Check for updates

Modelling and forecasting risk dependence and portfolio VaR for cryptocurrencies

Jie Cheng[1]

Abstract

In this paper, we investigate the co-dependence and portfolio value-at-risk of cryptocurrencies, with the Bitcoin, Ethereum, Litecoin and Ripple price series from January 2016 to December 2021, covering the crypto crash and pandemic period, using the generalized autoregressive score (GAS) model. We find evidence of strong dependence among the virtual currencies with a dynamic structure. The empirical analysis shows that the GAS model smoothly handles volatility and correlation changes, especially during more volatile periods in the markets. We perform a comprehensive comparison of out-of-sample probabilistic forecasts for a range of financial assets and backtests and the GAS model outperforms the classic DCC (dynamic conditional correlation) GARCH model and provides new insights into multivariate risk measures.

1 Introduction

During the last years, cryptocurrencies gain more and more attention not only from ordinary investors but also from regulatory authorities and policy makers. Cryptocurrencies are decentralized currencies that are powered by their users with no central authority and therefore are independent of monetary politics and not controlled by the

✉ Jie Cheng
    J.Cheng@keele.ac.uk

1   School of Computing and Mathematics, Keele University, MacKay Building, Keele ST5 5BG, UK

 Springer

existing banking system[1]. Bitcoin, the largest cryptocurrencies was created in 2009 and since then numerous other cryptocurrencies have been created. After a stable period of development, most of the cryptocurrencies started to climb and dramatically increased in the period 2016 to 2020 with pricing bubbles in 2018 (Corbet et al. 2018). After that, all major cryptocurrencies' prices have exhibited tremendous fluctuation with the sharpest drop during March 2020 selloff, as a result of the COVID-19 outbreak.

Existing literature on the cryptocurrencies market includes studies focusing on hedging and safe-haven properties of cryptocurrencies (e.g. Bouri et al. 2017; Conlon and McGee 2020), market efficiency (e.g. Nadarajah and Chu 2017; Tran and Leirvik 2020), volatility patterns and portfolio of cryptocurrency markets (Katsiampa 2017), most of which provide the within-sample fit for univariate cases. On the other hand, to account for the structure linkage and interdependencies among the cryptocurrencies and other financial assets, different multivariate approaches including the GARCH-DCC models (Guesmi et al. 2019; Ghabri et al. 2021), the GARCH-BEKK models (Katsiamp et al. 2019; Stavroyiannis and Babaros 2017) and GARCH-copula models (Bouri et al. 2018; Boako et al. 2019; Syuhada and Hakim 2020) have documented for volatility forecasting and risk management.

While these studies provide useful analyses, they also confirm that both the conditional volatilities and the correlations of the cryptocurrencies change over time, especially during the bubble period in 2018 and the pandemic era in 2020. Therefore, we pay attention to the observation-driven time-varying multivariate generalized autoregressive score (GAS) model to examine the price dependency relationships and portfolio value-at-risk (VaR) of cryptocurrencies; particularly, Bitcoin (BTC), Ethereum (ETH), Litecoin (LTC) and Ripple (XRP) are considered. The generalized autoregressive score-driving model (GAS) is proposed by Creal et al. (2013), and it nests many well-known models, including GARCH (Bollerslev 1986) and ACD (Engle and Russell 1998) models. Tafakori et al. (2018) consider an asymmetric exponential GAS model to predict Australian electricity returns. Chen and Xu (2019) use both univariate and bivariate GAS models to analyse and forecast volatilities and correlations between Brent, WTI and gold prices. To the best of our knowledge, no other study has ever used the multivariate GAS model to forecast the volatility and correlation of cryptocurrencies.

Due to the relatively young literature on cryptocurrency, there are few studies related to out-of-sample forecasting performance for both dependence structure and volatility. Amongst those, Syuhada and Hakim (2020) construct a dependence model through vine copula and provide the value-at-risk (VaR) forecasts. Chi and Hao (2021) show GARCH model's volatility forecast is better than the option implied volatility using the BTC and ETH prices. In our paper, we conduct out-of-sample forecasting performance for both point forecasts (e.g. VaR) and density forecasts. In order to see how effectively the GAS model treats different dynamic features simultaneously in a unified way, we compare the forecasting results with those of the classic dynamic conditional correlation generalized autoregressive conditional heteroskedasticity (DCC-GARCH) model (Engle 2002).

---

[1] Since 2019, China's central bank has announced that all transactions of cryptocurrencies are illegal, effectively banning digital tokens such as Bitcoin. As one of the consequences, the price of major cryptocurrencies has dropped sharply.

Our main findings are as follows: First, beside the most applied volatility models, GARCH, asymmetric GARCH specifications including GJR-GARCH and APARCH models are also considered for the univariate ETH, LTC, BTC and XRP return series. Interestingly, the additional parameters in these models, which are supposed to show the asymmetric volatility response to past returns (so-called leverage effect), are not significant for all the cryptocurrencies in this paper. These results are consistent with those found in Chi and Hao (2021) and Syuhada and Hakim (2020). Several studies apply the asymmetric GARCH models to cryptocurrencies' return series; however, they either use a GARCH-type model with Gaussian innovation (Cheikh et al. 2020) or show rather weak significant additional terms, which are supposed to reflect the asymmetry (Apergis 2021). One possible explanation is that the traders or investors from the cryptocurrency market are different to those from the stock market. Unlike the stock market which is usually dominated by well informed investors, the cryptocurrency market has more uninformed investors, and the volatility asymmetry, which can be traced to trading activity that has been guided by information asymmetry between well informed and uninformed traders in the market (Avramov et al. 2006), is not significant as it did in the stock market.

Second, we find empirical evidence to show that the forecasting ability of the GAS model is better than those of the DCC-GARCH model. More specifically, the GAS model accounts for large price changes in a very natural way when updating the correlations and volatilities over time, especially during extreme events. This is particularly important when we form a portfolio risk and estimate the corresponding VaR forecasts. Through a sequence of statistical tests, our results prefer the GAS model to the DCC-GARCH models in terms of point (volatilities and correlations) forecasts, quantile (value-at-risk) forecasts and density forecasts.

This paper is organized as follows: Section 2 describes the multivariate GAS model and the DCC-GARCH model. Section 3 provides the data source and preliminary analysis. In Sect. 4, we applied the two multivariate models to the daily cryptocurrencies and present the estimation results for the within-sample period. Moreover, we conduct out-of-sample forecasting performance for volatilities, correlations, VaRs and probability distributions for the two models. Section 5 concludes.

2 Empirical models

2.1 The multivariate GAS model

Let r_t be an N-dimensional random vector at time t with conditional distribution

$$r_t | F_{t-1} \sim p(r_t, \theta_t), \tag{1}$$

where F_{t-1} contains all the information up to time $t-1$, θ_t is a vector of time-varying parameters depending on F_{t-1} and a set of static parameters ϕ for all time t. The GAS(p,q) model is an observation-driven model, and the time-varying parameters θ_t are governed by the score of the conditional density in (1) and an autoregressive

updating equation

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\kappa} + \sum_{i=1}^{p} A_i \boldsymbol{s}_{t-i+1} + \sum_{j=1}^{q} B_j \boldsymbol{\theta}_{t-j+1}, \tag{2}$$

where $\boldsymbol{\kappa}$, A and B are the coefficient matrices with proper dimensions and $\boldsymbol{s}_t$ is the scaled score function

$$\boldsymbol{s}_t = \boldsymbol{S}_t \nabla_t(\boldsymbol{r}_t, \boldsymbol{\theta}_t), \tag{3}$$

with

$$\nabla_t = \frac{\partial}{\partial \boldsymbol{\theta}_t} p(\boldsymbol{r}_t, \boldsymbol{\theta}_t),$$
$$\boldsymbol{S}_t = I_t(\boldsymbol{\theta}_t)^{-\gamma},$$
$$I_t(\boldsymbol{\theta}_t) = E_{t-1}\left[\nabla_t \nabla_t^T\right] = -E_{t-1}\left[\frac{\partial^2 \log p(\boldsymbol{r}_t, \boldsymbol{\theta}_t)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right],$$

where the expectation is taken with respect to the conditional distribution in (1). The additional parameter γ is fixed. By choosing different values of γ, the GAS model encompasses some well-known models (e.g. GARCH, ACD and ACM models, see Creal et al. 2013, for a detailed discussion).

In the application, we consider a GAS(1,1) model with $\gamma = 0$ and the conditional distribution in (1) follows a multivariate standardized Student-t distribution (Ardia et al. 2019). Therefore, the time-varying parameter vector $\boldsymbol{\theta}$ (including location μ, scale σ, correlation ρ and shape ν parameters) is given by:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\kappa} + A\boldsymbol{s}_t + B\boldsymbol{\theta}_t,$$

and a natural choice for S_t is identity matrix.

2.2 The multivariate DCC-GARCH model

Following Engle (2002), the DCC-GARCH(1,1) model is as follows. Let $\boldsymbol{r}_t$ be an N-dimensional random vector at time t, we consider

$$Var(\boldsymbol{r}_t | F_{t-1}) = Q_t = D_t R_t D_t, \tag{4}$$

where F_{t-1} is the information available up to time $t-1$, D_t is a diagonal matrix such that $D_t = \text{diag}(\sqrt{h_{11,t}}, \cdots, \sqrt{h_{nn,t}})$ and $h_{ii,t}$, $i = 1, 2, \cdots, N$ is the conditional variance obtained from the univariate model, which is usually GARCH-type model and R_t is the dynamic conditional correlation matrix. More specifically, let

$$\boldsymbol{r}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\psi}_t, \tag{5}$$

$$\boldsymbol{\psi}_t = Q_t^{1/2}\boldsymbol{\varepsilon}_t, \tag{6}$$

then the time-varying correlation matrix Q_t can be updated by

$$Q_t = (1 - a - b)\bar{Q} + a\mathbf{Z}_{t-1}\mathbf{Z}_{t-1}^T + bQ_{t-1}$$

where $\bar{Q}$ is a symmetric time-invariant unconditional covariance matrix and $\mathbf{Z}_t = D_t^{-1}\boldsymbol{\varepsilon}_t$. In our application, we assume $\boldsymbol{\varepsilon}_t$ follows a multivariate standardized Student-t distribution, as we did in GAS(1,1) model.

