	2.331	2.300	2.332	1.836	1.829	1.837
MAPE($\hat{\omega}$)	28.935	29.328	28.720	17.008	17.283	17.233
MAPE($\hat{\alpha}$)	24.115	24.595	24.250	14.611	14.893	14.562
MAPE($\hat{\beta}$)	8.100	8.182	8.019	4.072	4.245	4.187
Mix of Student distributions (bimodal)						
MAPE($\hat{\mu}$)	2.152	2.177	2.161	2.240	2.178	2.224
MAPE($\hat{\omega}$)	18.243	19.408	18.192	17.199	17.751	17.673
MAPE($\hat{\alpha}$)	11.759	12.899	11.961	11.395	10.837	11.093
MAPE($\hat{\beta}$)	4.248	4.715	4.291	4.426	4.448	4.407
Mix of noncentral Student distribution (bimodal)						
MAPE($\hat{\mu}$)	1.998	1.911	1.980	1.702	1.804	1.741
MAPE($\hat{\omega}$)	18.522	18.443	18.627	12.772	11.984	12.502
MAPE($\hat{\alpha}$)	11.203	11.490	11.268	7.707	7.431	7.636
MAPE($\hat{\beta}$)	4.795	4.816	4.833	3.129	2.969	3.047

Table 4.

Accuracy metrics for conditional volatilities estimates, 300 observations

Metric	GARCH model				
	Classic	SGED	SSTD	TRUE	KDE
Student distribution					
$100 \cdot \mathrm{RMSE}(\hat{\sigma})$	12.526	11.914	13.067	7.988	21.871
$\mathrm{MAPE}(\hat{\sigma})$	9.060	8.532	9.897	6.049	22.334
$100 \cdot \mathrm{MAE}(\hat{\sigma})$	8.835	8.412	9.632	5.794	17.627
Victories, %		0.58	0.50	0.77	0.23
AIC (average)	2.560	2.518	2.509	2.507	2.578
BIC (average)	2.639	2.592	2.583	2.557	2.615
HQC (average)	2.610	2.548	2.539	2.527	2.592
Noncentral Student distribution					
$100 \cdot \mathrm{RMSE}(\hat{\sigma})$	18.874	11.171	13.823	4.331	23.880
$\mathrm{MAPE}(\hat{\sigma})$	12.591	9.410	11.570	3.493	23.217
$100 \cdot \mathrm{MAE}(\hat{\sigma})$	12.180	8.476	10.754	3.120	18.625
Victories, %		0.75	0.70	0.95	0.16
AIC (average)	2.479	1.995	1.981	1.969	2.108
BIC (average)	2.529	2.069	2.055	2.019	2.145
HQC (average)	2.450	2.024	2.011	1.989	2.122
Mix of Student distributions (bimodal)					
$100 \cdot \mathrm{RMSE}(\hat{\sigma})$	6.055	7.343	7.163	3.707	14.964
$\mathrm{MAPE}(\hat{\sigma})$	4.769	6.083	5.900	3.026	14.157
$100 \cdot \mathrm{MAE}(\hat{\sigma})$	4.704	5.994	5.755	2.945	12.211
Victories, %		0.40	0.38	0.77	0.26
AIC (average)	2.737	2.567	2.591	2.302	2.407
BIC (average)	2.787	2.642	2.665	2.351	2.444
HQC (average)	2.757	2.598	2.620	2.321	2.421
Mix of noncentral Student distribution (bimodal)					
$100 \cdot \mathrm{RMSE}(\hat{\sigma})$	6.507	7.686	9.293	2.226	14.640
$\mathrm{MAPE}(\hat{\sigma})$	4.947	6.979	8.777	1.833	13.639
$100 \cdot \mathrm{MAE}(\hat{\sigma})$	4.891	6.724	8.340	1.757	11.884
Victories, %	0.39	0.46	0.30	0.91	0.18
AIC (average)	2.701	2.201	2.231	1.841	2.043
BIC (average)	2.751	2.275	2.305	1.890	2.080
HQC (average)	2.721	2.230	2.260	1.861	2.058

