ARTICLE IN PRESS

International Journal of Forecasting xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

# Forecasting cryptocurrency volatility

Leopoldo Catania [a,b,*], Stefano Grassi [b,c]

[a] Aarhus University (Aarhus Universitet), Denmark
[b] CREATES, Denmark
[c] Universita degli Studi di Roma Tor Vergata, Dipartimento di Economia e Finanza, Italy

article info

*Keywords:*
Cryptocurrency
Bitcoin
Score-driven model
Density prediction
Volatility prediction
Leverage effect
Long memory
Higher-order moments

abstract

This paper studies the behavior of cryptocurrencies' financial time series, of which Bitcoin is the most prominent example. The dynamics of these series are quite complex, displaying extreme observations, asymmetries, and several nonlinear characteristics that are difficult to model and forecast. We develop a new dynamic model that is able to account for long memory and asymmetries in the volatility process, as well as for the presence of time-varying skewness and kurtosis. The empirical application, carried out on 606 cryptocurrencies, indicates that a robust filter for the volatility of cryptocurrencies is strongly required. Forecasting results show that the inclusion of time-varying skewness systematically improves volatility, density, and quantile predictions at different horizons.



## 1. Introduction

Bitcoin (BTC) is the first decentralized cryptocurrency. It was created in 2009 and documented in Nakamoto (2009). Since its inception, it has gained growing attention from the media, academics, and the finance industry. There are several reasons for this: Japan and South Korea have recognized BTC as a legal method of payment (Bloomberg, 2017a; Cointelegraph, 2017); some central banks are exploring the use of cryptocurrencies (Bloomberg, 2017b); and a large number of companies and banks created the Enterprise Ethereum Alliance[1] to make use of cryptocurrencies and the related technology, called blockchain Forbes (2017). Moreover, the Chicago Mercantile Exchange (CME) started to negotiate BTC futures on the 18th of December 2017 (Chicago Mercantile Exchange, 2017), and Facebook has launched a cryptocurrency, called Libra (The Economist, 2019). All this interest has been reflected in cryptocurrencies' market capitalization, which exploded from $19 billion in February 2017 to a peak of around $800 billion in December 2017. The market capitalization in February 2020 was around $280 billion, with around 5000 cryptocurrencies available, and BTC traded at around $9680.[2] Although BTC can be considered relatively new, there has already been some initial analysis of the cryptocurrency. Hencic and Gourieroux (2014) applied a non-causal autoregressive model to detect the presence of bubbles in the BTC/USD exchange rate. Sapuric and Kokkinaki (2014) measured the volatility of the BTC exchange rate against six major currencies. Bianchi (2020) analyzed whether cryptocurrencies can be considered a standard asset class, and Schilling and Uhlig (2019) studied the monetary policy implications of cryptocurrency.

From a statistical perspective, Chu et al. (2015) analyzed log returns of the exchange rate of BTC versus the USD and found that the generalized hyperbolic distribution seems to be the most appropriate choice to model the unconditional distribution of cryptocurrency time series. Núñez et al. (2019) show the ability of the normal inverse Gaussian distribution (NIG) to fit the returns of BTC.

* Corresponding author.
*E-mail address:* leopoldo.catania@econ.au.dk (L. Catania).

[1] Source: https://entethalliance.org/members/.

[2] Source: https://coinmarketcap.com/.

https://doi.org/10.1016/j.ijforecast.2021.06.005
0169-2070/

Please cite this article as: L. Catania and S. Grassi, Forecasting cryptocurrency volatility. International Journal of Forecasting (2021), https://doi.org/10.1016/j.ijforecast.2021.06.005.

A few studies have analyzed cryptocurrency predictability. For instance, Hotz-Behofsits et al. (2018) applied a time-varying parameter VAR with $t$-distributed measurement errors and stochastic volatility to model cryptocurrencies, Catania et al. (2018a) studied cryptocurrency
predictability using several alternative univariate and multivariate models. Griffin and Shams (2020) investigated whether Tether (another cryptocurrency backed by the USD) is directly manipulating the price of BTC, increasing its predictability; see also Gandal et al. (2018). Trucíos (2019) compared the predictability of the one-step-ahead volatility and value at risk (VaR) of BTC using several volatility models, showing that robust procedures outperform non-robust ones in forecasting the volatility and estimating the VaR. His results confirm previous findings by Catania et al. (2018b) and Charles and Darné (2019) where it was noted that a careful treatment of extreme observations for the analysis of cryptocurrency returns is key. Specifically, he found that a score-driven model with time-varying volatility and Student's $t$ distributed innovations outperforms a large number of GARCH-type models. We stress that while everyone acknowledges the presence of extreme observations in cryptocurrency returns, very few researchers incorporate this stylized fact in their modeling framework.

The contribution of the paper is twofold. First, we develop a dynamic updating equation for the volatility of a time series, exploiting the information contained in the score of the conditional distribution of the data, which is assumed to follow the generalized hyperbolic skewed Student's $t$ law (GHSKT); this distribution is detailed in Aas and Haff (2006) and obtained as a limiting case of the generalized hyperbolic family (Barndorff-Nielsen, 1977). The resulting model is compared with the simpler one with Student's $t$ shocks, which Trucíos (2019) finds particularly suited for predicting BTC volatility. This dynamic model can be thought of as a special case of the multivariate model introduced by Lucas et al. (2014) and Lucas et al. (2017) and belongs to the score-driven class of models recently introduced by Creal et al. (2013) and Harvey (2013).

Our contribution extends that of Lucas et al. (2014) and Lucas et al. (2017) in several directions. First, we derive expressions in order to apply the model in a univariate framework. These cannot be easily recovered from (Lucas et al., 2014). Second, we investigate the characteristics of the filter implied for the volatility update. In particular, we investigate the implications of the asymmetric tail behavior of the GHSKT distribution. Third, we extend the model to account for several features, such as the presence of leverage effects (Black, 1976), the long memory of the volatility process, and time-varying higher-order moments. Furthermore, in the supplementary material accompanying this paper, we report a Monte Carlo experiment to investigate the finite sample properties of the maximum likelihood estimator (MLE).

The second contribution of this paper is empirical. Specifically, we concentrate on the dynamic properties of cryptocurrencies' financial time series, rather than on their unconditional distribution, as in previous studies; see Chu et al. (2015). Moreover, we investigate and forecast a large set of cryptocurrencies, which to the best of our knowledge has not been done yet. We do so by reporting in-sample and out-of-sample aggregate results for the universe of cryptocurrencies, as well as for subgroups according to their market capitalization. Detailed results for the four major cryptocurrencies are also reported. Besides studying the volatility process of hundreds of cryptocurrencies, we also report a comparison of the proposed volatility specification to the well-known GARCH model of Bollerslev (1986). It is generally acknowledged that score-driven models report better volatility estimates than GARCH in the presence of extreme observations (see, e.g., Creal et al. 2013 and Bernardi and Catania 2016). However, many scientific articles that use GARCH-type models with the BTC series (which is notoriously characterized by a large number of extreme observations) have appeared.[3] This point was recently stressed by Alexander and Dakos (2020), who showed that the presence of extreme observations in cryptocurrency returns can lead to misleading results for risk assessment. Therefore, another contribution of our study is to warn researchers and practitioners regarding the use of non-robust volatility models (like GARCH) with cryptocurrencies, as these can lead to misleading results in financial applications like option pricing, portfolio optimization, and forecasting. In this sense, our results extend those of Trucíos (2019) in several directions: we find that the score-driven volatility filter arising from the GHSKT distributional assumption outperforms the one built according to the simpler Student's $t$ density; we consider 606 cryptocurrencies, and not only BTC; we report multi-step-ahead volatility predictions; together with volatility and VaR predictions, we also consider density predictions; and we investigate the contribution of including time-varying higher-order moments in the modeling framework.

Among other results, our large-scale analysis indicates that incorporating time-varying skewness significantly improves volatility, risk measures, and density predictions. The results are robust when considering different levels of market capitalization and forecast horizons. The paper proceeds as follows: Section 2 presents the basic model. Section 3 reports the extensions of the model to incorporate leverage effects, long memory in the volatility process, and time-varying higher-order moments. Section 4 provides the main empirical contributions of the paper. Section 5 draws some conclusions.

## 2. Model

Let $y_t$ be the realization of the random variable $Y_t \in \Re$ at time $t$. We assume:

$$Y_t = \varsigma + \exp(h_t)\varepsilon_t, \qquad \varepsilon_t \overset{iid}{\sim} \mathcal{GHSKT}(0, 1, \boldsymbol{\theta}), \tag{1}$$
$$h_t = \phi(h_{t-1}, y_{t-1}; \boldsymbol{\phi}),$$

[3] See, for example, Dyhrberg (2016), Chu et al. (2017), Katsiampa (2017), and Caporale and Zekokh (2019). We also acknowledge that investment banks and private investors have also started operating with cryptocurrencies; see, e.g., Fortune (2018). Furthermore, it is likely that many investors are using GARCH-type models with cryptocurrencies.

where $\varsigma$ is a constant, and $h_t$ is the log volatility obtained as a function of the past log volatility level, $h_{t-1}$, and the realization $y_{t-1}$, via the known function $\phi : \Re \times \Re \to \Re^+$. The function $\phi$ constitutes the filter for the volatility and characterizes the model specification. For example, when $\phi(h_{t-1}, y_{t-1}, \boldsymbol{\psi}) = (1/2)\log(\phi_0 + \phi_1(y_{t-1} - \varsigma)^2 + \phi_2 \exp(2h_{t-1}))$, where $\boldsymbol{\psi} = (\phi_0, \phi_1, \phi_2, \varsigma)'$, we obtain the popular GARCH specification of Bollerslev (1986). To conclude the model specification, the only source of error affecting the dynamic system, $\varepsilon_t$, is assumed to be distributed independently as a standardized GHSKT. We briefly detail how to construct this distribution in the next subsection, and for the sake of notation we drop the time index $t$.