3 Empirical application

Daily Cryptocurrencies data, Ethereum (ETH), Litecoin (LTC), Bitcoin (BTC) and Ripple (XRP), in US dollars, are obtained from https://www.cryptocompare.com[2] using a Python script. Our sample period is from 1 January 2016 till 31 December 2021. We split the sample into two parts, a within-sample period from 1 January 2016 to 31 December 2018, which includes a total of 1096 daily prices and out-of-sample period from 1 January 2019 to 31 December 2021. For each of the datasets, the returns r_t of ETH, LTC, BTC and XRP are calculated as

$$r_t = 100\left[(\log(P_t) - \log(P_{t-1}))\right],$$

where P_t is the daily closing price at time t.

Cryptocurrency returns are extremely volatile, so we winsorized them at the 0.005% and 99.5% levels. Figure 1 displays the winsorized return series for ETH, LTC, BTC and XRP during the full sample period, i.e. from January 2016 to December 2021. We observe multiple volatile periods for different returns series, but they behave more similarly after 2018. During the March 2020 selloff, all of them experienced the most negative changes. It is worth mentioning that XRP suffered significant price fluctuations during first half of 2021 due to an SEC lawsuit Ripple faced at the end of 2020. Therefore, volatility changes of XRP were mostly caused by updates on the SEC lawsuits after 2021. Table 1 reports the descriptive statistics for the ETH, LTC, BTC and XRP return series. All of them have positive mean returns and leptokurtic empirical distributions for both sample periods. Moreover, the skewness for BTC (XRP) is negative (positive) across the full sample, while ETH and LTC present positive skewness before 2019 and negative one after 2019. For all returns series, the augmented Dickey and Fuller statistics reject the unit root null at 1% significance level, in favour of the stationary time series. The normality is significantly rejected by the enormous Jarque–Bera statistics, indicating the fat-tailed distribution. Engle's ARCH test (Engle 1982) results reveal the significant ARCH effect, highlighting the application of GARCH-type models.

---

[2] CryptoCompare's real-time aggregate index methodology (CCCAGG) calculates the market price of cryptocurrency pairs traded across exchanges. Aggregating transaction data from more than 250 exchanges, CryptoCompare uses a 24-hour volume-weighted average for every currency pair.

Table 1 Descriptive statistics for ETH, LTC, BTC and XRP return series

Descriptive Statistics	In sample				Out-of-sample			
	ETH	LTC	BTC	XRP	ETH	LTC	BTC	XRP
Mean	0.449	0.181	0.192	0.3255	0.325	0.158	0.255	0.0894
Std.	6.534	5.774	4.016	7.1633	4.843	5.350	3.703	5.9866
Min	−31.350	−36.358	−18.256	−42.5689	−31.350	−36.358	−18.256	−42.5689
Max	27.802	35.116	17.983	49.0847	23.069	26.873	17.182	41.1113
Skewness	0.228	0.939	−0.260	1.7132	−0.532	−0.531	−0.007	0.1702
Kurtosis	6.545	11.362	6.556	16.1768	8.233	9.226	6.256	16.0209
JB stat($\times 10^3$)	0.583**	3.354**	0.590**	8.465**	1.302**	1.821**	0.484**	7.748**
ADF(5) stat	−14.176**	−14.018**	−14.305**	−11.191**	−14.645**	−14.923**	−14.465**	−12.623**
ARCH(5) stat	102.981**	66.626**	92.205**	123.750**	57.904**	41.339**	27.665	72.759**

JB stat, ADF stat and ARCH stat are the statistics testing for normal distribution, stationarity and heteroskedastic effects, respectively. The selected lag number is in parentheses.

*Denote rejections of null hypothesis at 5% significance level.

**Denote rejections of null hypothesis at 1% significance level
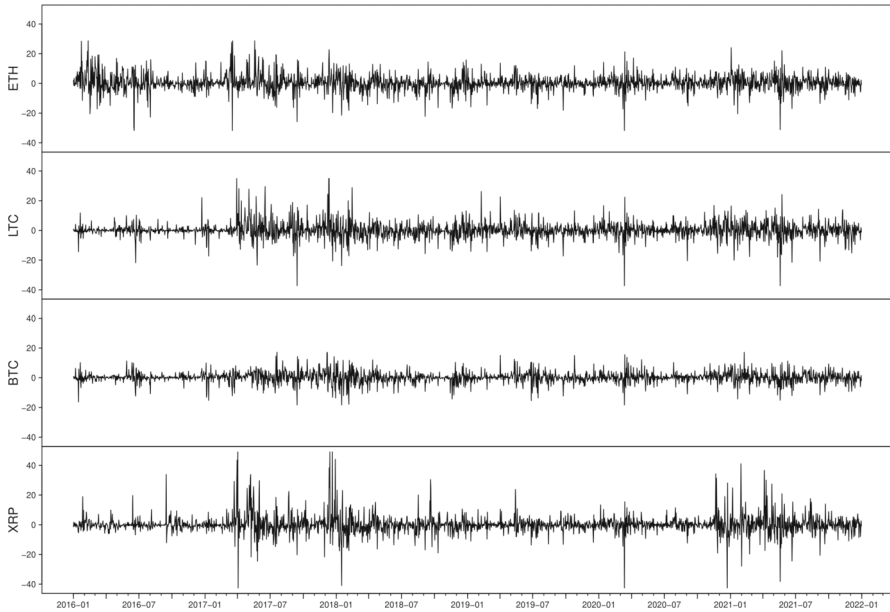
**Fig. 1** Cryptocurrency return series from January 2016 to December 2021

Following Tang and Xiong (2012), we first study the full sample rolling uncon-
ditional correlations between the ETH, LTC, BTC and XRP return series using a
bivariate approach. We rescale the return series by subtracting their means and divid-
ing by their standard deviations and specify the regression of the rescaled return $r^r_{m,t}$
on the rescaled return $r^r_{l,t}$, with $l, m = 1, 2, 3, 4$ and $l \neq m$:

$$r^r_{m,t} = \mu + \tilde{\rho} r^r_{l,t} + \eta_t$$

and $\hat{\tilde{\rho}}$ is the estimated unconditional correlation between the two cryptocurrencies
returns r_m and r_l. The time-varying estimated correlation is obtained by using a rolling
window of fix length equal to 30 days. The rolling correlations of full-sample return
series are plotted in Fig. 2.

Before 2017, the correlation between BTC and LTC stays high and positive while
those between ETH, BTC and XRP are low and negative. This is not surprising as
Litecoin was one of the first "altcoins" to draw from Bitcoin's original open-source
code to create a new cryptocurrency, therefore one of the most correlated altcoins with
Bitcoin, while Ethereum is launched based on the platform which enables building and
deploying smart contracts and decentralized applications, and compete against Bitcoin
for market shares; XRP is created as a faster, cheaper, and more energy-efficient digital
asset that can process transactions within seconds and consume less energy than some
counterpart cryptocurrencies.

From the beginning of 2017 to the middle of 2018, distinct spikes in the correla-
tion can be generally found between the cryptocurrencies. Such spikes may reflect
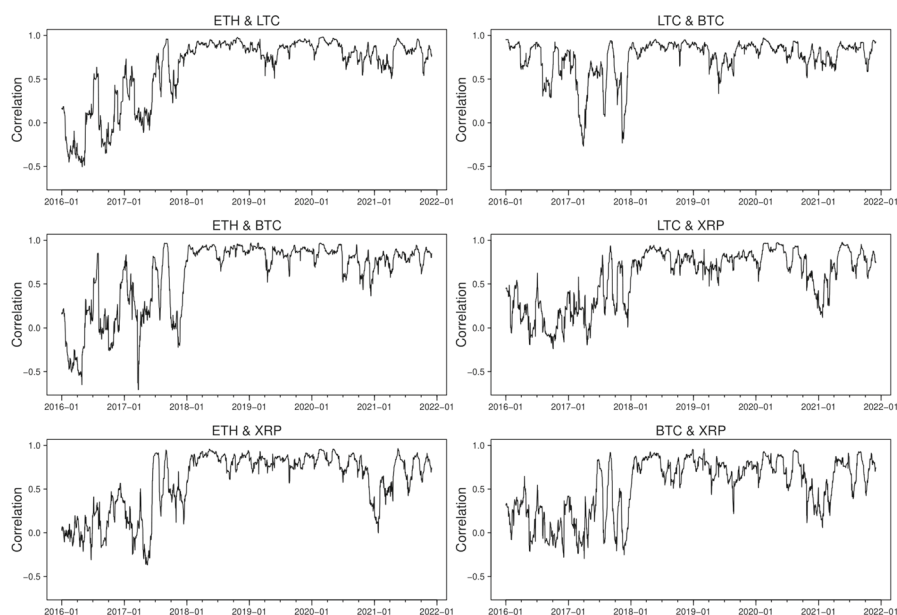
Fig. 2 Estimated unconditional correlations (30-day rolling window) from January 2016 to December 2021

the presence of significant uncertainty during the stage of the development of cryptocurrency market. All the correlations drastically go up at the middle of 2018 and remain positive and strong until the end of the sample. This finding is in line with the current literature (Katsiampa 2019; Katsiampa et al. 2019; Chowdhury et al. 2022; Pace and Rao 2023), and the connectedness between the cryptocurrencies is mainly caused by market uncertainty in response to the 2018 cryptocurrency crash (Aslanidis et al. 2019 and Antonakakis et al. 2019) and the launch on 10 December 2017 of the Bitcoin futures contracts at the Chicago Board Options Exchange (Blau et al. 2020). Moreover, a significant drop in rolling window correlations can be observed at the beginning of 2021 in the cryptocurrency pairs ETH-XRP, LTC-XRP, and BTC-XRP. Again, this is due to the SEC lawsuit Ripple faced. The above bivariate approach considers two return series at a time, as such, cannot exploit the dynamic interdependence simultaneously. To address this issue, we consider the multivariate GAS and DCC models in the next section.