Continues

Metric	GARCH model					
	PGN-AIC	PGN-BIC	PGN-HQC	SPL-AIC	SPL-BIC	SPL-HQC
	Student distribution					
$100 \cdot \text{RMSE}(\hat{\sigma})$	13.494	13.076	13.445	11.902	11.572	11.875
$\text{MAPE}(\hat{\sigma})$	9.967	9.493	9.992	8.847	8.644	8.768
$100 \cdot \text{MAE}(\hat{\sigma})$	9.622	9.247	9.643	8.573	8.356	8.515
Victories, %	0.51	0.50	0.49	0.52	0.54	0.52
AIC (average)	2.518	2.527	2.521	2.531	2.538	2.533
BIC (average)	2.613	2.600	2.603	2.660	2.649	2.652
HQC (average)	2.556	2.557	2.554	2.583	2.582	2.580
	Noncentral Student distribution					
$100 \cdot \text{RMSE}(\hat{\sigma})$	13.461	13.957	13.344	9.920	11.119	10.079
$\text{MAPE}(\hat{\sigma})$	10.872	11.224	10.898	8.683	9.614	8.914
$100 \cdot \text{MAE}(\hat{\sigma})$	9.842	10.176	9.816	7.714	8.628	7.902
Victories, %	0.79	0.77	0.79	0.87	0.74	0.81
AIC (average)	2.038	2.044	2.040	2.003	2.013	2.005
BIC (average)	2.173	2.168	2.169	2.151	2.141	2.145
HQC (average)	2.092	2.094	2.092	2.062	2.064	2.061
	Mix of Student distributions (bimodal)					
$100 \cdot \text{RMSE}(\hat{\sigma})$	5.280	5.339	5.218	5.046	5.183	5.219
$\text{MAPE}(\hat{\sigma})$	4.237	4.317	4.213	4.102	4.209	4.233
$100 \cdot \text{MAE}(\hat{\sigma})$	4.151	4.222	4.125	4.010	4.115	4.140
Victories, %	0.63	0.67	0.68	0.69	0.66	0.67
AIC (average)	2.318	2.326	2.320	2.346	2.348	2.346
BIC (average)	2.423	2.411	2.414	2.517	2.512	2.513
HQC (average)	2.360	2.360	2.358	2.414	2.414	2.413
	Mix of noncentral Student distribution (bimodal)					
$100 \cdot \text{RMSE}(\hat{\sigma})$	4.994	4.997	4.999	4.398	4.518	4.506
$\text{MAPE}(\hat{\sigma})$	4.120	4.109	4.118	3.964	4.052	4.070
$100 \cdot \text{MAE}(\hat{\sigma})$	3.986	3.988	3.989	3.720	3.811	3.821
Victories, %	0.73	0.76	0.75	0.69	0.70	0.67
AIC (average)	1.949	1.956	1.945	1.938	1.943	1.939
BIC (average)	2.087	2.083	2.084	2.121	2.113	2.115
HQC (average)	2.004	2.007	2.004	2.011	2.011	2.009

Table 5.

Accuracy metrics for parameters estimates, 300 observations

Metric	GARCH model				
	Classic	SGED	SSTD	TRUE	KDE
Student distribution					
MAPE($\hat{\mu}$)	3.681	3.505	3.309	2.263	4.541
MAPE($\hat{\omega}$)	78.663	58.524	61.982	89.313	109.287
MAPE($\hat{\alpha}$)	40.737	41.372	41.826	22.542	48.349
MAPE($\hat{\beta}$)	20.360	17.884	17.897	20.132	32.904
Noncentral Student distribution					
MAPE($\hat{\mu}$)	4.081	3.619	3.569	1.916	4.862
MAPE($\hat{\omega}$)	123.542	32.098	43.378	33.004	123.705
MAPE($\hat{\alpha}$)	66.829	28.020	32.447	12.174	54.221
MAPE($\hat{\beta}$)	30.956	9.694	10.197	8.346	24.248
Mix of Student distributions					
MAPE($\hat{\mu}$)	3.790	4.972	4.006	1.262	4.362
MAPE($\hat{\omega}$)	66.408	51.871	56.517	44.934	75.699
MAPE($\hat{\alpha}$)	27.986	27.603	28.897	13.865	34.257
MAPE($\hat{\beta}$)	14.945	12.513	14.236	9.995	18.121
Mix of Bimodal noncentral Student distribution					
MAPE($\hat{\mu}$)	3.396	4.229	4.916	0.921	3.845
MAPE($\hat{\omega}$)	66.276	42.683	36.116	29.713	88.018
MAPE($\hat{\alpha}$)	28.987	24.133	21.897	9.008	34.290
MAPE($\hat{\beta}$)	15.229	9.437	9.266	6.960	18.261