### 2.1. Distributional assumption

The GHSKT distribution detailed in Aas and Haff (2006) has the following probability density function (PDF):

$$p(y; \boldsymbol{\theta}^*) = \frac{2^{(1-\nu)/2}\delta^\nu |\beta|^{(\nu+1)/2} K_{(\nu+1)/2}\left(\sqrt{\beta^2(\delta^2 + (y-\mu)^2)}\right)\exp(\beta(y-\mu))}{\Gamma(\nu/2)\sqrt{\pi}(\sqrt{\delta^2 + (y-\mu)^2})^{(\nu+1)/2}}, \quad \beta \neq 0, \tag{2}$$

and

$$p(y; \boldsymbol{\theta}^*) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi}\delta\Gamma(\frac{\nu}{2})}\left[1 + \frac{(y-\mu)^2}{\delta^2}\right]^{-(\nu+1)/2}, \quad \beta = 0, \tag{3}$$

where $K_a(b)$ is a modified Bessel function of the third kind (Abramowitz & Stegun, 1964) evaluated in $b$ with (fractional) order $a$, and $\boldsymbol{\theta}^* = (\mu, \delta, \beta, \nu)'$ contains the location $\mu \in \Re$, scale $\delta > 0$, and the two shape parameters $\beta \in \Re$ and $\nu > 0$. When $\beta = 0$, GHSKT reduces to Student's $t$ distribution with density of Eq. (3), and to a Gaussian distribution if also $\nu = \infty$. The dynamic model for these two special cases is detailed in Creal et al. (2013) and Harvey (2013), to which we refer the reader for further details. For the rest of the paper, we analyze the case where $\beta \neq 0$.

From Eq. (1), exploiting the $\boldsymbol{\theta}$ parametrization of the GHSKT distribution reported in Appendix A, it follows that:

$$Y_t|\mathbf{y}_1^{t-1} \sim \mathcal{GHSKT}(\varsigma, \sigma_t, \bar{\beta}, \nu), \tag{4}$$

where $\sigma_t = \exp(h_t)$ is the volatility at time $t$, and $\mathbf{y}_1^{t-1} = (y_1, \ldots, y_{t-1})'$ denotes the history of the process up to time $t-1$; see Theorem 1(c) of Blæsild (1981). According to Eq. (4), the two moment conditions $\mathbb{E}_{t-1}(Y_t) = \varsigma$ and $\mathrm{Var}_{t-1}(Y_t) = \sigma_t^2$ hold.[4] The score with respect to $\varsigma$ and $\sigma_t$ is easily derived as follows:

$$\begin{aligned}\frac{\partial \log p(y_t; \varsigma, \sigma_t, \bar{\beta}, \nu)}{\partial \varsigma} &= -\frac{1}{\sigma_t}\nabla_\varepsilon(\varepsilon_t; \varsigma, \sigma_t, \bar{\beta}, \nu),\\ \frac{\partial \log p(y_t; \varsigma, \sigma_t, \bar{\beta}, \nu)}{\partial \sigma_t} &= -\frac{1}{\sigma_t}(1 + \varepsilon_t\nabla_\varepsilon(\varepsilon_t; \varsigma, \sigma_t, \bar{\beta}, \nu)),\end{aligned} \tag{5}$$

where $\nabla_\varepsilon(\varepsilon_t; \varsigma, \sigma_t, \bar{\beta}, \nu) = \partial \log p(\varepsilon_t; 0, 1, \bar{\beta}, \nu)/\partial\varepsilon_t$ is the derivative of the log density of a standardized GHSKT distribution with respect to $\varepsilon_t$, given by:

$$\nabla_\varepsilon(\varepsilon_t; \varsigma, \sigma_t, \bar{\beta}, \nu) = \beta\,(1 + \beta(\varepsilon_t - \mu)\zeta) - \frac{(\nu+1)(\varepsilon_t - \mu)}{2(\delta^2 + (\varepsilon_t - \mu)^2)}, \tag{6}$$

where $\zeta$ is given in Eq. (A.8) with $z = \varepsilon_t$, and $\beta$, $\mu$, and $\delta$ are computed from $\bar{\beta}$ according to the mapping $(\bar{\beta}, \nu) \to (\mu, \delta, \beta, \nu)$, as discussed in Appendix A.

### 2.2. Filter $\phi$

Let $u_t^h = \partial \log p(y_t; \varsigma, \sigma_t, \bar{\beta}, \nu)/\partial h_t = (\partial h_t/\partial\sigma_t)(\partial \log p(y_t; \varsigma, \sigma_t, \bar{\beta}, \nu)/\partial\sigma_t)$ be the score of the conditional distribution $(Y_t|\mathbf{y}_1^{t-1})$, with respect to the log volatility $h_t$. Following Creal et al. (2013) and Harvey (2013), we build the following score-driven filter for the log volatility level $h_t$:

$$h_t = \kappa^h(1 - \beta^h) + \alpha^h u_{t-1}^h + \beta^h h_{t-1}, \tag{7}$$

where $\kappa = \mathbb{E}[h_t]$ (since $u_t$ is a Martingale difference with respect to $\sigma(Y_{t-1}, Y_{t-2}, \ldots)$, the filtration generated by the process $\{Y_t\}$) is the unconditional log volatility level, $|\beta^h| < 1$ is an autoregressive parameter controlling the persistence of the log volatility process, and $\alpha^h > 0$ is the parameter associated to the score which determines the step of the log volatility update. The positiveness of $\alpha^h$ is assumed, in order to avoid distorting the signal given by the score; see Creal et al. (2013).[5]

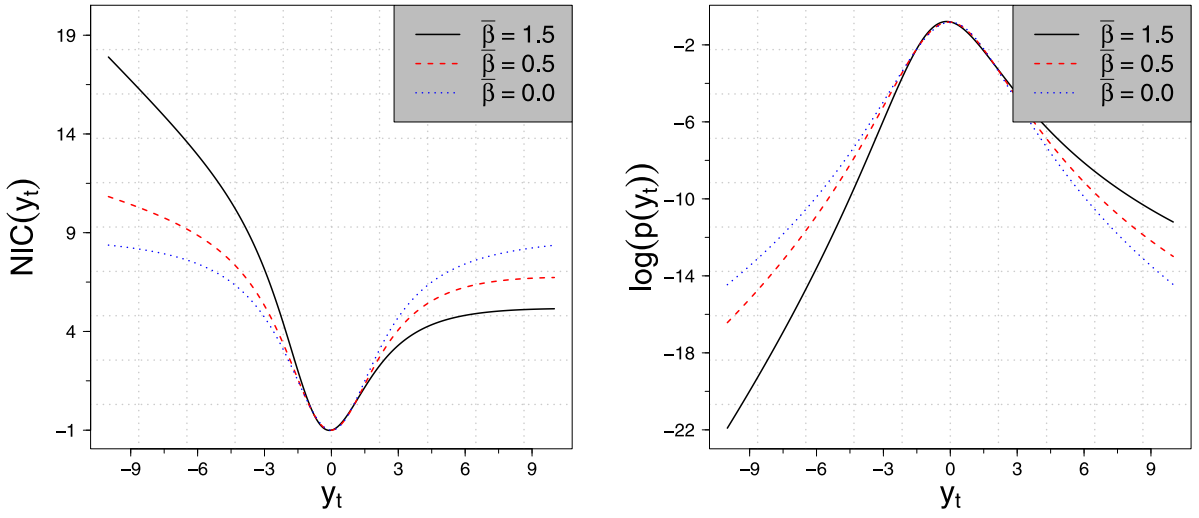
### 2.3. Volatility reaction to past returns

Engle and Ng (1993) introduced an important tool for time-varying volatility models, called a news impact curve (NIC). An NIC gives insight into the reaction of future volatility levels to past observations via a graphical representation. Formally, in Eq. (7) we set $h_t$ and $h_{t-1}$ to the unconditional level $\kappa^h$, and plot the function $\mathrm{NIC}(y_t) = \kappa^h + \alpha^h u_t^h$ for different values of $y_t$. Fig. 1(a) depicts the NIC for $\kappa^h = 0$, $\alpha^h = 1$ and different values of the shape parameter $\bar{\beta}$. Fig. 1(b) reports the associated log densities. As noted by Harvey and Sucarrat (2014) and Harvey and Lange (2017), the volatility reaction induced by score-driven models defined with a skewed conditional density is asymmetric and can represent the leverage effect phenomenon for selected values of the skewness parameter.

However, even if the volatility reaction is always asymmetric for $\bar{\beta} \neq 0$, the direction of the asymmetry depends on the sign of $\bar{\beta}$. Indeed, for $\bar{\beta} > 0$, the density is right

[4] We indicate by $\mathbb{E}_{t-1}[Y_t]$ the expectation $\mathbb{E}[Y_t|\mathbf{y}_1^{t-1}]$.

[5] In the context of score-driven models, the forcing variable $u_t$ is sometimes scaled by a positive quantity that accounts for its variance. For instance, $(\mathbb{E}[u_t^2])^{-1/2}u_t$ would imply unit variance of the resulting forcing variable of the filter. Unfortunately, however, the variance of the score, $\mathbb{E}[u_t^2]$, is not available in closed form for the GHSK distribution. In a related context, Lucas et al. (2017) replaced $\mathbb{E}[u_t^2]$ with the information matrix of Student's $t$ distribution instead of that of a GHSKT distribution. We tried with a similar approach and found that the results were qualitatively the same in our application on cryptocurrency time series.




**(a)** NIC **(b)** Log PDF

**Fig. 1.** Panel (a) reports the NIC for the volatility of the score-driven GHSKT model when $\alpha^h = 1$ and $\kappa^h = 0$ for selected values of $\bar{\beta}$ and $\zeta = 0$, $\sigma = 1$, $\nu = 7$. Panel (b) reports the associated log probability density functions.

skewed and we observe the usual leverage effect; i.e., past negative shocks increase the future volatility more than past positive shocks. By contrast, when $\bar{\beta} < 0$, the density is left skewed and the volatility reaction is opposed to the previous case. This behavior is explained by noting that when $\bar{\beta} > 0$, it is less likely to sample negative observations. Hence, when this event occurs, it is natural to have a stronger signal of increasing volatility compared to the realization of a positive observation, which is indeed more likely. The reverse holds. From Fig. 1(a) we also observe the winsorization property of the score-driven filter for finite values of $\nu$; see Harvey (2013). However, unlike the filter implied by the Student's $t$ score-driven model ($\beta = 0$ in Fig. 1(a)), in our case we observe that winsorization occurs only for positive or negative observations according to the direction of the skewness. Specifically, we have that:

$$\begin{aligned} \lim_{y\to\infty} u_t &= \begin{cases} \infty, & \text{if } \bar{\beta} < 0 \\ \text{const}, & \text{if } \bar{\beta} \geq 0., \end{cases} \\ \lim_{y\to-\infty} u_t &= \begin{cases} \text{const}, & \text{if } \bar{\beta} \leq 0 \\ \infty, & \text{if } \bar{\beta} > 0, \end{cases} \end{aligned} \tag{8}$$

which follows from the tail behavior of the GHSKT; see Aas and Haff (2006).