3.1 In-sample results

For notational convenience, let $r_t = (r_1, r_2, r_3, r_4)$ be the returns of the four assets ETH, LTC, BTC and XRP at time t and $\rho_{12}, \rho_{13}, \rho_{14}, \rho_{23}, \rho_{24}$ and ρ_{34} be the correlation of the return series ETH and LTC, ETH and BTC, ETH and XRP, LTC and BTC, LTC and XRP, and BTC and XRP, respectively. We use the multivariate GAS(1,1) model and the DCC-GARCH(1,1) model (hereafter GAS and DCC) we mentioned in the last section to fit the multivariate return series r_t, respectively. Based on the

Table 2 The LR test results for the multivariate GAS model

	Model 1	Model 2	Model 3	Model 4	Model 5
Log LL	-12265.63	-11682.97	-11540.51	-11540.80	-11533.75
No. of the param	15	17	19	21	23
LR test statistic		1165.32	284.92	0.58	14.10
p value		0.000	0.000	0.748	0.001

Model 1 is constant-parameter GAS model, Model 2 indicates time-varying volatility only model, Models 3, 4 and 5 are volatility and correlation time-varying model, volatility, correlation and location time-varying model and all time-varying model. If Model 4 is a constant-location only model, then the LR statistic is 3.62 (p value is 0.164)

fat-tail leptokurtic empirical distributions we obtained in Table 1, the conditional distribution of r_t in the GAS model is specified by the multivariate standardized Student-t distribution; the univariate and multivariate residuals in the DCC model are also specified by the t-distribution.

Asymmetric GARCH specifications including GJR and EGARCH models are also considered for both GAS and DCC models. Interestingly, the additional parameters in these models, which are supposed to show the asymmetric volatility response to past returns (so-called leverage effect), are not significant for all the cryptocurrencies in this paper. These results are consistent with those found in Chi and Hao (2021) and Syuhada and Hakim (2020). Several studies apply the asymmetric GARCH models to cryptocurrencies' return; however, they either use the GARCH-type model with Gaussian innovation (Cheikh et al. 2020) or show rather weak significant additional terms which are supposed to reflect the asymmetry (Apergis 2021).

For the GAS model, the conditional distribution parameters are as follows:

$$\boldsymbol{\theta} = (\mu_1, \mu_2, \mu_3, \mu_4, \sigma_1, \sigma_2, \sigma_3, \sigma_4, \rho_{12}, \rho_{13}, \rho_{14}, \rho_{23}, \rho_{24}, \rho_{34}, \nu)$$

where $(\mu_1, \mu_2, \mu_3, \mu_4)$, $(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$, $(\rho_{12}, \rho_{13}, \rho_{14}, \rho_{23}, \rho_{24}, \rho_{34})$, ν are location, scale/volatility, correlation and shape parameters of the conditional t-distribution, respectively. Following (Chen and Xu 2019), we conduct a series of likelihood ratio test (LRT) to see whether these parameters are time varying or not. We are interested in the null hypothesis $H_0 : M = M_i$ versus to the alternative hypothesis $H_1 : M = M_{i+1}$ for $i = 1, 2, 3, 4$, where Model 1 to Models 5 are a series of nested time-varying parameters models, i.e. Model 1 assumes all the parameters are time-invariant, Model 2 is the time-varying volatility-only model and Model 5 is the time-varying volatility, correlation, location and shape model. Clearly, $M_1 \subset M_2 \subset M_3 \subset M_4 \subset M_5$, and under the regular conditions, the test statistic LRT shall follow a Chi-square distribution χ_k^2 with degree of freedom k if H_0 is true. The LRT test results are listed in Table 2. It is clear that model 5 seems to be a reasonable choice, i.e. the GAS model with time-varying volatility and correlation, location and shape model is used for the return series r_t during 2016 to 2019.

The estimation results are presented in Table 3. All the parameters, especially the time-varying parameters of the model (left panel), are significant at the 5% level. We also present the unconditional parameters (right panel) by considering the long-term

Table 3 Parameters estimation of the GAS model

Parameters	Estimate	p value	Parameters	Unconditional
κ_{μ_1}	0.001	0.008	μ_1	0.068
κ_{μ_2}	−0.001	0.000	μ_2	−0.075
κ_{μ_3}	0.004	0.000	μ_3	0.234
κ_{μ_4}	−0.004	0.000	μ_4	−0.243
κ_{σ_1}	0.027	0.000	σ_1	4.337
κ_{σ_2}	0.014	0.000	σ_2	2.127
κ_{σ_3}	0.016	0.000	σ_3	2.374
κ_{σ_3}	0.019	0.000	σ_3	2.895
$\kappa_{\rho_{12}}$	0.009	0.000	ρ_{12}	0.388
$\kappa_{\rho_{13}}$	0.009	0.000	ρ_{13}	0.384
$\kappa_{\rho_{14}}$	0.009	0.000	ρ_{14}	0.442
$\kappa_{\rho_{23}}$	0.009	0.000	ρ_{23}	0.510
$\kappa_{\rho_{24}}$	0.009	0.000	ρ_{24}	0.499
$\kappa_{\rho_{34}}$	0.010	0.000	ρ_{34}	0.463
κ_ν	−0.585	0.000	ν	4.000
a_μ	0.012	0.000		
a_σ	0.058	0.000		
a_ρ	0.015	0.000		
a_ν	0.414	0.000		
a_μ	0.984	0.000		
a_σ	0.981	0.000		
a_ρ	0.992	0.000		
a_ν	0.982	0.000		

This table presents the parameter estimation results for ETH, LTC, BTC and XRP return series during January 2016 to December 2019 (within-sample period). The model assumes time-varying volatility and correlation, location and shape. The unconditional parameters are extracted using identity scaling

values of the parameters, i.e. $(I - \hat{B})^{-1}\hat{\kappa}$. With regard to the DCC model, similar estimation results are reported in Table 4. The parameters can be divided into two parts, the results of the GARCH model for each individual return series (upper panel) and the dynamic correlation using multivariate t distribution (lower panel).

In Figs. 3, 4, 5 and 6, we plot the estimated volatilities for ETH, LTC, BTC and XRP using both GAS and DCC models during the in-sample period, respectively. For all four return series, the DCC model seems to provide more fluctuant volatilities than the GAS model, especially during the 2018 crash period. Clearly, the extreme returns appear to have a strong effect on estimated volatilities for the GARCH models, whereas those for the GAS model appear to be robust.

The correlation estimates from the two models, which are presented in Figs. 7 and 8, show a substantial difference though both models identify a significant persistence of correlations in high positive values between the cryptocurrencies since 2018. The GAS model suggests, in general, positive correlations, varying from -0.15 to 1 between

Table 4 Parameters estimation of the DCC model

Parameters	ETH	p value	LTC	p value	BTC	p value	XRP	p value
μ	0.016	0.885	−0.029	0.595	0.200	0.000	−0.274	0.410
ω	2.880	0.055	0.116	0.252	0.077	0.250	0.579	0.012
α	0.274	0.000	0.124	0.000	0.128	0.000	0.183	0.000
β	0.725	0.000	0.874	0.000	0.870	0.000	0.816	0.000
ν	3.301	0.000	3.096	0.000	3.435	0.000	3.000	0.000
Joint parameters								p value
a			0.059					0.000
b			0.934					0.000
ν			4.000					0.000

This table presents the parameter estimation results for ETH, LTC, BTC and XRP return series during January 2016 to December 2019 (within-sample period). For each individual return series, GARCH model (with t innovation) is used and the dynamic correlation is captured using multivariate t distribution

three series, while the DCC model gives correlations fluctuating substantially over time, falling to extreme values around -0.6 during June 2016, which is mainly caused by the instability of the Ethereum prices due to the DAO hack. It is worth noting that the dynamic correlations we derive from DCC multivariate modelling approach appear to be similar to the rolling correlations we estimate in the previously described bivariate setting while those by GAS approach seem to produce more smoothed correlation estimates due to its desirable robust future.

3.2 Out-of-sample results

We now turn to the out-of-sample (OOS) forecast performance of the two models. We compare the one-step-ahead forecasting performance of the GAS model and DCC model using a rolling window scheme. The length of the rolling estimation window is set to be 1096 observations, such that 1096 observations (from January 1 2019, until December 31 2021) are left for out-of-sample forecast evaluation.