Continues

Metric	GARCH model					
	PGN-AIC	PGN-BIC	PGN-HQC	SPL-AIC	SPL-BIC	SPL-HQC
	Student distribution					
MAPE($\hat{\mu}$)	3.775	3.793	3.955	3.520	3.358	3.439
MAPE($\hat{\omega}$)	83.549	74.237	81.247	68.163	65.998	69.153
MAPE($\hat{\alpha}$)	42.161	43.348	42.772	37.469	37.276	38.055
MAPE($\hat{\beta}$)	22.162	20.488	21.777	18.486	18.530	19.217
	Noncentral Student distribution					
MAPE($\hat{\mu}$)	4.057	4.212	4.131	3.600	3.768	3.652
MAPE($\hat{\omega}$)	63.454	68.851	65.120	32.942	34.677	32.703
MAPE($\hat{\alpha}$)	35.845	37.847	35.592	24.340	26.207	23.838
MAPE($\hat{\beta}$)	13.294	14.194	13.130	9.055	9.878	8.960
	Mix of Student distributions					
MAPE($\hat{\mu}$)	3.639	3.757	3.605	3.663	3.707	3.659
MAPE($\hat{\omega}$)	45.080	44.612	43.931	42.989	47.520	47.700
MAPE($\hat{\alpha}$)	22.844	25.704	23.464	21.840	23.666	23.396
MAPE($\hat{\beta}$)	10.346	10.976	10.222	9.885	11.192	10.972
	Mix of Bimodal noncentral Student distribution					
MAPE($\hat{\mu}$)	3.369	3.319	3.346	3.213	3.348	3.194
MAPE($\hat{\omega}$)	40.760	40.399	40.668	29.359	31.365	29.737
MAPE($\hat{\alpha}$)	20.952	20.798	20.758	15.247	16.312	15.695
MAPE($\hat{\beta}$)	10.393	10.303	10.366	7.734	7.952	7.611

The key difference between the results for 300 and 1000 observations is that, in most cases, the advantage of the PGN-GARCH and SPL-GARCH model over classical GARCH has decreased, which is usual for semiparametric methods since they usually suffer from great estimators' variances for small sample sizes. For example, when random shocks follow Student's distribution, the number of victories of SPL-AIC-GARCH over classical GARCH has declined from

0.66 (for 1000 observations) to 0.52 (for 300 observations). However, for noncentral Student distribution corresponding decline was from 0.88 to 0.87, and seems to be insignificant. So, probably, for small sample sizes, the application of SPL-GARCH and PGN-GARCH models may result in substantial accuracy gains only when deviations from normality are rather strong.

Therefore, the results of simulated data analysis suggest that SPL-GARCH and PGN-GARCH models usually outperform the classical GARCH model. They also provide noticeably more efficient estimators than SGED-GARCH and SSTD-GARCH models in the case of bimodal distributions. Finally, we suggest using the SPL-GARCH model over PGN-GARCH since the former usually provides equally or noticeably more accurate results.

Bitcoin Volatility Analysis

During the past few years, there has been rapid growth in the cryptocurrency market. The increasing popularity of cryptocurrency is because it is not only an attractive investment asset but also a currency for international payments. Nevertheless, investments in cryptocurrency are risky. Therefore, effective risk management requires a detailed analysis of cryptocurrency returns volatility.

Bitcoin is the most popular cryptocurrency. Previous studies have indicated that shocks of returns of Bitcoin seem to deviate from normality [Katsiampa, 2017]. Specifically, heavy tail evidence has been found by [Troster et al., 2019]. These findings motivate the application of flexible volatility estimation technics, i.e., PGN-GARCH and SPL-GARCH models.

The sample consists of 1095 daily observations of log returns and covers the period from 01.01.2017 to 31.12.2019 (we omit the analysis of the 2020–2021 years period to avoid possible structural breaks probably associated with the influence of coronavirus). The data was retrieved from Investing.com. We have selected optimal specifications for PGN-GARCH and SPL-GARCH models following the AIC criterion (since the abovementioned simulations suggest that it is probably slightly more accurate than BIC), suggesting $K=8$ and $n_{\kappa}=8$, correspondingly.

Following [Katsiampa, 2017], during the preliminary analysis, we additionally accounted for the autoregressive (AR) part. However, we have found that for all specifications of random shocks' distribution, according to all information criteria, the inclusion of any number of lags (up to 8) decreases the quality of the model. Moreover, autocorrelation and partial autocorrelation graphs of the returns have not indicated the presence of autocorrelation. Therefore, we decided not to include the autoregressive part in the model.