## 3. Model extensions

In this section we propose different extensions to the model presented in Section 2. Specifically, we aim to incorporate long memory in the volatility process, leverage effects, and time-varying higher-order moments. These extensions are then used in our empirical analysis. Given the novelty of cryptocurrency time series in the financial econometric literature, we believe that the study of well-known volatility features (long memory and leverage effects) as well as conditional density characteristics (higher-order moments) is worth doing.

### 3.1. Long memory

The long memory of the volatility process is a very well-known feature of financial returns; see Andersen et al. (2001). The time-series literature developed three main methodologies to incorporate this feature in dynamic models: (i) fractional integration (Palma, 2007), (ii) multiple component models (Andersen et al., 2006), and (iii) via the so-called heterogeneous autoregressive framework introduced by Corsi (2009). Fractional integrated dynamic volatility models were first considered by Baillie et al. (1996), and recently extended by Janus et al. (2014, 2018) to the score-driven class of models. The first dynamic model with multiple volatility components was proposed by Engle and Lee (1999). Recently, Harvey and Sucarrat (2014) proposed a multiple component score-driven volatility model.[6] The heterogeneous autoregressive (HAR) model of Corsi (2009) was specifically developed to model realized volatility time series, even if its applicability clearly includes a wider range of time series.

From an unreported analysis, we found that the multiple component approach, when implemented with the model described in Section 2, reports better in-sample results than the two alternative aforementioned approaches. The inclusion of long memory via multiple components is straightforward. Essentially, we rewrite Eq. (7) as:

$$h_t = \kappa^h + h_{t,1} + h_{t,2}, \qquad h_{t,l} = \beta_l^h h_{t-1,l} + \alpha_l^h u_{t-1}^h, \quad l = 1, 2, \tag{9}$$

where $h_{1,l} = 0$ and $\beta_2^h > \beta_1^h$ are imposed for identifiability purposes. Namely, the log volatility process

[6] See also Alizadeh et al. (2002) for an application of this methodology in a stochastic volatility framework.

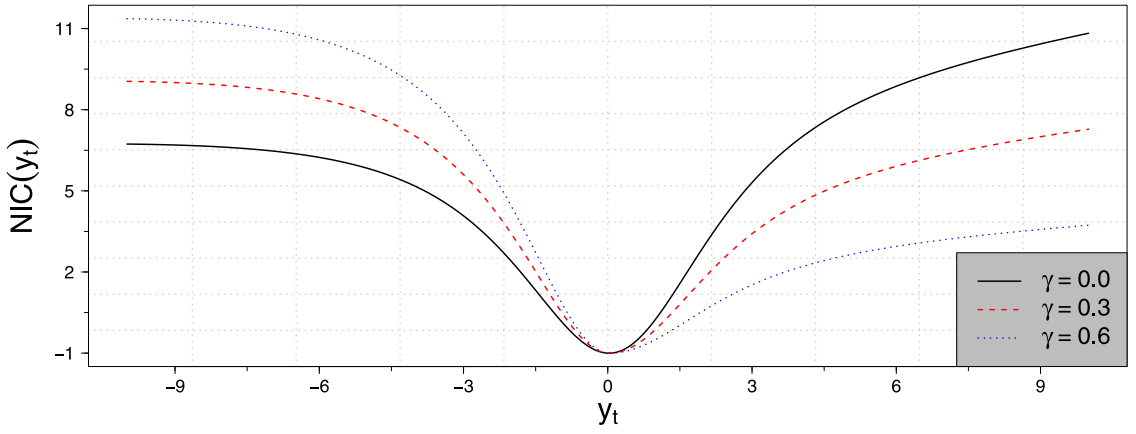



**Fig. 2.** NIC with the inclusion of the leverage-effect component for different values of $\gamma$: solid (black) line, $\gamma = 0$; dashed (red) line, $\gamma = 0.3$; dotted (blue) line, $\gamma = 0.6$. The other parameters are set as follows: $\alpha^h = 1$, $\kappa^h = 0$, $\bar{\beta} = -0.5$, and $\nu = 7.0$.

is now defined as a sum of two autoregressive processes with different autoregressive parameters and the same forcing variable $u_t^h$. The process $\{h_t, t > 0\}$ is still weakly stationary—provided that $\mathbb{E}[(u_t^h)^2] < \infty$—as long as $|\beta_l^h| < 1$ for $l = 1, 2$.

### 3.2. Leverage effects

Leverage effects refer to the asymmetric reaction of the volatility process to past positive and negative returns, a stylized fact which dates back to Crane (1959). This phenomenon was first investigated by Black (1976) and Christie (1982) and associated with the capital structure of firms in the spirit of the Modigliani and Miller (1958) framework. However, nowadays it is more common to attribute the asymmetric reaction of the volatility to different sources, such as the different reactions of investors to past positive and negative news.

Section 2.3, above, detailed how the inclusion of a skewed conditional density in score-driven volatility models naturally induces an asymmetric NIC. However, this kind of asymmetry is principally related to the parametric formulation of the model, i.e., how likely it is to sample from a particular region of the support. By contrast, leverage effects are usually related to the behavior of market participants. A score-driven volatility model with the inclusion of leverage effects was first introduced by Harvey (2013). Following Harvey's implementation, we rewrite Eq. (7) as:

$$h_t = \kappa^h(1 - \beta^h - \gamma^h\varpi) + \alpha^h u_{t-1}^h + \beta^h h_{t-1} + \gamma^h \text{sgn}(-(y_{t-1} - \varsigma))(u_{t-1} + 1), \quad (10)$$

where $\gamma^h$ is an additional parameter and $\text{sgn}(\cdot)$ is the sign function.[7] When $\gamma^h > 0$, we observe a higher increase of the volatility after negative (excess) returns. If $\gamma^h < \alpha^h$, we observe a decrease in volatility after positive (excess) returns. That is, the filter (10) can represent those situations when the volatility decreases after capital gains from investors. Note that the inclusion of the constant $\gamma^h\varpi$ with $\varpi = 2P(\varepsilon_t < 0) - 1$ is made to preserve the moment condition $\mathbb{E}[h_t] = \kappa^h$. Fig. 2 shows the implication of the leverage effect to the NIC for different values of $\gamma = \{0, 0.3, 0.6\}$. We set $\alpha^h = 1$, $\kappa^h = 0$, $\bar{\beta} = -0.5$, and $\nu = 7.0$. Since $\bar{\beta} < 0$, the conditional density is negatively skewed, and volatility reacts more to positive than negative past returns. This behavior is represented in Fig. 2 for $\gamma = 0$. However, for values of $\gamma > 0$, the direction of the asymmetry is allowed to change. Indeed, in Fig. 2, when $\gamma = 0.3$ or $\gamma = 0.6$, we recover the usual leverage effect. That is, the volatility increases more after negative returns. This means that, independently from the sign of $\bar{\beta}$, the filter reported in Eq. (10) is able to incorporate any kind of asymmetry in the volatility update via the inclusion of the additional parameter $\gamma^h$.

### 3.3. Time-varying higher-order moments

The inclusion of time-varying higher-order moments for financial time series dates back to the autoregressive conditional density model of Hansen (1994). Extensions include the independent factor autoregressive conditional density model of Ghalanos et al. (2015) and the autoregressive conditional kurtosis model of Brooks et al. (2005), among others. With regards to score-driven models, the inclusion of time-varying higher-order moments for financial time series was investigated by Lucas and Zhang (2016) and Bernardi and Catania (2016) for univariate and multivariate frameworks, respectively.

Before the development of score-driven models, the modeling of higher-order moments was performed via ad hoc procedures. The usual methodology was to include arbitrary transformations of past returns (say, $y_{t-1}$, $|y_{t-1}|$, $y_{t-1}^d$, for $d = 2, 3, 4$) and then see if these had any sort of predictive power for the shape parameters of the conditional distribution assumed for the returns. This strategy facilitates the tractability of the model, even if it is clearly open to a wide range of criticism. Indeed, a coherent statistical answer to the choice of such transformations was missing until the development of the score-driven framework.

[7] The use of $(u_{t-1} + 1)$ instead of just $u_{t-1}$ is due to the fact that omitting the $+1$ term would imply a jump of the NIC at $\varsigma$.

The score-driven methodology naturally suggests performing filtering using the score of the conditional distribution assumed for the data. Following this general approach, any parameter in a dynamic model, for which a parametric distributional assumption is available, can be time varying. Specifically, the quantities $u_t^{\bar{\beta}} = \partial \log p(y_t;\boldsymbol{\theta})/\partial\bar{\beta}$ and $u_t^{\nu} = \partial \log p(y_t;\boldsymbol{\theta})/\partial\nu$ collected in the vector $\nabla_{\boldsymbol{\theta}}(y_t;\boldsymbol{\theta}_t)$ reported in Eq. (A.9) can be used to build a filter for $\bar{\beta}_t$ and $\nu_t$. For the volatility parameter, we can write:

$$\bar{\beta}_t = \kappa^{\bar{\beta}}(1-\beta^{\bar{\beta}}) + \alpha^{\bar{\beta}} u_{t-1}^{\bar{\beta}} + \beta^{\bar{\beta}}\bar{\beta}_{t-1}, \tag{11}$$

for the time-varying skewness parameter, $\bar{\beta}_t$, and $\nu_t = \exp(\widetilde{\nu}_t) + 4$, where

$$\widetilde{\nu}_t = \kappa^{\nu}(1-\beta^{\nu}) + \alpha^{\nu} u_{t-1}^{\nu}\nu_{t-1} + \beta^{\nu}\widetilde{\nu}_{t-1}, \tag{12}$$

for the time-varying shape parameter, $\widetilde{\nu}_t$. Note that we have employed an exponential link function with a lower bound of 4 in order to ensure the positiveness of $\nu_t$ and the existence of the variance of the conditional distribution $Y_t|\mathbf{y}_{1:t-1}$. It is also worth noting that, in the filter (12), we have multiplied the conditional score $u_{t-1}^{\nu}$ by $\nu_t$, which is the derivative of the mapping $\widetilde{\nu}_t \to \nu_t$, such that the forcing variable $u_t^{\nu}\nu_t$ is equivalent to $\partial \log p(y_t;\boldsymbol{\theta})/\partial\widetilde{\nu}_t$.