### 3.2.1 Volatility and correlation forecast evaluation

To evaluate the forecasting performance of the two models, we construct two measures of realized volatility and correlation using intraday data. The realized volatility is computed as the sum of intraday returns (see, e.g. Andersen et al. (2001)),

$$RV_t = \sum_{i=1}^{N_t} r_{t,i}^2 \tag{7}$$

where $r_{t,i}$ is the intraday return on day t for intraday period i ($i = 1, 2, \cdots, N_t$). We use transaction prices of ETH, LTC, BTC and XRP from January 2019 to December
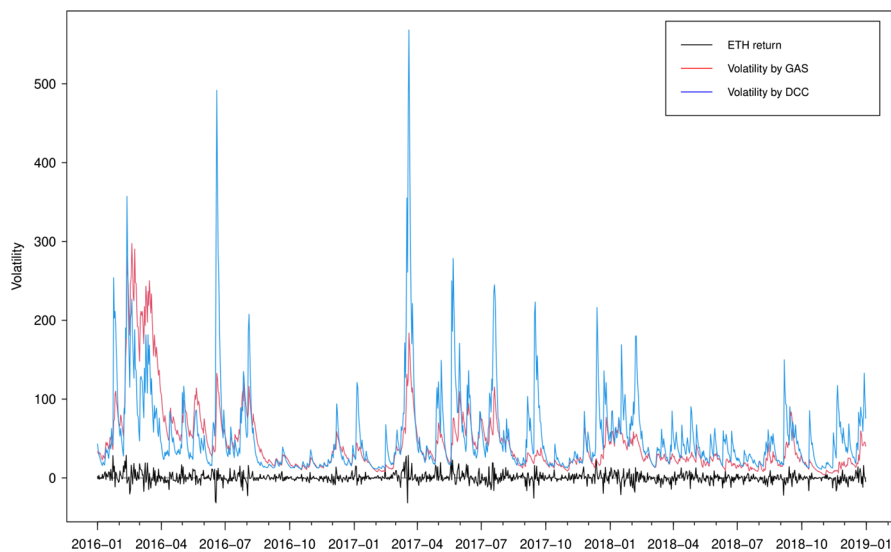
Fig. 3 Estimated volatilities of the ETH return using GAS and DCC models
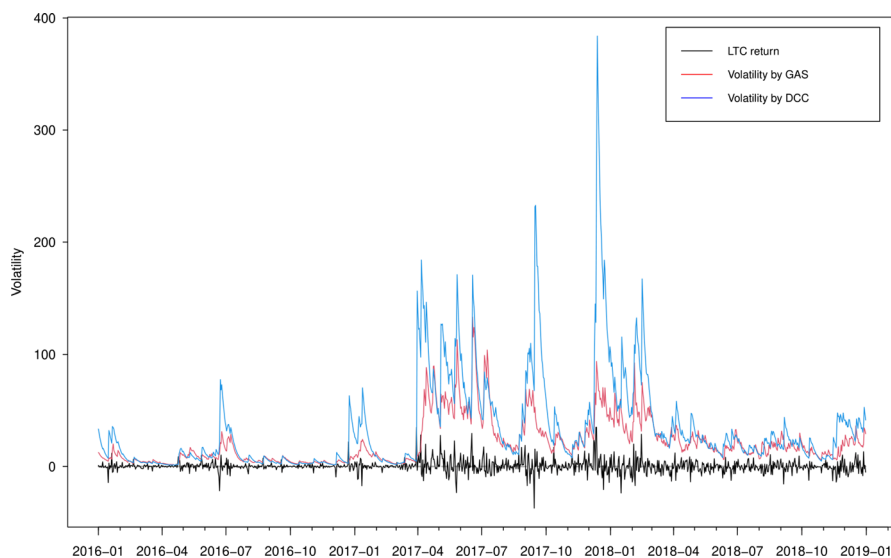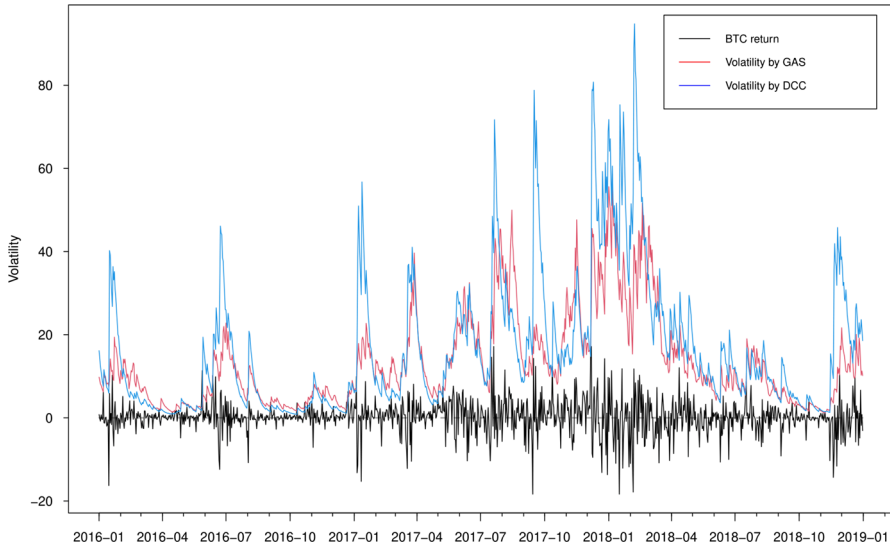

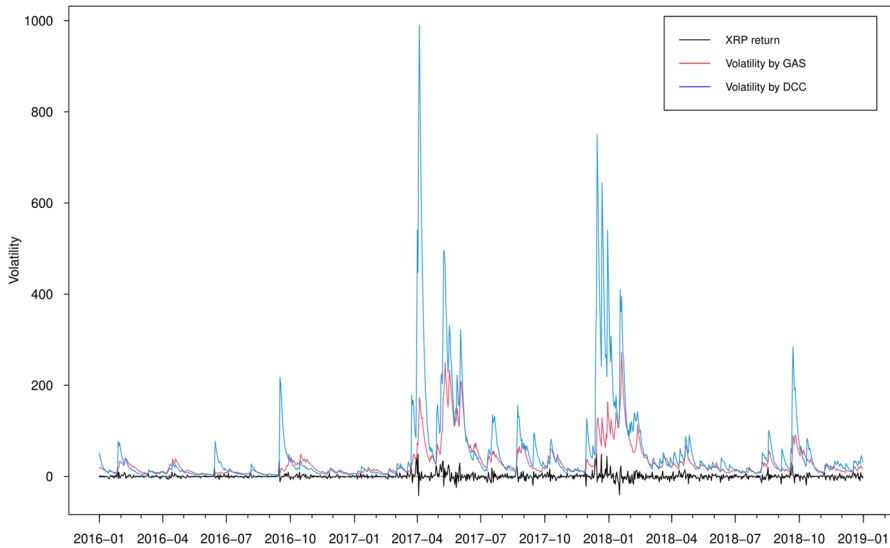**Fig. 4** Estimated volatilities of the LTC return using GAS and DCC models

2021, sampled in calendar time and tick-time with 5-minute sampling frequency[3]. The intraday return data are obtained from Bitfinex exchange[4], using a Python code. The

---

[3] Liu et al. (2015) finds that it is difficult to significantly beat the realized volatility (RV) using 5-minute intervals.

[4] we downloaded the intraday data from different exchanges (e.g. Bitstamp, Coinbase) and the results using these data are the same.

Fig. 5 Estimated volatilities of the BTC return using GAS and DCC models

Fig. 6 Estimated volatilities of the XRP return using GAS and DCC models

realized correlation[5] is calculated as:

$$RC_{xy,t} = \frac{\sum_{i=1}^{N_t} r_{x,t,i} r_{y,t,i}}{\sqrt{RV_{x,t}} \sqrt{RV_{y,t}}}$$

---

[5] Andersen et al. (2001) introduced the realized covariance and the realized correlation comes from the realized covariance divided by the square roots of the realized volatilities in (7).
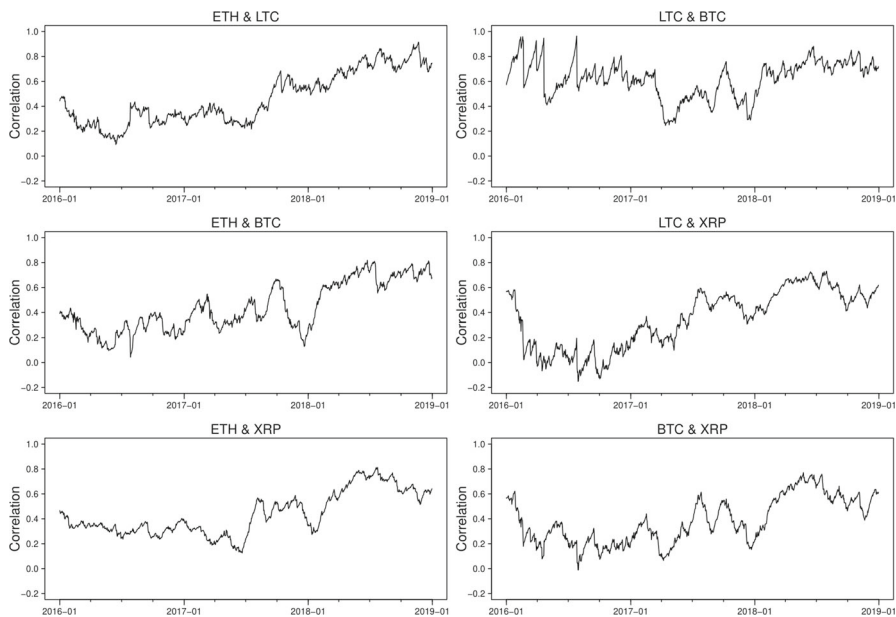
Fig. 7 Estimated correlation using GAS models

where $r_{x,t,i}$ and $r_{y,t,i}$ are the intraday return series for cryptocurrencies X and Y on day t for intraday period i $(i = 1, 2, \cdots, N_t)$ and $RV_{x,t}$ and $RV_{x,t}$ are the realized volatility for X and Y on day t.

Following (Patton 2011), we use two popular and robust loss functions, mean square error (MSE) and Gaussian quasi-likelihood (QLIKE) to compare the forecast accuracy of the GAS and DCC models on the out-of-sample data. These two loss functions are given by,

$$\text{MSE}_{\sigma^2} = \frac{1}{N} \sum_{i=1}^{N} \left(\sigma_i^2 - \hat{\sigma}_i^2 \right)^2, \quad \text{MSE}_{\rho} = \frac{1}{N} \sum_{i=1}^{N} \left(\rho - \hat{\rho}_i \right)^2 \tag{8}$$

and

$$\text{QLIKE}_{\sigma^2} = \frac{1}{N} \sum_{i=1}^{N} \left(\log(\hat{\sigma}_i^2) + \frac{\sigma_i^2}{\hat{\sigma}_i^2} \right), \quad \text{QLIKE}_{\rho} = \frac{1}{N} \sum_{i=1}^{N} \left(\log(\hat{\rho}_i) + \frac{\rho_i}{\hat{\rho}_i} \right), \tag{9}$$

where $\hat{\sigma}_i^2$, $\hat{\rho}_i$ are the rolling forecasts on volatility and correlation of day i by the two models, σ_i^2, ρ_i are the realized volatility and correlation at day i, respectively. N is the total number of volatility/correlation forecasts. We also use the (Diebold and Mariano 1995) method to test for the null hypothesis that the forecasts by the GAS model are less accuracy than or equal to the forecasts by the DCC model.