The estimation results are presented in Table 10. We do not report standard errors of the estimates since their accurate estimation for KDE-GARCH, PGN-GARCH, and SPL-GARCH models require analytical hessian implementation (numeric hessian seems to be rather inaccurate), which is a rather complicated technical task (we left it for a future research).

The difference in parameter estimates among models seems to be rather small. However, the mean absolute percentage deviations (MAPD) of conditional volatility estimates of the classical GARCH model from the corresponding estimates of other models are rather high, varying approximately from 6.2 to 9.7 percent. It indicates that despite parameter estimates similarity, there could be substantial differences in volatility estimates depending on the model being applied.

Table 6.

The results of Bitcoin's volatility analysis

	GARCH model					
	Classic	SGED	SSTD	KDE	PGN	SPL
$\hat{\mu}$	0.00171	0.00116	0.00118	0.00181	0.00063	0.00048
$\hat{\omega}$	0.00013	0.00005	0.00003	0.00003	0.00006	0.00005
$\hat{\alpha}$	0.13401	0.13979	0.12290	0.13225	0.13341	0.13750
$\hat{\beta}$	0.79998	0.84784	0.87610	0.85075	0.85038	0.85496
AIC	–3.577	–3.759	–3.754	–3.801	–3.722	–3.757
BIC	–3.558	–3.732	–3.726	–3.782	–3.667	–3.702
HQC	–3.570	–3.749	–3.743	–3.785	–3.708	–3.754
MAPD	0	7.19648	9.5344	9.66300	6.20688	8.19992

The values of information criteria suggest that PGN-GARCH and SPL-GARCH models outperform the classical GARCH model, providing evidence in favor of normality assumption violation. The graphical illustration (see Fig. 2) for the SPL-GRACH model reinforces this evidence suggesting that there are deviations from normality both in terms of asymmetry and tail thickness.

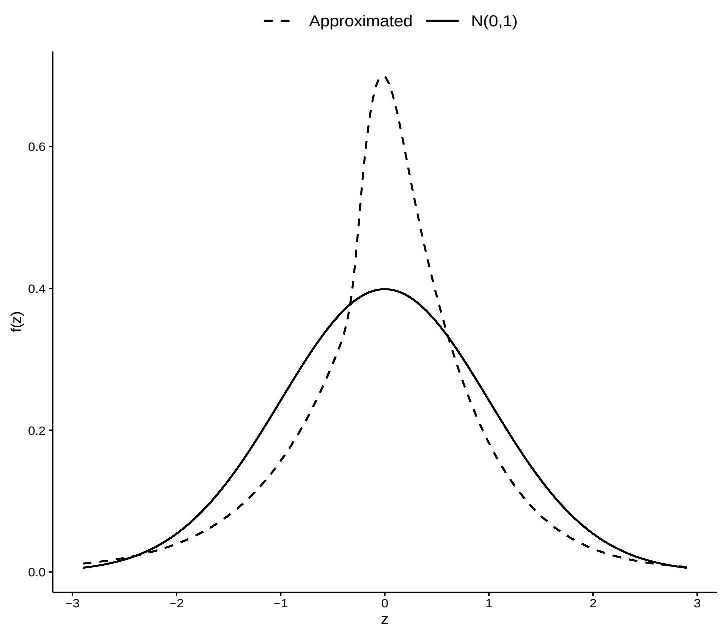

Fig. 2. Random shocks density approximation

Even though the KDE-GARCH model provides the lowest AIC and BIC values we can't conclude that there is a statistical evidence that this model outperforms the alternatives because the abovementioned simulated data analysis suggests that information criteria are not informative when comparing the KDE-GARCH model with GARCH models based on flexible distributions.

Both the SGED-GARCH and SSTD-GARCH models provide lower AIC and BIC values than the SPL-GARCH model. But the difference in terms of AIC criterion is rather small, while (according to the abovementioned simulated data analysis) the difference in BIC may be misleading. Therefore, the SGED-GARCH, SSTD-GARCH, and SPL-GARCH models seem to be approximately of the same quality.