## 4. Empirical application

The data used in this study are cryptocurrency closing prices. It is important to underline that the crypto-market is open 24 h a day, seven days a week; hence for the closing price, we use prices at midnight (UTC). Data are freely available from CoinMarketCap at https://coinmarketcap.com/. CoinMarketCap provides historical data for most of the cryptocurrencies by aggregating the prices from different exchanges. As pointed out by Alexander and Dakos (2020), historical data from CoinMarketCap are affected by artificial volume figures, which are deliberately inflated by many exchanges to increase their profit. As a result, we stress that our analysis may be sensitive to the dataset used, and specifically to the scheme CoinMarketCap uses to construct the series.

Since the introduction of BTC in 2009, hundreds of other cryptocurrencies have been created and, as of February 2020, 5000 cryptocurrencies exist. Some of them, though, do not represent serious attempts at establishing a foothold in the market. Others are quite young with just a few useful observations. For this reason, we focus on coins with at least 700 observations, which correspond to almost two years of data; this leaves us with 606 series to analyze. We group each series into three categories based on the market capitalization recorded during the month of September 2019: (i) Large Cap (top 20%), (ii) Medium Cap (between 40% and 60%), and (iii) Small Cap (lower 20%).[8] Hence, we report results at the aggregate level (All), for the three categories (Large, Medium, and Small), and the four major cryptocurrencies: (i) BTC, (ii) Ethereum (ETH), (iii) Ripple (XRP), and (iv) Litecoin (LTC).

[8] Note that, series in the ranges (20% − 40%) and (60% − 80%) are not included in the Small, Medium, and Large categories in order to increase the clusters' separability.

### 4.1. Data and unconditional distribution of cryptocurrency financial time series

Table 1 reports descriptive statistics for the four main cryptocurrencies: BTC, ETH, XRP, and LTC. The table also reports the median value of the statistics computed over all cryptocurrencies, as well as for the three subgroups: Large Cap, Medium Cap, and Small Cap. The last column reports the same statistics for the Standard and Poor 500 index (S&P 500) and provides strong evidence of the large differences between cryptocurrencies and equity series. For example, the largest positive and negative returns observed in the S&P 500 series are 4.84% and −4.18%, while for all cryptocurrencies the median values are 140.36% and −118.46%, respectively, clearly showing the extreme behavior of those series.

Among the four main currencies, we note that ETH experienced the most extreme behavior, displaying a kurtosis coefficient of 68.61 and a negative skewness coefficient of −3.39. Other currencies also report descriptive statistics which might be considered extreme if compared with the S&P500 series. We also note that the cryptocurrencies are heterogeneous with respect to their volatility level. Indeed, we find that volatility decreases with the increase of market capitalization.

In Fig. 3 we report the median, standard deviation, skewness, and kurtosis for all the cryptocurrencies and for the Large, Medium, and Small groups. As the figure shows, the median (Panel a), skewness (Panel c), and kurtosis (Panel d) are very similar between the Large, Medium, and Small groups. The kurtosis, concentrated around 15, is quite high with extreme values near 100. The majority of the series show a positive skewness coefficient, concentrated around 1.7, with the presence of extreme values in the positive and negative region of the order −6 to +8. Finally, the standard deviations (Panel b) are heterogeneous across groups, with the biggest values for the Large group and relatively moderate values for the Small group. Interestingly, cryptocurrency financial time series also exhibit serial autocorrelation, as documented by the empirical autocorrelation coefficients (not reported). We believe this phenomenon requires an extended analysis and decided to leave it for future research. Consequently, we pre-filter the series using a simple autoregressive filter of order $p$ estimated by OLS, with $p$ chosen according to the Akaike information criterion.[9]

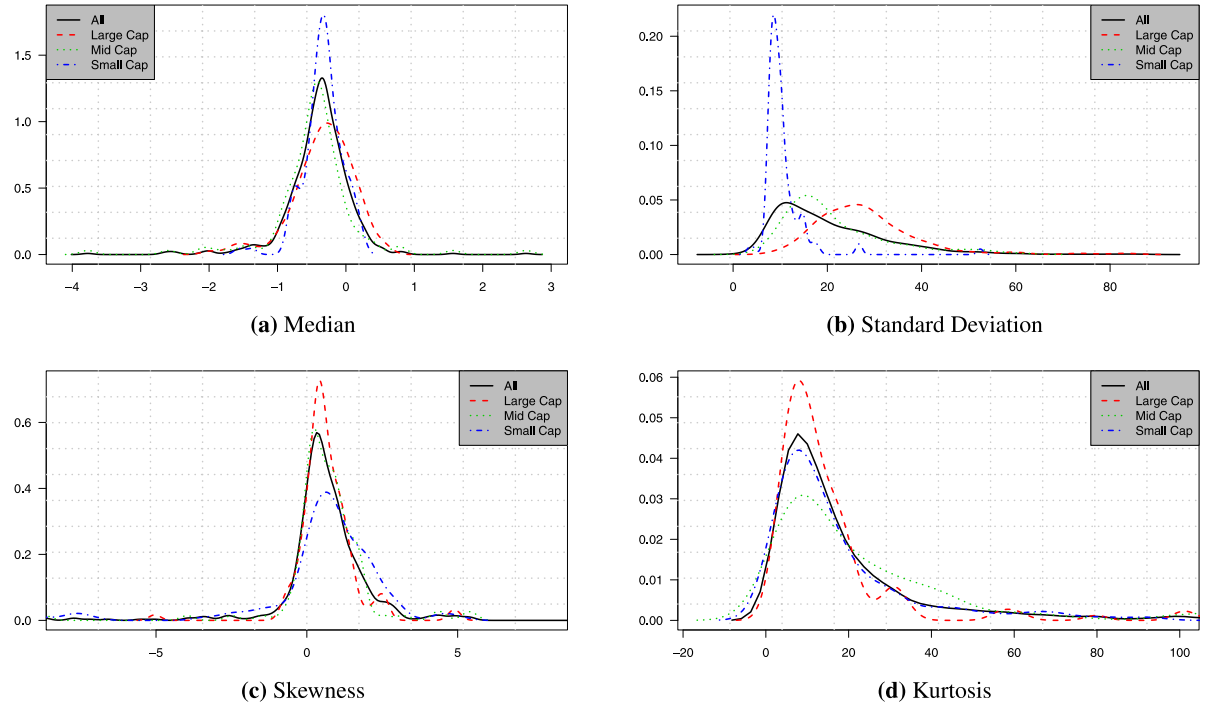
### 4.2. Importance of a robust volatility filter for cryptocurrency time series

We begin by studying the evolution of the volatility process of cryptocurrency time series and by emphasizing the importance of a robust filter for the volatility update. To this end, we compare the score-driven GHSKT

[9] OLS estimates are affected by extreme observations; see Huber (2004). In the case of cryptocurrencies, extreme observations are likely to arise as a result of large speculative trades that do not propagate over time. Then, extreme observations might be considered as additive outliers, implying that AR coefficients estimated by OLS are likely to be downward biased; see Huber (2004).

**Table 1**
The table reports the four larger cryptocurrencies by market capitalization: Bitcoin (BTC), Ethereum (ETH), Ripple (XRP), and Litecoin (LTC). For each cryptocurrency the table reports the number of observations, the creation date, the maximum number of coins in millions (Mil.) and billions (Bil.), the market capitalization as in September 2019 reported at https://coinmarketcap.com/, and descriptive statistics. Ethereum has a total supply of 18 million coins per year. The other three have a total prefix amount.

| Coin | BTC | ETH | XRP | LTC | All | Small | Medium | Large | S&P 500 |
|---|---|---|---|---|---|---|---|---|---|
| Created | 03-Jan-09 | 01-Aug-14 | 01-Jul-13 | 01-Nov-13 | – | – | – | – | – |
| Series begin | 30-May-13 | 13-Aug-15 | 17-Aug-13 | 07-May-13 | – | – | – | – | 04-March-14 |
| No. Obs. | 2292 | 1484 | 2211 | 2239 | – | – | – | – | 1393 |
| Supply | 21 Mil. Total | 18 Mil. Yearly | 100 Bil. Total | 84 Mil. | – | – | – | – | – |
| Market Cap. | 182 Billion | 19 Billion | 11 Billion | 4 Billion | – | – | – | – | – |
| Max. | 35.75 | 41.23 | 102.74 | 82.90 | 140.36 | 191.21 | 155.67 | 75.11 | 4.84 |
| Min. | −26.62 | −130.21 | −61.63 | −51.39 | −118.46 | −159.36 | −145.98 | −51.27 | −4.18 |
| Mean | 0.19 | 0.28 | 0.17 | 0.12 | −0.10 | −0.18 | −0.11 | −0.01 | 0.04 |
| Median | 0.20 | −0.09 | −0.30 | −0.25 | −0.36 | −0.30 | −0.42 | −0.32 | 0.05 |
| Std. Dev. | 4.33 | 7.35 | 7.42 | 6.66 | 17.26 | 26.51 | 18.14 | 9.39 | 0.84 |
| Skewness | −0.17 | −3.39 | 2.05 | 1.68 | 0.55 | 0.51 | 0.63 | 0.81 | −0.50 |
| Kurtosis | 7.58 | 68.61 | 28.84 | 24.07 | 12.40 | 10.75 | 14.92 | 11.69 | 3.67 |




**(a)** Median **(b)** Standard Deviation

**(c)** Skewness **(d)** Kurtosis

**Fig. 3.** Kernel densities of the median (a), standard deviation (b), skewness (c), and kurtosis (d) coefficients of the full sample of cryptocurrency time series. Kernel densities associated to the cryptocurrency groups are also reported: (i) Large (red dashed line), (ii) Medium (green dotted line), and (iii) Small (blue dot-dashed line).

model with a GARCH model with the same conditional distributional assumption, defined as