Table 5 reports the OOS losses for volatility and correlation, using the loss functions in (8) and (9), for the GAS and DCC models. The Diebold–Mariano statistics
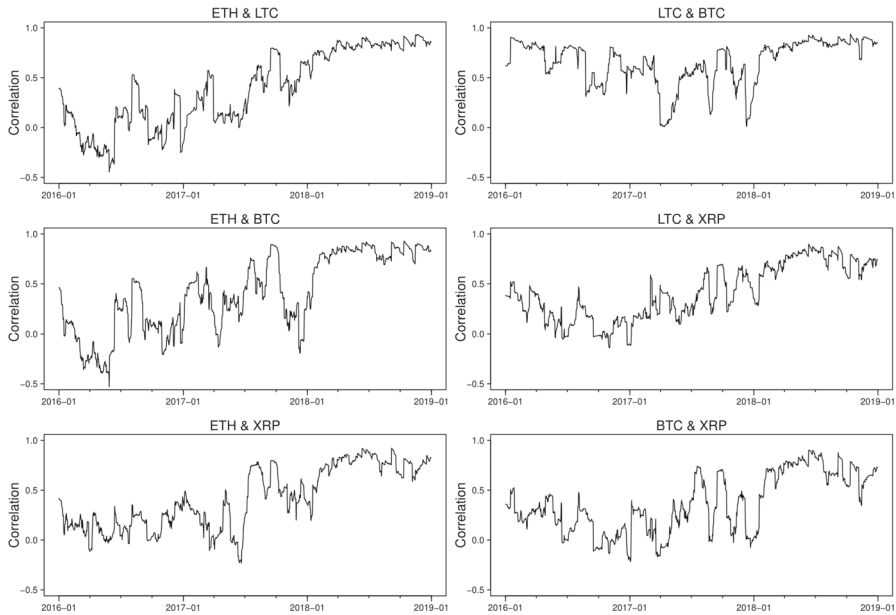
Fig. 8 Estimated correlation using DCC models

on the loss differences are also presented to see whether the gains are statistically significant. Overall, the forecasting ability of volatility and correlation in the GAS model is superior to those of the DCC model. Judging by the MSE and QLIKE, it is significant that the GAS model delivers substantially better correlation forecasts than the DCC model though the two models provide similar correlation forecasts between the BTC and XRP return series in terms of MSE.

The volatility forecasts comparison of MSE and QLIKE between the two models are mixed. The MSE favours the GAS model for all volatilities, while the QLIKE supports the GAS model for XRP volatility only. There is no evidence to show a significant difference of volatility forecasts for ETH, LTC and BTC in terms of QLIKE. These results can be further confirmed in the plots. The difference of correlation forecasts between the two models can be found across the whole OOS period (Figs. 9 and 10), while the volatility forecasts of BTC are similar for both models (Figs. 11, 12, 13 and 14). Noted that the DCC model continuously gives large volatility forecasts for all three return series when there are large changes in the return series.

Interestingly, we find that, on average, for both models, the dynamic correlation forecasts between cryptocurrencies behave similarly in all pairs. The correlations remain positive and at high levels with a few fluctuations across the whole OOS period using GAS model, while those using DCC models gives more sensitive dynamics, especially after January 2020. This could be considered as the consequence of the COVID-19 effect on cryptocurrencies. In particular, during January 2020 to May 2020, weak correlation forecasts can be observed between XRP and other cryptocurrencies using both models, which is, again, due to the SEC lawsuit.

Table 5 Results of out-of-sample forecasting accuracy

| | ETH volatility | | LTC volatility | | BTC volatility | |
	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
GAS	3960.097	3.624	6888.701	4.121	968.498	4.340
DCC	4135.622	3.679	7034.982	4.145	971.500	4.396
DM	−2.292*	−0.443	−1.686*	−0.961	−2.817*	−0.840

| | XRP volatility | | | | | |
	MSE	QLIKE				
GAS	18960.519	4.272				
DCC	19618.484	4.456				
DM	−2.471*	−1.892*				

| | ETH LTC correlation | | ETH BTC correlation | | ETH XRP correlation | |
	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
GAS	0.296	0.242	0.229	0.366	0.176	0.372
DCC	0.238	0.261	0.356	0.397	0.292	0.376
DM	−1.849*	−5.499*	−3.252*	−3.076*	−4.722*	−4.731*

| | LTC BTC correlation | | LTC XRP correlation | | BTC XRP correlation | |
	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
GAS	0.188	0.072	0.189	0.116	0.223	0.133
DCC	0.319	0.124	0.244	0.157	0.224	0.147
DM	−4.194*	−3.007*	−2.264*	−4.778*	−0.010	−2.149*

This table presents the t-statistics from Diebold–Mariano (DM) tests of equal predictive accuracy for the rolling window out-of-sample forecasts with different models using corresponding loss functions. A t-statistic greater than 1.96 in absolute value indicates a rejection of the null of equal predictive accuracy at the 0.05 level. These statistics are marked with an asterisk. The sign of the t-statistics indicates which forecast performed better for each loss function: a negative t-statistic indicates that the GAS forecast produced smaller average loss than the DCC forecast, while a positive sign indicates the opposite

3.2.2 Density forecast evaluation

To conduct further the comparison experiment, we use the estimated results for each of the models in the previous section to get one-step-ahead density forecasts and the evaluation is based on scoring rules, which are widely used in weather and climate prediction (Palmer 2012) and financial risk management (Groen et al. 2013). Let $\mathbf{y} = (y^{(1)}, \cdots, y^{(N)})$ be an observation of the N-dimensional random vector, let $f(.)$ denote a forecast density of $\mathbf{y}$, let Ω denote the set of possible values of $\mathbf{y}$, and let $\mathcal{F}$ denote a convex class of probability distribution on Ω. A scoring rule is a loss function:
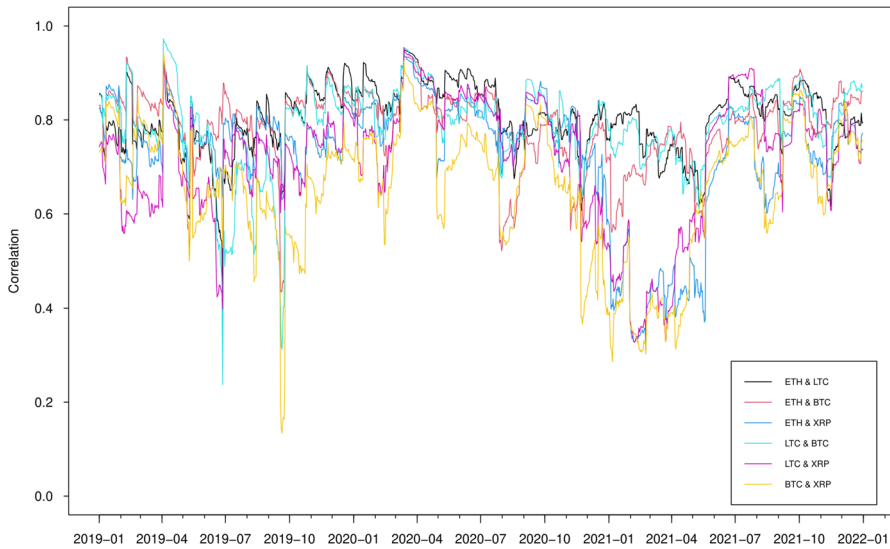
$$S(f, y) : \mathcal{F} \times \Omega \to \mathbb{R} \cup \{\infty\}$$

such that better forecast yields a lower score. A scoring rule S is said to be proper if the expected score is optimized, while the true distribution of the observation is issued
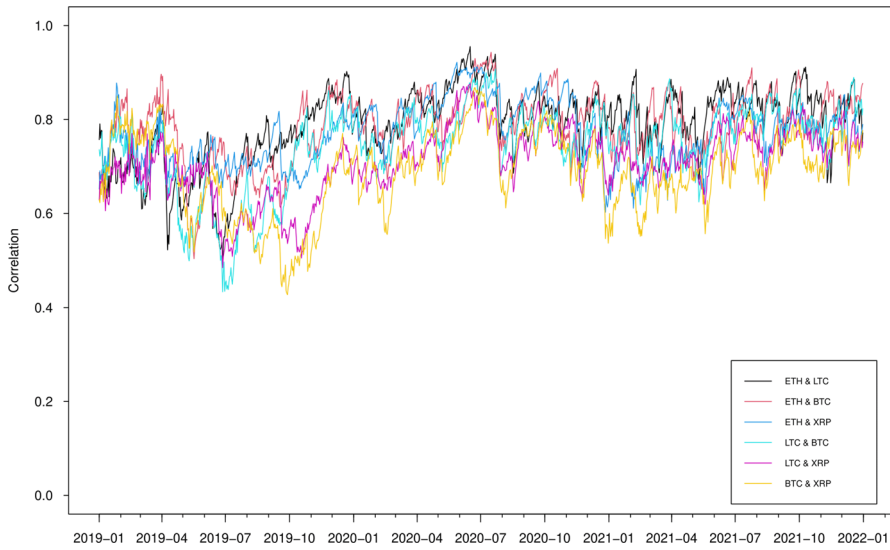
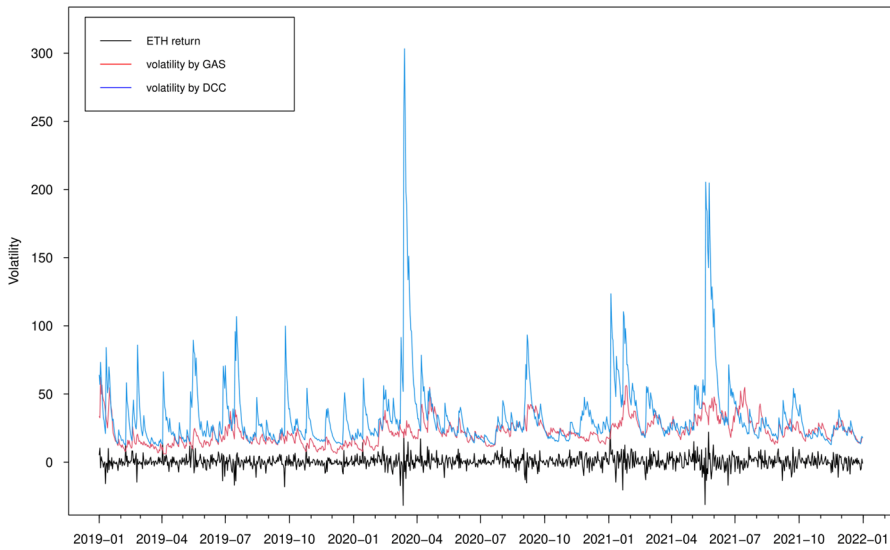**Fig. 9** Out-of-sample estimated correlation using GAS models