Conclusions

We have proposed PGN-GARCH and SPL-GARCH models based on flexible distribution introduced by [Gallant, Nychka, 1987]. These models allow relaxing distributional assumptions in GARCH models. According to the simulated data analysis, these models may outperform other well-known alternatives, including the classical GARCH model, GARCH models based on some flexible distributions (SSTD-GARCH and SGED-GRACH), and kernel-based GARCH model (KDE-GARCH). Bitcoin log-returns analysis is in line with this evidence since PGN-GARCH and SPL-GARCH notably outperform the classical GARCH model in terms of information criteria being approximately equally accurate as SSTD-GARCH and SGED-GARCH models. Also, both simulated and real data analyses suggest that the SPL-GARCH model (which is based on splines) probably outperforms the PGN-GARCH model (which is based on polynomials).

Finally, we would like to emphasize possible extensions of our research. First, it is possible to adopt PGN and SPL distributions to other types of univariate and multivariate GARCH models, i.e., EGARCH, DCC-GARCH, and so forth. Also, it is possible to consider other time series models, including the autoregressive integrated moving average (ARIMA) model. Moreover, one may adopt SPL distribution to econometric models beyond the scope of the times series analysis. For example, one may implement an SPL-based truncated or binary choice regression model.

Second, it seems that the numeric hessian for PGN-GARCH and SPL-GARCH models is inaccurate, which complicates hypothesis testing. Therefore, it is necessary to get the analytical expression for a hessian for these models or to implement a bootstrap procedure.

Finally, it is interesting to implement structural breaks' identification procedures for PGN-GARCH and SPL-GARCH models.

* *
*

References

Bollerslev T. (1986) Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics,* 31, 3, pp. 307–327. Doi: 10.1016/0304-4076(86)90063-1

Bollerslev T., Wooldridge J.M. (1992) Quasi-maximum Likelihood Estimation and Inference in Dynamic Models with Time-varying Covariances. *Econometric Reviews,* 11, 2, pp. 143–172. Doi: 10.1080/07474939208800229

Carl de Boor C. (1978) A Practical Guide to Spline. *Mathematics of Computation,* 27, 149. Doi: 10.2307/2006241

Feng L., Shi Y. (2017) A Simulation Study on the Distributions of Disturbances in the GARCH Model. *Cogent Economics & Finance,* 5, 1. Doi: 10.1080/23322039.2017.1355503

Gallant A., Nychka D. (1987) Semi-nonparametric Maximum Likelihood Estimation. *Econometrica,* 55, 2, pp. 363–390. Doi: 10.2307/1913241

Gemmill G., Saflekos A. (2000) How Useful Are Implied Distributions? Evidence from Stock-index Options. *The Journal of Derivatives,* 7, 3, pp. 83–98. Doi: 10.2307/1913241

Ghalanos A. (2020) rugarch: Univariate GARCH Models. R package version 1.4–4.

Katsiampa P. (2017) Volatility Estimation for Bitcoin: A Comparison of GARCH Models. *Economics Letters,* 158(C), pp. 3–6. Doi: 10.1016/j.econlet.2017.06.023

Phillips P. (1983) ERA's: A New Approach to Small Sample Theory. *Econometrica,* 51, 5, pp. 1505–1525. Doi: 10.2307/1912287

Scrucca L. (2017) On Some Extensions to GA Package: Hybrid Optimisation, Parallelisation and Islands Evolution. *The R Journal,* 9, 1, pp. 187–206. Doi: 10.32614/RJ-2017-008

Sheather S.J., Jones M.C. (1991) A Reliable Data-based Bandwidth Selection Method for Kernel Density Estimation. *Journal of the Royal Statistical Society Series B,* 53, pp. 683–690. Doi: 10.2307/2345597

Skoglund J. (2001) *A Simple Efficient GMM Estimator of GARCH Models*. SSE/EFI Working Paper Series in Economics and Finance 0434, Stockholm School of Economics.

Troster V., Tiwari A.K., Shahbaz M., Macedo D.N. (2019) Bitcoin Returns and Risk: A General GARCH and GAS Analysis. *Finance Research Letters,* 30, pp. 187–193. Doi: 10.1016/j.frl.2018.09.014

Wang J., Jiang Y., Zhu Y., Yu J. (2020) Prediction of Volatility Based on Realized-GARCH-kernel-type Models: Evidence from China and the U.S. *Economic Modelling,* 91, pp. 428–444. Doi: 10.1016/j.econmod.2020.06.004

Zhang X., King M. (2013) *Gaussian Kernel GARCH Models.* Monash Econometrics and Business Statistics Working Papers 19/13.

[dataset] Investing.com, 2021, Bitcoin historical data. Available at: https://www.investing.com/crypto/bitcoin/historical-data