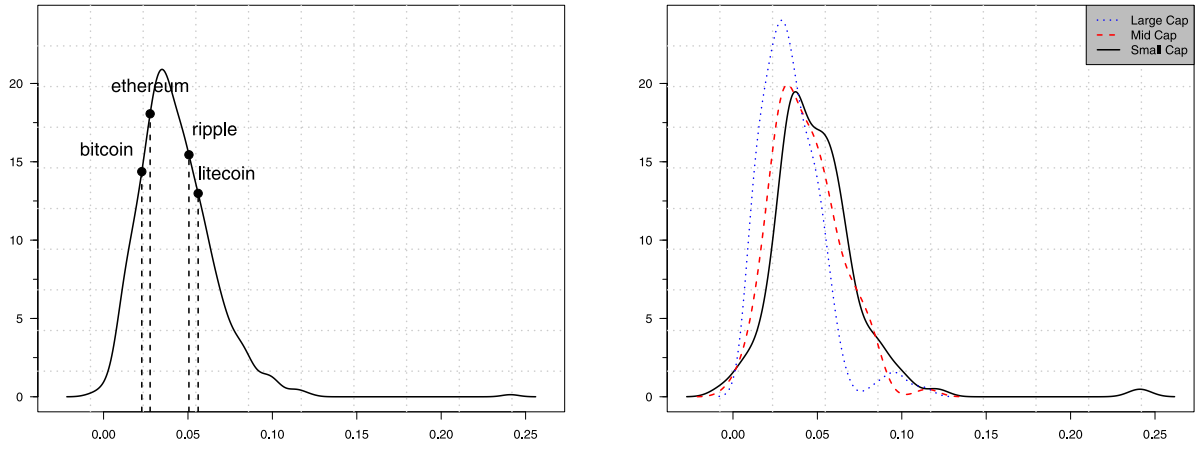
$$Y_t = \varsigma + \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \mathcal{GHSKT}(0, 1, \bar{\beta}, \nu), \tag{13}$$

$$\sigma_t^2 = \zeta_0 + \zeta_1 (y_{t-1} - \varsigma)^2 + \zeta_2 \sigma_{t-1}^2, \tag{14}$$

where $\zeta_i > 0$ for $i = 0, 1, 2$ and $\zeta_1 + \zeta_2 < 1$; see Bollerslev (1986). It is worth noting that the NIC implied by the filter in Eq. (13) is quadratic in the residual $y_t - \varsigma$, and hence different from the one implied by the score-driven filter; it is not robust to extreme observations. Fig. 4 reports a Gaussian kernel fitted on the likelihood percentage increases of the score-driven model over GARCH across the 606 cryptocurrencies.[10]

Panel (a) reports the percentage increase for the whole cross-section, while Panel (b) reports that for the three

[10] Note that the number of parameters for the GARCH and score-driven models is the same, such that in-sample model comparison can be performed by a simple likelihood comparison. We use a Gaussian kernel density estimator with bandwidth selected by Silverman's rule of thumb (see Silverman, 1986, p. 48). Alternative kernels like Epanechnikov and rectangular led to similar results.




**(a)** All **(b)** Subgroups

**Fig. 4.** Kernel density estimated on the likelihood percentage increase between the GHSKT score-driven model and the GARCH–GHSKT model. Positive values indicate outperformance of the GHSKT score-driven model with respect to the GARCH–GHSKT alternative. Panel (a) reports the density for all cryptocurrencies (All), and Panel (b) for the three groups of cryptocurrencies.

groups. We find that, on average, the GHSKT score-driven model reports a likelihood increase in the order of 5%. Furthermore, the GARCH always reports a lower likelihood than the score-driven specification. Across the four major cryptocurrencies, LTC is the one which benefits most from the score-driven filter, with a gain of 8.5%, followed by XRP (6.6%), ETH (4.5%), and BTC (2.8%). Looking at Fig. 4(a), we note that the likelihood increments vary with the size of the market capitalization. Specifically, likelihood increments are on average 7.2% for Small, 5.6% for Medium, and 5.4% for Large cryptocurrencies. Interestingly, we find that for Small and Medium cryptocurrencies the use of a robust volatility filter is more important than for Large Cap. Hence, for those series the robustness of the filter is an essential requirement.

To show the difference implied by the robust score-driven volatility filter, in Fig. 5 we report the filtered variances for a sub-sample of data delivered by the two models for Bitcoin (Panel (a)) and Ripple (Panel (b)). The lower panel of each figure shows the log-return series. The plot is very clear; whenever there is an extreme observation, the GARCH filter reports a huge increase in the volatility level, even if the signal of increased volatility is fairly weak. For instance, in Fig. 5(b), we see that on the 15th of September 2016, XRP exhibited a log return of roughly 30%, due to received funding of USD $55 million from private banks.[11] The reaction of GARCH was that of increasing the conditional variance from 5.8 to 408.9, while the increase delivered by the score-driven specification was from 2.8 to 15.6. Indeed, after that extreme observation, the overall volatility level did not change much, indicating that the huge increase in volatility reported by GARCH is not justified by the data.

For less capitalized cryptocurrencies, GARCH results deteriorate substantially. For this reason, we decided not to consider the GARCH specification (or other non-robust volatility models) in our empirical analysis. However, we note that a large number of articles report volatility estimates for cryptocurrencies based on the non-robust GARCH model. Our suggestion is to be careful with the interpretation of those results, especially when predictions are considered.

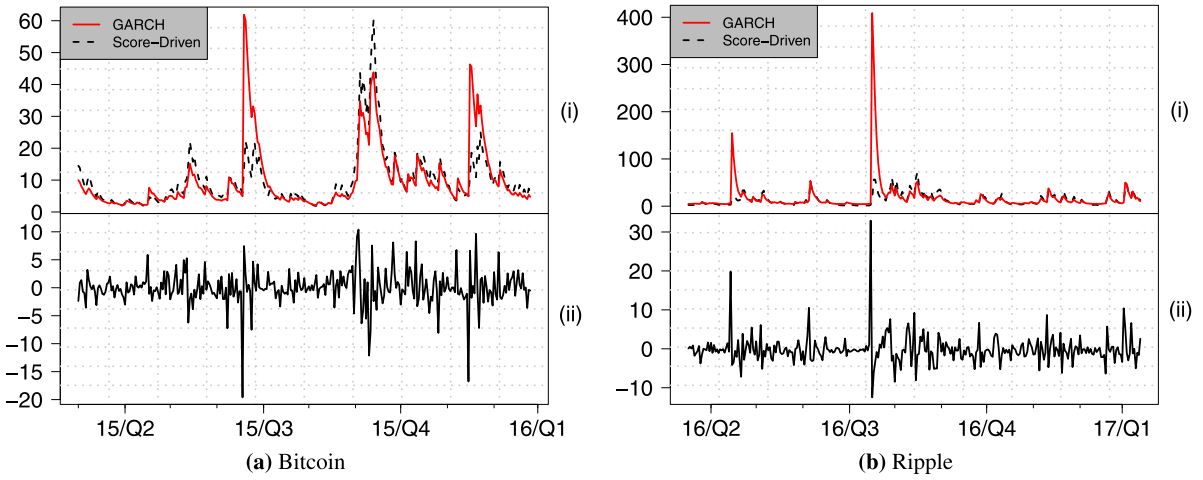
[11] Source: http://www.cnbc.com/2016/09/15/google-backed-blockchain-start-up-ripple-raises-55-million-from-big-banks.html.

**Table 2**
The table reports all the models considered, where $\mathcal{M}_0$ is the benchmark model in the forecasting exercise. The first column is the label of the model. The second column provides a brief description of each model.

| Label | Description |
|---|---|
| $\mathcal{M}_0$ | The Beta-Skew-$t$-EGARCH model with a leverage component. |
| $\mathcal{M}_1$ | The score-driven GHSKT model of Eq. (1) |
| $\mathcal{M}_2$ | $\mathcal{M}_1$ with leverage, as in Eq. (10) |
| $\mathcal{M}_3$ | $\mathcal{M}_1$ with two volatility components, as in Eq. (9) |
| $\mathcal{M}_4$ | $\mathcal{M}_1$ with time-varying $\bar{\beta}_t$, as in Eq. (11) |
| $\mathcal{M}_5$ | $\mathcal{M}_1$ with time-varying $\nu_t$, as in Eq. (12) |

### 4.3. Model selection and in-sample results

In the following, we employ the Beta-Skew-$t$-EGARCH model of Harvey and Sucarrat (2014) with a leverage component as our benchmark model. Beta-Skew-$t$-EGARCH is a score-driven model for volatility based on the skewed Student's $t$ distribution with the skewness mechanism built according to the two-piece method of Fernández and Steel (1998). This model displays features which are similar to those of our specification reported in Section 2 and thus turns out to be an ideal candidate. Indeed, as for score-driven GHSKT, Beta-Skew-$t$-EGARCH exhibits: (i) robustness with respect to extreme observations, (ii) an asymmetric response to the sign of past observations, and (iii) a skewed and fat-tailed conditional distribution. To increase the comparability with the score-driven GHSKT specification, we parameterize the skewed Student's $t$




**(a)** Bitcoin **(b)** Ripple

**Fig. 5.** Filtered conditional variances for BTC (Panel (a)) and XRP (Panel (b)) delivered by the score-driven GHSKT model (black dashed line) and the GARCH model (red solid line). For each figure, sub-panel (i) reports the conditional variance, and sub-panel (ii) reports the log return in percentage points.

distribution of Fernández and Steel (1998) in terms of its conditional variance following the discussion in Trottier and Ardia (2016). We then define a score-driven recursion for the logarithm of the variance of the skewed Student's $t$ distribution, giving rise to a specification which is slightly different from that of Harvey and Sucarrat (2014) and closer to the $t$-GAS specification in Creal et al. (2013). In the case of absence of skewness, the difference between the two specifications, $t$-GAS and Beta-$t$-EGARCH, which is due to the fact that variances are modeled instead of scales, is highlighted by Gao and Zhou (2016). Trucíos (2019) finds that modeling variances ($t$-GAS) provides better results than modeling scales (Beta-$t$-EGARCH) for cryptocurrency series. Note that, in the absence of skewness, the skewed Student's $t$ density we define for the benchmark model collapses to the Student's $t$ distribution parameterized in terms of the variance, giving rise to the $t$-GAS model of Creal et al. (2013). Also, the same applies in our case when $\beta = 0$ in Eq. (2).