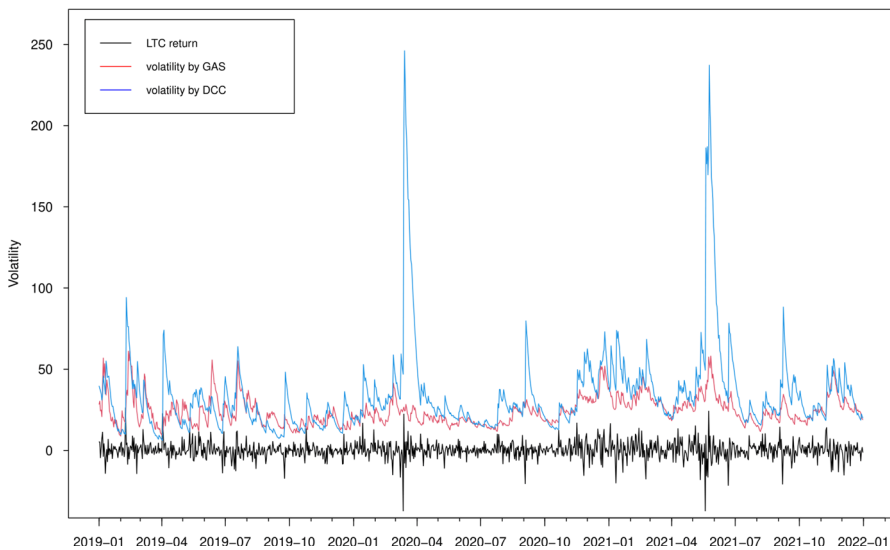**Fig. 10** Out-of-sample estimated correlation using DCC models

as a forecast, i.e.

$$\mathbb{E}_g S(g, \cdot) \leq \mathbb{E}_g S(f, \cdot) \tag{10}$$

for all $f, g \in \mathcal{F}$. Furthermore, a scoring rule is called strictly proper if equality (10) holds only if $f = g$.

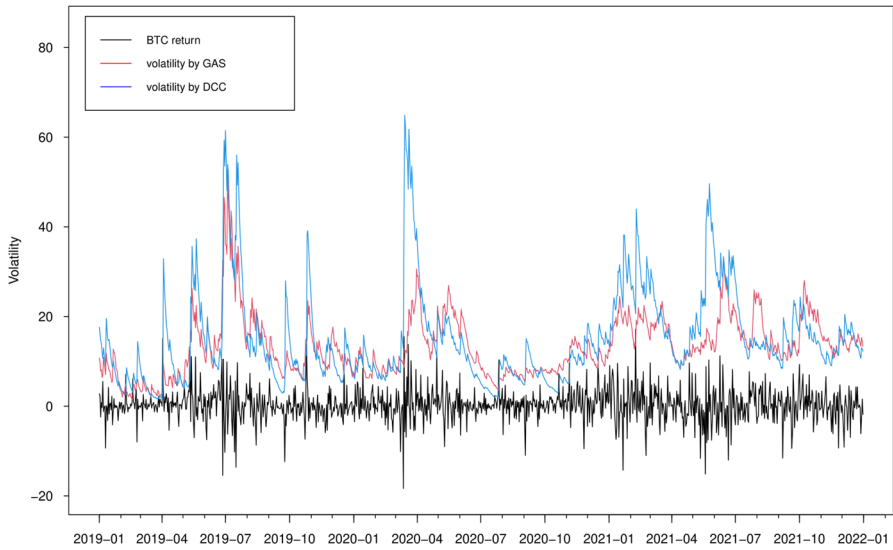**Fig. 11** Out-of-sample estimated volatilities of the ETH return using GAS and DCC models

Fig. 12 Out-of-sample estimated volatilities of the LTC return using GAS and DCC models

A natural approach is the logarithmic score (Good 1952; Mitchell and Hall 2005; Amisano and Giacomini 2007), which is defined as:

$$LogS(f, y) = -\log f(\mathbf{y}). \tag{11}$$

However, the logarithmic score is not sensitive to distance, which means it only rewards the predictive densities for assigning high probabilities to realized values but

Fig. 13 Out-of-sample estimated volatilities of the BTC return using GAS and DCC models
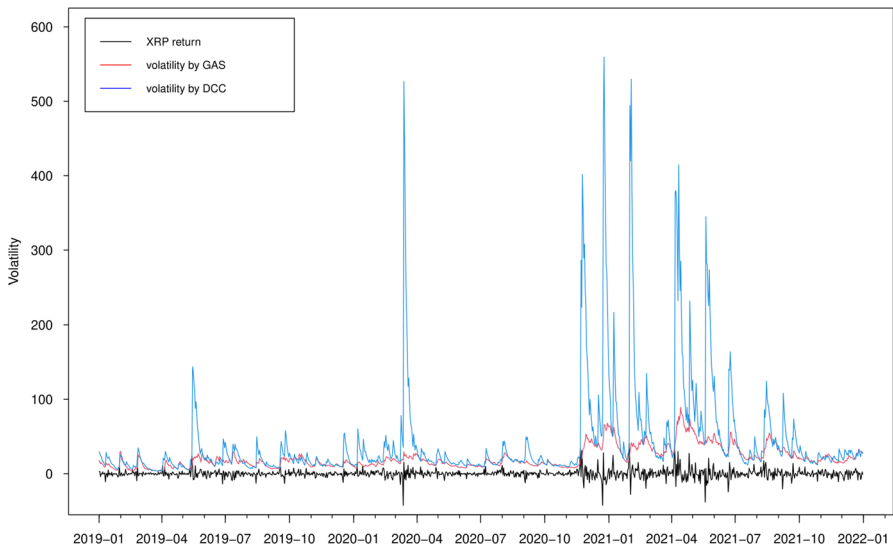

Fig. 14 Out-of-sample estimated volatilities of the XRP return using GAS and DCC models

not the neighbourhood values. To overcome this problem, (Gneiting and Raftery 2007) introduce the energy score which is a generalization of the univariate continuous ranked probability score (CRPS) and allows for a direct comparison of density forecasts. The energy score is defined as:

$$ES(f, y) = E\left(\|Y - \mathbf{y}\|^\beta\right) - \frac{1}{2}E\left(\|Y - \tilde{Y}\|^\beta\right) \tag{12}$$
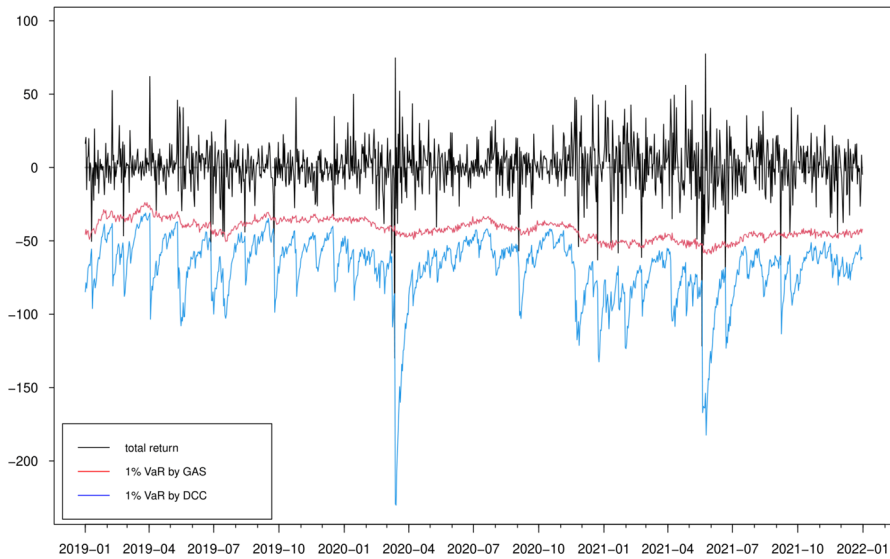
Fig. 15 Estimated 1% value-at-risk (VaR) for portfolio 1
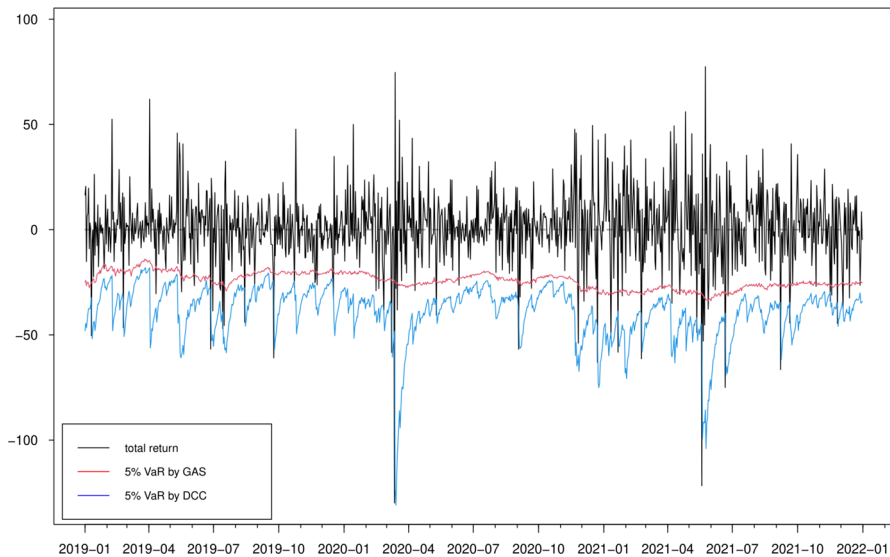
Fig. 16 Estimated 5% value-at-risk (VaR) for portfolio 1

where $\tilde{Y}$ is an independent copy of Y, so it is drawn independently from the same distribution $f(.)$ as Y, $\|.\|$ is the Euclidean norm. Gneiting and Raftery (2007) show that the energy score is strictly proper with $\beta \in (0, 2)$. In application, $\beta = 1$ seems to be a standard choice and the score is usually calculated through Monte Carlo methods.