Together with Beta-Skew-$t$-EGARCH, we consider the five specifications, $\mathcal{M}_i$, $i = 1, \ldots, 5$, reported in Table 2. $\mathcal{M}_1$ corresponds to the specification of Eq. (1). The specifications $\mathcal{M}_2$ and $\mathcal{M}_3$ extend $\mathcal{M}_1$ by including the additional leverage component (as in Eq. (10)) and the two volatility components (as in Eq. (9)), respectively. The specifications $\mathcal{M}_4$ and $\mathcal{M}_5$ extend $\mathcal{M}_1$ by including time-varying skewness ($\bar{\beta}_t$) and shape ($\nu_t$) parameters, respectively. The benchmark Beta-Skew-$t$-EGARCH specification is denoted by $\mathcal{M}_0$.

We perform model selection using the Bayesian information criteria (BIC) over the specifications reported in Table 2. Given our large cross-section, we compute the percentage of times each specification is selected across the three subgroups and over the whole cross-section in Table 3. The results indicate that models $\mathcal{M}_4$ and $\mathcal{M}_5$ (time-varying $\bar{\beta}_t$ and $\nu_t$, respectively) are those most often selected according to the BIC. The Akaike and Hannan–Quinn information criteria tend to select more often the multiple component specification, $\mathcal{M}_3$, however, leading to qualitatively similar conclusions. Results from a direct comparison between $\mathcal{M}_1$ and $\mathcal{M}_2$ indicate that $\mathcal{M}_2$ is selected approximately 20% of the time according to BIC, irrespective of the level of capitalization. In the supplementary material accompanying this paper, we report additional detailed results about model selection.

**Table 3**
This table reports the percentage of times each specification is selected according to the BIC. The results are reported for the three capitalization groups and over the whole cross-section of cryptocurrencies.

| | $\mathcal{M}_0$ | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ |
|---|---|---|---|---|---|---|
| Small | 12 | 0 | 0 | 0 | 36 | 52 |
| Medium | 7 | 0 | 0 | 2 | 29 | 62 |
| Large | 6 | 0 | 0 | 0 | 39 | 55 |
| All | 8 | 0 | 0 | 1 | 36 | 56 |

### 4.4. Out-of-sample analysis

We now report an extensive out-of-sample analysis to investigate which features should be incorporated in a volatility model to increase the accuracy of predictions for cryptocurrency time series. We divide each series in two parts: an in-sample period where models are estimated, and an out-of-sample period where predictions are made and models are compared. The length of the in-sample and out-of-sample periods is equal and ranges between 350 and 1163 depending on the availability of data. Model parameters are updated every time a new observation becomes available, using an expanding window, and predictions are made for one day ($h = 1$), one week ($h = 7$), and two weeks ($h = 14$) ahead. We focus on (i) volatility, (ii) density, and (iii) risk measure predictions. Specifically, we generate predictions for the $h$-step-ahead conditional variance ($\text{Var}(Y_{t+h}|\mathcal{F}_t)$), conditional density ($p_{t+h|t}(y_{t+h})$), value at risk (VaR) ($\text{VaR}_{t+h|t}(\alpha) = F^{-1}_{t+h|t}(\alpha)$), and expected shortfall ($\text{ES}_{t+h|t}(\alpha) = \mathbb{E}[Y_{t+h}|Y_{t+h} \leq \text{VaR}_{t+h|t}(\alpha), \mathcal{F}_t]$) for the two risk levels $\alpha = 1\%$ and $\alpha = 5\%$. One-step-ahead

predictions ($h = 1$) are available in closed form while multi-step-ahead predictions ($h > 1$) are obtained by sampling 100,000 draws from the predictive distribution $Y_{t+h}|\mathcal{F}_t$.[12]

In the following, we consider a set of proper scoring rules (Gneiting & Raftery, 2007) to compare the predictive ability of the competing specifications. Losses are averaged over the out-of-sample period and their difference with respect to the benchmark specification tested, by employing the two-sided (Diebold & Mariano, 1995) (DM) test at the 5% significance level. Heteroskedasticity and autocorrelation consistent (HAC) standard errors following the procedure of Andrews (1991) are employed in the testing procedure.

*4.4.1. Variance prediction results*

We first analyze results for the variance prediction. Let $\widehat{\sigma}^2_{i,j,t+h|t}$ be the $h$-step-ahead variance prediction computed according to model $i$ for the series $j$, for $h \in \{1, 7, 14\}$, $i = 0, 1, \ldots, 5$, and $j = 1, \ldots, 606$. Variance predictions are compared with the squared returns volatility proxy $\tilde{\sigma}^2_{t+1} = y^2_{t+1}$. Squared returns are valid volatility proxies, but they can be rather noisy. We follow (Patton, 2011) and use the quasi-likelihood (QLIKE) and mean squared error (MSE) loss functions to compare different variance predictions. QLIKE and MSE are defined as:

$$\begin{aligned} \text{MSE}_{i,j,t+h|t} &= (\widehat{\sigma}^2_{i,j,t+h|t} - \tilde{\sigma}^2_{t+1})^2, \\ \text{QLike}_{i,j,t+h|t} &= \log \widehat{\sigma}^2_{i,j,t+h|t} + \frac{\tilde{\sigma}^2_{t+1}}{\widehat{\sigma}^2_{i,j,t+h|t}}, \end{aligned} \tag{15}$$

respectively. Both loss functions induce a perfect ranking of competing volatility forecasts, even for noisy volatility proxies like $\tilde{\sigma}^2_{t+1} = y^2_{t+1}$. QLIKE is more robust than MSE with respect to large observations which occur frequently in cryptocurrency returns series; see Patton (2011). A pairwise comparison between each specification and the benchmark $\mathcal{M}_0$ is computed according to the DM test. Specifically, we compute the loss differential $d^{0,i}_{j,t+h|t} = \text{L}_{0,j,t+h|t} - \text{L}_{i,j,t+h|t}$ for $i = 1, \ldots, 5$, where L = MSE or L = QLike, and test $\mathbb{E}[d^{0,i}_{j,t+h|t}] = 0$ versus the bilateral alternative $\mathbb{E}[d^{0,i}_{j,t+h|t}] \neq 0$ for each $j = 1, \ldots, 606$.

Table 4 reports the percentage of times each specification, $\mathcal{M}_i$, $i = 1, \ldots, 5$, outperforms or underperforms the benchmark Beta-Skew-$t$-EGARCH model. Outperformance (Underperformance) is defined as a rejection of the null hypothesis of equal predictive ability associated with a positive (negative) value of the DM test statistics. Grey cells indicate whether the percentage of outperformance is greater than the percentage of underperformance, and vice versa. The results are very clear and indicate that all score-driven GHSKT models generally outperform the benchmark more often than the times they underperform it. Exceptions are in the case of the time-varying shape parameter model ($\mathcal{M}_5$), when $h = 1$, and the QLike loss is considered and for small capitalized currencies when $h = 7$ and $h = 14$. The results also show that most of the time the null hypothesis of equal predictive ability cannot be rejected at the chosen confidence level of 5%. Indeed, outperformance and underperformance ratios are systematically below 50%. We would like to stress that given the specification of our benchmark, the huge uncertainty implied by the nature of cryptocurrency series, and the relatively short out-of-sample period, it is expected that rejections of the null hypothesis of equal predictive ability between the benchmark and our model are not frequent. Clearly, by specifying a less sophisticated benchmark model, the results would be different. However, we believe that employing a state-of-the-art benchmark provides us with more insightful results.

Looking at Table 4, a clear pattern which links the percentage of outperformance with the amount of market capitalization emerges from the results. Indeed, we see that outperformance is more evident when market capitalization increases. However, it should be noted that larger capitalized series are usually characterized by a longer time series, and thus less uncertainty in the model comparison, which might explain this result.

By comparing the results among the five score-driven GHSKT models, we note that the specification with time-varying skewness coefficient, $\mathcal{M}_4$, is the one that reports higher outperformance rates, especially when the MSE loss is considered. Thus, we suggest implementing this specification for volatility prediction of cryptocurrency time series.

*4.4.2. Risk measure prediction results*

We now focus on the ability of the different models to predict the left tail of the conditional distribution of cryptocurrencies returns. Let $\text{VaR}_{i,j,t+h|t}(\alpha)$ and $\text{ES}_{i,j,t+h|t}(\alpha)$ be the $h$-step-ahead VaR and ES prediction, respectively, at the risk level $\alpha \in \{1\%, 5\%\}$ computed according to model $i$, for series $j$. To compare models in terms of their VaR predictions we employ the so-called quantile loss (QL) function, defined as:

$$\begin{aligned} &\text{QL}_{i,j,t+h|t}(\alpha) \\ &\quad = (\alpha - \mathbb{1}_{\{y_{t+h} < \text{VaR}_{i,j,t+h|t}(\alpha)\}})(y_{t+h} - \text{VaR}_{i,j,t+h|t}(\alpha)). \end{aligned}$$

Evidently, $\text{QL}_{i,j,t+h|t}(\alpha)$ is an asymmetric loss function which more heavily penalizes observations below the predicted VaR level. It can easily be shown that $\mathbb{E}[\text{QL}_{i,j,t+h|t}(\alpha)]$ achieves its minimum at zero when $\text{VaR}_{i,j,t+h|t}(\alpha)$ coincides with the true $h$-step-ahead quantile; see Koenker and Bassett (1978). This property implies that VaR is an *elicitable* risk measure, which means that it can be obtained as the unique minimizer of a loss function. Unfortunately, ES does not share the elicitability properties with VaR. However, recently (Fissler & Ziegel, 2016) (FZ) showed that even if ES is not elicitable, it is a component of a higher-order elicitable functional also defined in terms of VaR. This property implies that a family of functions exists where their unique minimizer

[12] Specifically, we start by sampling 100,000 observations from the one-step-ahead distribution, $Y_{t+1}|\mathcal{F}_t$. This is a GHSKT from which we sample using the stochastic representation reported in Aas and Haff (2006). Subsequently, for each $l = 1, \ldots, h$, we update the time-varying model's parameters using the updating equations reported in Section 2, and we iterate sampling from the one-step-ahead distribution conditional on previous simulated draws until the end of the forecast horizon. Last simulated observations are draws from the distribution $Y_{t+h}|\mathcal{F}_t$.

**Table 4**
This table reports the percentage of times each specification, $\mathcal{M}_i$, $i = 1, \ldots, 5$, outperforms or underperforms the benchmark Beta-Skew-$t$-EGARCH model for volatility prediction. Results are reported for the quasi-likelihood (QLike) and mean squared error (MSE) loss functions, where squared returns are used to proxy the realized volatility, and for three forecast horizons: (i) one day ($h = 1$), (ii) one week ($h = 7$), and (iii) two weeks ($h = 14$). Gray cells indicate when the percentage of outperformance is higher than the corresponding percentage of underperformance. Outperformance (Underperformance) is evaluated as a positive (negative) value of the Diebold–Mariano test statistics evaluated with HAC standard errors associated with a $p$-value lower than 5%.

| | Outperformance | | | | | | | | | | Underperformance | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | QLike | | | | | MSE | | | | | QLike | | | | | MSE | | | | |
| | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ |
| | $h = 1$ | | | | | | | | | | | | | | | | | | | |
| Small | 0.15 | 0.13 | 0.12 | 0.12 | 0.08 | 0.20 | 0.14 | 0.14 | 0.21 | 0.16 | 0.01 | 0.00 | 0.00 | 0.01 | 0.12 | 0.01 | 0.02 | 0.01 | 0.02 | 0.03 |
| Medium | 0.11 | 0.12 | 0.12 | 0.16 | 0.06 | 0.17 | 0.13 | 0.14 | 0.18 | 0.08 | 0.02 | 0.02 | 0.02 | 0.00 | 0.14 | 0.02 | 0.03 | 0.03 | 0.01 | 0.08 |
| Large | 0.23 | 0.21 | 0.21 | 0.18 | 0.13 | 0.27 | 0.30 | 0.30 | 0.31 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.18 | 0.00 | 0.02 | 0.02 | 0.00 | 0.12 |
| All | 0.16 | 0.16 | 0.16 | 0.16 | 0.09 | 0.20 | 0.18 | 0.18 | 0.21 | 0.15 | 0.01 | 0.01 | 0.01 | 0.00 | 0.15 | 0.02 | 0.02 | 0.02 | 0.01 | 0.09 |
| | $h = 7$ | | | | | | | | | | | | | | | | | | | |
| Small | 0.05 | 0.04 | 0.04 | 0.18 | 0.06 | 0.07 | 0.07 | 0.07 | 0.17 | 0.10 | 0.16 | 0.17 | 0.17 | 0.01 | 0.17 | 0.11 | 0.15 | 0.14 | 0.03 | 0.15 |
| Medium | 0.10 | 0.10 | 0.11 | 0.17 | 0.11 | 0.17 | 0.15 | 0.17 | 0.25 | 0.16 | 0.06 | 0.06 | 0.06 | 0.00 | 0.10 | 0.10 | 0.11 | 0.10 | 0.04 | 0.08 |
| Large | 0.21 | 0.22 | 0.22 | 0.22 | 0.19 | 0.31 | 0.31 | 0.28 | 0.39 | 0.25 | 0.03 | 0.06 | 0.05 | 0.02 | 0.12 | 0.11 | 0.12 | 0.12 | 0.03 | 0.10 |
| All | 0.12 | 0.13 | 0.13 | 0.16 | 0.12 | 0.17 | 0.17 | 0.16 | 0.26 | 0.17 | 0.09 | 0.09 | 0.10 | 0.01 | 0.14 | 0.11 | 0.12 | 0.12 | 0.03 | 0.12 |
| | $h = 14$ | | | | | | | | | | | | | | | | | | | |
| Small | 0.07 | 0.07 | 0.05 | 0.19 | 0.19 | 0.07 | 0.08 | 0.07 | 0.17 | 0.19 | 0.21 | 0.25 | 0.21 | 0.05 | 0.19 | 0.19 | 0.25 | 0.23 | 0.07 | 0.18 |
| Medium | 0.21 | 0.18 | 0.18 | 0.22 | 0.23 | 0.19 | 0.20 | 0.18 | 0.26 | 0.25 | 0.10 | 0.08 | 0.09 | 0.02 | 0.12 | 0.12 | 0.12 | 0.12 | 0.02 | 0.09 |
| Large | 0.33 | 0.28 | 0.29 | 0.32 | 0.27 | 0.40 | 0.39 | 0.38 | 0.43 | 0.39 | 0.07 | 0.07 | 0.07 | 0.03 | 0.12 | 0.13 | 0.11 | 0.11 | 0.05 | 0.10 |
| All | 0.19 | 0.18 | 0.17 | 0.22 | 0.23 | 0.21 | 0.22 | 0.20 | 0.29 | 0.26 | 0.13 | 0.13 | 0.13 | 0.03 | 0.15 | 0.15 | 0.15 | 0.15 | 0.04 | 0.12 |

is equal to the true pair of VaR and ES. Among these functions, we employ the one used by Patton et al. (2019), which assumes that $\text{VaR}_{i,j,t+h|t}(\alpha) < 0$ and $\text{ES}_{i,j,t+h|t}(\alpha) < \text{VaR}_{i,j,t+h|t}(\alpha)$, two conditions that hold by construction in our case. The joint loss function for ES and VaR is defined as:

$$\text{FZ}_{i,j,t+h|t}(\alpha) = \frac{\mathbb{1}_{\{y_{t+h} < \text{VaR}_{i,j,t+h|t}(\alpha)\}}[y_{t+h} - \text{VaR}_{i,j,t+h|t}(\alpha)]}{\alpha \text{ES}_{i,j,t+h|t}(\alpha)} + \frac{\text{VaR}_{i,j,t+h|t}(\alpha)}{\text{ES}_{i,j,t+h|t}(\alpha)} + \log(-\text{ES}_{i,j,t+h|t}(\alpha)) - 1.$$

The reader is referred to Patton et al. (2019) for additional details. Losses are averaged over the out-of-sample period, and model comparison is performed similarly to the volatility case by employing the DM test. Tables 5 and 6 report the percentages of outperformance and underperformance with respect to the benchmark model for the two probability risk levels $\alpha = 1\%$ and $\alpha = 5\%$, respectively. In the case where $\alpha = 1\%$, the results are very clear and match those obtained in the volatility case. VaR and ES predictions from the score-driven GHSKT model are most of the time preferred over those computed from the benchmark model. This result is robust across the different model specifications, forecast horizons, and loss functions. Also in this case, the top performer is the $\mathcal{M}_4$ specification, which includes the time-varying skewness component. The results for the FZ loss function strongly indicate the score-driven GHSKT specification as the top performer with outperformance ratios in the range of 60%–80%, well above 50%. The results for the $\alpha = 5\%$ risk levels are a bit mixed, indicating that when we focus on a less extreme part of the conditional distribution, the Beta-Skew-$t$-EGARCH and the score-driven GHSKT specifications behave similarly. However, we also observe in this case that $\mathcal{M}_4$ performs very well, irrespective of the level of market capitalization and forecast horizon.

To conclude our risk measure prediction analysis, we assess whether VaR and ES predictions provide good coverage of the left tail of the one-step-ahead predictive distribution of cryptocurrency returns. For VaR predictions we employ the dynamic quantile (DQ) test of Engle and Manganelli (2004) which tests the null hypothesis of correct conditional and unconditional coverage of VaR violations.[13] To backtest ES predictions, we follow McNeil and Frey (2000) and define the sequence of "exceedance residuals" $\varepsilon^*_{t+1} = \frac{y_{t+1} - \text{ES}_{t+1|t}(\alpha)}{\widehat{\sigma}_{t+1|t}}$. Conditional on the event $\{y_{t+1} < \text{VaR}_{t+1|t}(\alpha)\}$, it is easy to see that, under the assumption of correct ES, VaR, and $\sigma_{t+1}$ predictions, $\varepsilon^*_{t+1}$ has zero mean. So, the null hypothesis of McNeil and Frey (2000) can be formulated as $\mathbf{H}_0 : \mathbb{E}[\varepsilon^*_{t+1} | y_{t+1} < \text{VaR}_{t+1|t}(\alpha)] = 0$. We test this against the one-sided alternative hypothesis $\mathbb{E}[\varepsilon^*_{t+1} | y_{t+1} < \text{VaR}_{t+1|t}(\alpha)] < 0$, which is the one most relevant for risk management applications. We follow McNeil and Frey (2000) and approximate the test statistics distribution with 1000 independent bootstrap replications, as in Algorithm 16.2 of Tibshirani and Efron (1993).

Table 7 reports the percentage of times we fail to reject the null hypothesis of the DQ and ES tests at the 5% significance level for all specification, $\mathcal{M}_i$, $0 = 1, \ldots, 5$, at the $\alpha = 5\%$ and $\alpha = 1\%$ risk levels. The results are homogeneous with respect to market capitalization. Regarding the DQ test, we find that for the benchmark model ($\mathcal{M}_0$), we fail to reject $\mathbf{H}_0$ approximately 63% of the times when $\alpha = 5\%$ and 60% of the times when $\alpha = 1\%$. When $\alpha = 5\%$, the results for $\mathcal{M}_i$ for $i =$

[13] The DQ test has more power against some forms of model misspecification compared to other tests, like the conditional coverage test of Christoffersen (1998).

**Table 5**
This table reports the percentage of times each specification, $\mathcal{M}_i$, $i = 1, \ldots, 5$, outperforms or underperforms the benchmark Beta-Skew-$t$-EGARCH model for VaR and ES predictions at the $\alpha = 1\%$ risk level. The results are reported for the quantile loss (QL) and the joint VaR and ES loss (FZ) for three forecast horizons: (i) one day ($h = 1$), (ii) one week ($h = 7$), and (iii) two weeks ($h = 14$). Grey cells indicate when the percentage of outperformance is higher than the corresponding percentage of underperformance. Outperformance (Underperformance) is evaluated as a positive (negative) value of the Diebold–Mariano test statistics evaluated with HAC standard errors associated with a $p$-value lower than 5%.