Pinson and Tastu (2013) show that the discrimination ability of energy score may be limited, while the dependence structure of multivariate probabilistic forecasts is

misspecified. To overcome this problem, Scheuerer and Hamill (2015) propose the variogram score which is based on pairwise differences:

$$VS(f, y) = \sum_{i,j=1}^{N} w_{ij} \left(|y_i - y_j|^p - E|x_i - x_j|^p \right)^2 \qquad (13)$$

where N is the dimension of random vector $\mathbf{y}$, x_i and x_j are the ith and jth component of a random vector $\mathbf{x}$ that is from the distribution f, w_{ij} are nonnegative weights that allows one to emphasize pairs of component combinations and standard choice for weights is $w_{ij} = 1$. $p > 0$ is the order of the variogram score. The variogram score is proper relative to the class of distributions for which the $2p$-th moments of all elements are finite and it is not strictly proper (Scheuerer and Hamill 2015). In application, the choice of p is a trade-off between all relative moments of the pairwise deviation and outliers. Typical choices of p include 0.5 and 1.

To test the null hypothesis of equal predictive ability of two competing models based on a given scoring rule, we consider (Diebold and Mariano 1995) type tests using score difference. Given a scoring rule S, the score difference is defined as:

$$d_t = S(\hat{f}_1, \mathbf{y}_t) - S(\hat{f}_2, \mathbf{y}_t)$$

where $\hat{f}_1$ and $\hat{f}_2$ are the density forecasts. The null hypothesis of equal scores is:

$$H_0 : E(d_t) = 0, \text{ for all } t$$

versus the alternative $H_1 : E(d_t) \neq 0$. It can be shown that, under the null hypothesis, with certain conditions (e.g. see Giacomini and White 2006), the statistic

$$DM = \frac{\bar{d}}{\sqrt{\hat{\sigma}^2/n}} \rightarrow N(0, 1) \qquad (14)$$

where n is the forecast sample size, $\bar{d} = \frac{1}{n} \sum_{t=1}^{n} d_t$ and $\hat{\sigma}^2$ is a heteroskedasticity and autocorrelation-consistent variance estimator of $\sigma^2 = var(\sqrt{n}\bar{d})$.

We applied the above three scores to evaluate and compare the density forecasts by GAS and DCC models. For variogram score, we present the results with different p values ($p = 0.5, 1, 2$) as used in Scheuerer and Hamill (2015)). The overall density forecast can be evaluated using average score $\bar{d}$ during the whole out-of-sample period [6] and the DM statistics are obtained using the log score in (11), the energy score in (12) and the variogram score in (13). The score difference d_t is computed by subtracting the score of the DCC model density forecast from the score of the GAS density forecast, such that negative values of d_t indicate the better predictive ability of the forecast method based on the GAS model. Table 6 shows the average score differences $\bar{d}_n$ with

---

[6] Noted that all the scores we discussed above are proper (Gneiting and Raftery 2007), which means that any incorrect density forecasts $\hat{f}_t$ do not receive a lower average score (negatively oriented score) than the true density.

Table 6 Average score differences and tests of equal predictive accuracy

Scoring rule		$\bar{d}$	DM test stat.	p value
Log score		0.004	0.114	0.454
Energy score		-0.198	-6.506	0.000
Variogram score	$p = 0.5$	-1.742	-6.380	0.000
Variogram score	$p = 1$	-31.980	-3.882	0.000
Variogram score	$p = 2$	-38247.031	-4.355	0.000

The table presents the average score difference $\bar{d}^*$ and the corresponding test statistics (with p values in the parentheses) for the log score in (11), the energy score in (12) and the variogram score in (13). The variogram scores are presented with $p = 0.5, 1$ and 2. The score difference d_t is computed for density forecasts obtained from a GAS model with multivariate t innovations relative to the DCC model with same innovations, for daily ETH, LTC, BTC and XRP returns over the evaluation period 1 January 2019–31 December 2021

the accompanying tests of equal predictive accuracy as in (14). These results clearly demonstrate that both energy and variogram scoring rules suggest superior density predictive ability of the GAS model. The large values of average variogram score difference with $p = 2$ are caused by the nature of quadratic form, and the results are in accord with the simulation studies by Scheuerer and Hamill (2015).

From the risk management point of view, it is also important to focus on the performance of density forecasts in the region of interest. Therefore, we compare the models in terms of correctly forecasting the 1% and 5% value-at-risk (VaR) at 1-day horizons for both individual cryptocurrencies and different portfolios that can be constructed from the three cryptocurrencies. We define five different arbitrary portfolios, $p_{jt} = g_j r_t$ for given 4×1 weight vectors g_j and for $j = 1, 2, 3, 4, 5$. By ordering the cryptocurrencies as ETH, LTC, BTC and XRP, we construct the following long-only and long-short portfolios: $g_1 = (1/4, 1/4, 1/4, 1/4)$, $g_2 = (1/4, 1/4, 1/4, -1/4)$, $g_3 = (1/4, 1/4, -1/4, 1/4)$, $g_4 = (1/4, -1/4, 1/4, 1/4)$ and $g_5 = (-1/4, 1/4, 1/4, 1/4)$. The long-short positions reflect the relative value bets among these cryptocurrencies.

We simulate 10000 sample paths for $\mathbf{r}_{t+1} = (r_1, r_2, r_3, r_4)'$, denoted by $\mathbf{r}_{t+1}^s$ for $s = 1, 2, \cdots, 10000$ using the multivariate t distribution by the GAS and DCC models. We then construct the simulated individual returns $r_{i,t+1}^s$ for $i = 1, 2, 3, 4$ and portfolio returns $p_{j,t+1}^s = g_j' \mathbf{r}_{t+1}^s$ for $j = 1, 2, 3, 4, 5$. We use the sample of 10000 simulated paths to estimate the quantiles of the forecasting distribution at the 1-day horizon. The out-of-sample VaR accuracy is assessed through the unconditional coverage (UC) test (Kupiec 1995) and the conditional coverage (CC) test (Christoffersen 1998).

Table 7 presents the UC and CC test statistics and the corresponding p values of the 5% and 1% VaR forecasts for both individual returns (upper panel) and four portfolios (lower panel). For the individual VaR forecasts, all results, except for BTC returns series, suggest that GAS model performs better than DCC model at the 1% and 5% quantile levels. The GAS and DCC models provide same results for the BTC return: the 1% VaRs forecasts perform reasonably well, but the 5% VaR forecasts are rejected for both tests. Meanwhile, the GAS model outperforms the DCC model in general

Table 7 Results of out-of-sample VaR forecasting performance

		5% VaR		1% VaR	
		UC test	CC test	UC test	CC test
ETH	GAS	9.192 (0.003)	10.310 (0.006)	0.362 (0.547)	2.522 (0.283)
	DCC	44.968 (0.000)	45.331 (0.000)	15.223 (0.000)	15.224 (0.000)
LTC	GAS	2.268 (0.132)	3.293 (0.193)	6.055 (0.014)	6.841 (0.033)
	DCC	24.624 (0.000)	25.070 (0.000)	8.204 (0.004)	8.221 (0.016)
BTC	GAS	16.057 (0.000)	16.068 (0.000)	3.826 (0.051)	4.428 (0.109)
	DCC	14.036 (0.000)	14.075 (0.000)	0.891 (0.345)	1.009 (0.604)
XRP	GAS	2.678 (0.102)	9.890 (0.007)	2.713 (0.099)	2.779 (0.249)
	DCC	24.624 (0.000)	25.611 (0.000)	15.012 (0.000)	17.141 (0.000)
Portfolio 1	GAS	2.268 (0.132)	2.563 (0.278)	1.349 (0.245)	1.766 (0.414)
	DCC	42.233 (0.000)	42.649 (0.000)	8.204 (0.004)	8.221 (0.016)
Portfolio 2	GAS	13.312 (0.000)	14.538 (0.000)	3.826 (0.050)	4.911 (0.086)
	DCC	28.401 (0.000)	29.060 (0.000)	5.901 (0.015)	5.930 (0.052)
Portfolio 3	GAS	1.891 (0.169)	2.123 (0.346)	2.878 (0.090)	4.137 (0.126)
	DCC	30.432 (0.000)	31.176 (0.000)	8.204 (0.004)	8.221 (0.016)
Portfolio 4	GAS	3.592 (0.058)	5.183 (0.075)	8.692 (0.003)	9.238 (0.010)
	DCC	28.400 (0.000)	29.060 (0.000)	11.189 (0.001)	11.196 (0.003)
Portfolio 5	GAS	7.745 (0.005)	9.179 (0.010)	1.658 (0.198)	1.748 (0.417)
	DCC	22.871 (0.000)	23.947 (0.000)	8.691 (0.003)	9.594 (0.008)

The one-step-ahead VaR forecasts for both individual ETH, LTC, BTC and XRP returns and arbitrary portfolios are found simultaneously based on simulated innovation. By ordering the cryptocurrencies as ETH, LTC, BTC and XRP, portfolios 1, 2, 3, 4 and 5 are constructed using weight vectors $g_1 = (1/4, 1/4, 1/4, 1/4)$, $g_2 = (1/4, 1/4, 1/4, -1/4)$, $g_3 = (1/4, 1/4, -1/4, 1/4)$, $g_4 = (1/4, -1/4, 1/4, 1/4)$ and $g_5 = (-1/4, 1/4, 1/4, 1/4)$, respectively. The column labelled UC and CC reports the unconditional coverage test of Kupiec (1995) and the conditional coverage test of Christoffersen (1998) with p values in the parentheses, respectively

for all portfolios in the forecasting experiment. The only exception is the portfolio with weights $g_2 = (1/4, 1/4, 1/4, -1/4)$ and $g_5 = (-1/4, 1/4, 1/4, 1/4)$ at the 5% significant level and the portfolio with weight $g_4 = (1/4, -1/4, 1/4, 1/4)$ at the 1% significant level, for which both model perform poorly.

In Figs. 15 and 16, we show the 1% and 5% VaR estimates against the realized returns for portfolio 1, i.e. the long-only portfolio with equal weights for the four cryptocurrencies. We observe that typically the VaR estimates based on the DCC models are more extreme, confirming that the DCC model significantly overestimates the risk at both 5% and 1% quantile levels, especially when the return changes are large (e.g. April 2020 and May 2021). These results are in accordance with previous findings (Creal et al. 2011). The estimates of the DCC model are based on lagged squared returns and the forecasts thus move stochastically every day. However, the updating equation in the GAS model with the Student-t density provides a more moderate increase in the variance/correlation for a large absolute realization of return.

The forecasts using the GAS model naturally inherit the return information. Overall, we conclude that the GAS model has better out-of-sample forecasting behavior.

4 Conclusion

We have investigated the co-dependence and portfolio VaR of cryptocurrencies using four popular virtual currencies (Bitcoin, Ethereum, Litecoin and Ripple). The results of the multivariate GAS model show strong dynamic interdependence among the cryptocurrencies throughout the sample period. Our out-of-sample forecasting period notably included the COVID-19 outbreak period, which lasted from early 2020 to the end of 2021. Thus, it sheds new light on the multivariate risk measures of cryptocurrencies for global investors.

We examine the out-of-sample predictive performance of the multivariate GAS model for a range of financial assets at various quantile levels. Using a battery of scoring rules and backtesting procedures, our results show that the GAS model significantly outperforms the traditional DCC-GARCH model. These results still hold if different cryptocurrencies are considered. There is plenty of room for future research on the analysis of cryptocurrencies, especially during financial turmoil. We can extend the existing scoring rules (especially in multivariate cases) to a more flexible form to cover a particular region of the density. An alternative extension could explore the safe-haven properties of cryptocurrencies, stablecoins and traditional assets. Under this framework, the dynamic correlations and the portfolio diversification can be studied systematically.

Declarations

Competing interest The author has no competing interests to declare that are relevant to the content of this article.

References

Amisano G, Giacomini R (2007) Comparing density forecasts via weighted likelihood ratio tests. J Bus Econ Stat 25:177–190

Andersen T, Bollerslev T, Diebold F, Labys P (2001) The distribution of realized exchange rate volatility. J Am Stat Assoc 96:42–55

Antonakakis N, Chatziantoniou I, Gabauer D (2019) Cryptocurrency market contagion: market uncertainty, market complexity, and dynamic portfolios. J Int Financ Markets, Inst Money 61:37–51

Apergis N (2021) COVID-19 and cryptocurrency volatility: evidence from asymmetric modelling. Financ Res Lett. https://doi.org/10.1016/j.frl.2021.102659

Ardia D, Boudt K, Catania L (2019) Generalized autoregressive score models in R: the GAS package. J Stat Softw 88(6):1–28

Aslanidis N, Bariviera AF, Martinez-Ibanez O (2019) An analysis of cryptocurrencies conditional cross correlations. Financ Res Lett 31:130–137

Avramov D, Chordia T, Goyal A (2006) The impact of trades on daily volatility. Rev Financ Stud 19:1241–1277

Blau B, Griffith T, Whitby R (2020) Comovement in the Cryptocurrency market. Econom Bull 40(1):448–455

Boako G, Tiwari AK, Roubaud D (2019) Vine copula-based dependence and portfolio value-at-risk analysis of the cryptocurrency market. Int Econom 158:77–90

Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. J Econom 31:307–327

Bouri E, Gupta R (2017) Does Bitcoin hedge global uncertainty? Evidence from wavelet-based quantile-in-quantile regression. Finance Res Lett 23:87–95

Bouri E, Gupta R, Lau CKM, Roubaud D, Wang S (2018) Bitcoin and global financial stress: a copula-based approach to dependence and causality in the quantiles. Q Rev Econ Finance 69:297–307

Cheikh NB, Zaied YB, Chevallier J (2020) Asymmetric volatility in cryptocurrency markets: new evidence from smooth transition GARCH models. Financ Res Lett 35:101293

Chen R, Xu J (2019) Forecasting volatility and correlation between oil and gold prices using a novel multivariate GAS model. Energy Econom 78:379–391

Chi Y, Hao W (2021) Volatility models for cryptocurrencies and applications in the options market. J Int Financial Mark Inst Money 75:101421

Christoffersen PF (1998) Evaluating Interval Forecasts. Int Econ Rev 39(4):841–862

Chowdhury MSR, Damianov DS, Elsayed AH (2022) Bubbles and crashes in cryptocurrencies: interdependence, contagion, or asset rotation? Financ Res Lett 46:102494

Conlon T, McGee R (2020) Safe Haven or Risky Hazard? Bitcoin during the COVID-19 Bear Market. Financ Res Lett 35:3560361

Corbet S, Lucey B, Yarovya L (2018) Datastamping the Bitcoin and Ethereum bubbles. Financ Res Lett 26:81–88

Creal DD, Koopman SJ, Lucas A (2011) A dynamic multivariate Heavy-Tailed model for time-varying volatilities and correlations. J Bus Econom Stat 29(4):552–563

Creal DD, Koopman SJ, Lucas A (2013) Generalized autoregressive score models with applications. J Appl Economet 28(5):777–795

Diebold FX, Mariano RS (1995) Comparing predictive accuracy. J Bus Econom Stat 13(3):253–263

Engle RF (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50(4):987–1008

Engle RF (2002) Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. J Bus Econom Stat 20(3):339–350

Engle RF, Russell JR (1998) Autoregressive conditional duration: a new model for irregularly spaced transaction data. Econometrica 66(5):1127–1162

Ghabri Y, Guesmi K, Zantour A (2021) Bitcoin and liquidity risk diversification. Financ Res Lett 40:101679

Giacomini R, White H (2006) Tests of conditional predictive ability. Econometrica 74:1545–1578

Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc 102(477):359–378

Good IJ (1952) Rational Decisions. J Royal Stat Soc B 14(1):107–114

Groen JJ, Paap R, Ravazzolo F (2013) Real-time inflation forecasting in a changing world. J Bus Econom Stat 31(1):29–44

Guesmi K, Saadi Samir, Abid I, Ftiti Z (2019) Portfolio diversification with virtual currency: Evidence from bitcoin. International Review of Financ Anal 63:431–437

Guesmi K, Saadi S, Abid I, Ftiti Z (2019) Portfolio diversification with virtual currency: evidence from bitcoin. Int Rev Financ Anal 63:431–437

Katsiampa P (2017) Volatility estimation for bitcoin: a comparison of GARCH models. Econ Lett 158:3–6

Katsiampa P (2019) An empirical investigation of volatility dynamics in the cryptocurrency market. Res Int Bus Financ 50:322–333

Katsiampa P, Corbet S, Lucey B (2019) High frequency volatility co-movements in cryptocurrency markets. J Int Financ Mark, Inst Money 62:35–52

Kupiec PH (1995) Techniques for verifying the accuracy of risk measurement models. J Deriv 3(2):73–84. https://doi.org/10.3905/jod.1995.407942

Liu L, Patton A, Sheppard K (2015) Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. J Econom 187(1):293–311

Mitchell J, Hall SG (2005) Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESR 'fan' charts of inflation. Oxford Bull Econ Stat 67:995–1033

Nadarajah S, Chu J (2017) On the inefficiency of Bitcoin. Econ Lett 150:6–9

Pace PD, Rao J (2023) Comovement and instability in cryptocurrency markets. Int Rev Econ Financ 8:173–200

Palmer T (2012) Towards the probabilistic earth-system simulator: a version for the future of climate and weather prediction. Q J R Meteorol Soc 138(665):841–861

Patton AJ (2011) Volatility forecast comparison using imperfect volatility proxies. J Econom 160(1):246–256

Pinson P, Tastu J (2013) Discrimination ability of the Energy score, Technical University of Denmark. DTU Compute-Technical Report-2013 No. 15

Scheuerer M, Hamill TM (2015) Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. Mon Weather Rev 143(4):1321–1334

Stavroyiannis S, Babalos V (2017). Dynamic Properties of the Bitcoin and the US Market. https://doi.org/10.2139/ssrn.2966998

Syuhada K, Hakim A (2020) Modeling risk dependence and portfolio VaR forecast through vine copula for cryptocurrencies. PLoS ONE 15(12):e0242102

Tafakori L, Pourkhanali A, Fard FA (2018) Forecasting spikes in electricity return innovations. Energy 150:508–526

Tang K, Xiong W (2012) Index investment and the financialization of commodities. Financ Anal J 68(6):54–74

Tran VL, Leirvik T (2020) Efficiency in the markets of crypto-currencies. Financ Res Lett 35:101382

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.