| | Outperformance | | | | | | | | | | Underperformance | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | QL 1% | | | | | FZ 1% | | | | | QL 1% | | | | | FZ 1% | | | | |
| | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ |
| | $h = 1$ | | | | | | | | | | | | | | | | | | | |
| Small | 0.08 | 0.08 | 0.08 | 0.05 | 0.04 | 0.74 | 0.79 | 0.79 | 0.78 | 0.36 | 0.07 | 0.09 | 0.09 | 0.09 | 0.05 | 0.03 | 0.02 | 0.02 | 0.05 | 0.16 |
| Medium | 0.12 | 0.12 | 0.12 | 0.09 | 0.07 | 0.64 | 0.69 | 0.70 | 0.74 | 0.39 | 0.04 | 0.04 | 0.04 | 0.02 | 0.07 | 0.06 | 0.04 | 0.04 | 0.05 | 0.17 |
| Large | 0.15 | 0.12 | 0.11 | 0.14 | 0.04 | 0.69 | 0.79 | 0.79 | 0.83 | 0.41 | 0.01 | 0.02 | 0.02 | 0.01 | 0.04 | 0.12 | 0.07 | 0.07 | 0.09 | 0.13 |
| All | 0.12 | 0.10 | 0.10 | 0.10 | 0.06 | 0.70 | 0.77 | 0.77 | 0.79 | 0.43 | 0.03 | 0.04 | 0.04 | 0.03 | 0.07 | 0.07 | 0.05 | 0.05 | 0.07 | 0.16 |
| | $h = 7$ | | | | | | | | | | | | | | | | | | | |
| Small | 0.02 | 0.03 | 0.03 | 0.04 | 0.07 | 0.52 | 0.60 | 0.60 | 0.66 | 0.44 | 0.07 | 0.09 | 0.09 | 0.04 | 0.07 | 0.20 | 0.19 | 0.21 | 0.12 | 0.23 |
| Medium | 0.10 | 0.10 | 0.08 | 0.14 | 0.10 | 0.58 | 0.60 | 0.60 | 0.64 | 0.47 | 0.09 | 0.08 | 0.08 | 0.04 | 0.07 | 0.20 | 0.16 | 0.15 | 0.17 | 0.21 |
| Large | 0.12 | 0.12 | 0.11 | 0.14 | 0.17 | 0.62 | 0.70 | 0.71 | 0.74 | 0.52 | 0.07 | 0.07 | 0.05 | 0.02 | 0.03 | 0.20 | 0.14 | 0.14 | 0.12 | 0.17 |
| All | 0.08 | 0.08 | 0.08 | 0.11 | 0.11 | 0.59 | 0.65 | 0.65 | 0.67 | 0.49 | 0.09 | 0.09 | 0.09 | 0.04 | 0.08 | 0.20 | 0.16 | 0.16 | 0.15 | 0.22 |
| | $h = 14$ | | | | | | | | | | | | | | | | | | | |
| Small | 0.08 | 0.07 | 0.05 | 0.14 | 0.11 | 0.47 | 0.54 | 0.54 | 0.64 | 0.44 | 0.13 | 0.14 | 0.12 | 0.06 | 0.07 | 0.25 | 0.23 | 0.22 | 0.14 | 0.27 |
| Medium | 0.12 | 0.13 | 0.12 | 0.17 | 0.21 | 0.56 | 0.59 | 0.59 | 0.63 | 0.51 | 0.11 | 0.11 | 0.12 | 0.03 | 0.07 | 0.23 | 0.24 | 0.25 | 0.15 | 0.22 |
| Large | 0.13 | 0.17 | 0.16 | 0.17 | 0.24 | 0.60 | 0.69 | 0.69 | 0.74 | 0.64 | 0.10 | 0.08 | 0.08 | 0.03 | 0.03 | 0.23 | 0.17 | 0.17 | 0.12 | 0.17 |
| All | 0.11 | 0.13 | 0.12 | 0.15 | 0.19 | 0.56 | 0.62 | 0.62 | 0.66 | 0.54 | 0.11 | 0.11 | 0.10 | 0.04 | 0.07 | 0.24 | 0.21 | 0.21 | 0.15 | 0.23 |

**Table 6**
This table reports the percentage of times each specification, $\mathcal{M}_i$, $i = 1, \ldots, 5$, outperforms or underperforms the benchmark Beta-Skew-$t$-EGARCH model for VaR and ES predictions at the $\alpha = 5\%$ risk level. The results are reported for the quantile loss (QL) and the joint VaR and ES loss (FZ) for the three forecast horizons: (i) one day ($h = 1$), (ii) one week ($h = 7$), and (iii) two weeks ($h = 14$). Grey cells indicate when the percentage of outperformance is higher than the corresponding percentage of underperformance. Outperformance (Underperformance) is evaluated as a positive (negative) value of the Diebold–Mariano test statistics evaluated with HAC standard errors associated with a $p$-value lower than 5%.

| | Outperformance | | | | | | | | | | Underperformance | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | QL 5% | | | | | FZ 5% | | | | | QL 5% | | | | | FZ 5% | | | | |
| | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ |
| | $h = 1$ | | | | | | | | | | | | | | | | | | | |
| Small | 0.03 | 0.07 | 0.07 | 0.06 | 0.06 | 0.15 | 0.25 | 0.24 | 0.33 | 0.12 | 0.04 | 0.03 | 0.03 | 0.03 | 0.22 | 0.25 | 0.17 | 0.16 | 0.27 | 0.59 |
| Medium | 0.07 | 0.07 | 0.07 | 0.15 | 0.03 | 0.20 | 0.23 | 0.23 | 0.34 | 0.12 | 0.03 | 0.02 | 0.02 | 0.02 | 0.27 | 0.33 | 0.22 | 0.22 | 0.31 | 0.53 |
| Large | 0.07 | 0.05 | 0.05 | 0.09 | 0.05 | 0.19 | 0.30 | 0.30 | 0.49 | 0.13 | 0.02 | 0.02 | 0.02 | 0.01 | 0.13 | 0.33 | 0.17 | 0.17 | 0.28 | 0.60 |
| All | 0.07 | 0.07 | 0.07 | 0.09 | 0.05 | 0.16 | 0.25 | 0.25 | 0.37 | 0.14 | 0.04 | 0.03 | 0.03 | 0.03 | 0.22 | 0.30 | 0.18 | 0.18 | 0.29 | 0.53 |
| | $h = 7$ | | | | | | | | | | | | | | | | | | | |
| Small | 0.05 | 0.07 | 0.05 | 0.15 | 0.02 | 0.18 | 0.22 | 0.21 | 0.32 | 0.15 | 0.12 | 0.11 | 0.12 | 0.02 | 0.16 | 0.42 | 0.39 | 0.38 | 0.24 | 0.59 |
| Medium | 0.09 | 0.06 | 0.06 | 0.17 | 0.07 | 0.21 | 0.29 | 0.30 | 0.35 | 0.22 | 0.11 | 0.07 | 0.10 | 0.02 | 0.22 | 0.41 | 0.29 | 0.29 | 0.27 | 0.55 |
| Large | 0.06 | 0.07 | 0.06 | 0.17 | 0.07 | 0.26 | 0.32 | 0.31 | 0.45 | 0.16 | 0.17 | 0.10 | 0.11 | 0.03 | 0.13 | 0.47 | 0.30 | 0.29 | 0.23 | 0.62 |
| All | 0.06 | 0.06 | 0.05 | 0.14 | 0.06 | 0.21 | 0.30 | 0.30 | 0.37 | 0.19 | 0.14 | 0.10 | 0.11 | 0.03 | 0.19 | 0.43 | 0.31 | 0.31 | 0.26 | 0.56 |
| | $h = 14$ | | | | | | | | | | | | | | | | | | | |
| Small | 0.07 | 0.07 | 0.09 | 0.13 | 0.08 | 0.21 | 0.24 | 0.24 | 0.34 | 0.16 | 0.18 | 0.18 | 0.16 | 0.07 | 0.17 | 0.46 | 0.42 | 0.43 | 0.25 | 0.59 |
| Medium | 0.07 | 0.09 | 0.09 | 0.21 | 0.13 | 0.28 | 0.37 | 0.37 | 0.40 | 0.22 | 0.13 | 0.09 | 0.09 | 0.02 | 0.19 | 0.45 | 0.34 | 0.32 | 0.27 | 0.54 |
| Large | 0.10 | 0.07 | 0.07 | 0.15 | 0.07 | 0.26 | 0.36 | 0.36 | 0.50 | 0.17 | 0.09 | 0.08 | 0.08 | 0.03 | 0.21 | 0.50 | 0.35 | 0.34 | 0.22 | 0.64 |
| All | 0.07 | 0.09 | 0.09 | 0.16 | 0.10 | 0.25 | 0.34 | 0.34 | 0.41 | 0.20 | 0.15 | 0.13 | 0.13 | 0.05 | 0.20 | 0.47 | 0.35 | 0.35 | 0.26 | 0.57 |

$1, \ldots, 5$ are better than those of the benchmark $\mathcal{M}_0$. For example, looking at the whole cross-section of data, for $\mathcal{M}_3$ we fail to reject the null 78% of the time, while for $\mathcal{M}_0$ we fail to reject it only 60% of the time. When $\alpha = 1\%$, the results are similar across all specifications. Overall, according to the DM test, $\mathcal{M}_3$ reports the best results. Looking at the ES test, we find that the results for $\mathcal{M}_i$ for $i = 1, \ldots, 5$ are very good, with percentages around 95% when $\alpha = 5\%$ and 90% when $\alpha = 1\%$. The percentages for the benchmark $\mathcal{M}_0$ model are around 70% for $\alpha = 5\%$ and 65% for $\alpha = 1\%$. Overall, we find that all specifications constructed, starting from the score-driven GHSKT model, report good coverage of the tail of the one-step-ahead conditional distribution. In the supplementary material accompanying this paper, we report additional backtesting results for the ES predictions using the tests of Bayer and Dimitriadis (2020) and Nolde and Ziegel (2017). The results are similar to those reported in Table 7.

To assess the economic gains of the score-driven GHSKT model and its generalizations with respect to the benchmark model, we compute the average absolute deviations of violations from VaR forecasts (AD) and the average daily capital requirements (CR) with multiplicative factor 3, as in McAleer and Da Veiga (2008).

**Table 7**
This table reports the percentage of times we fail to reject the null hypothesis of the correct coverage of the left tail of the one-step-ahead ($h = 1$) conditional distribution of the returns according to the dynamic quantile (DQ) test of Engle and Manganelli (2004) and the null hypothesis of zero mean of exceedance residuals (ES test), as in McNeil and Frey (2000), at the 5% significance level. The distribution of the test statistics of McNeil and Frey (2000) is approximated with 1000 independent bootstrap replications. The results are reported for each specification, $\mathcal{M}_i$, $0 = 1, \ldots, 5$, at the $\alpha = 1\%$ and $\alpha = 5\%$ risk levels.

| | DQ Test | | | | | | ES Test | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\mathcal{M}_0$ | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ | $\mathcal{M}_0$ | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ |
| | $\alpha = 5\%$ | | | | | | | | | | | |
| Small | 0.68 | 0.81 | 0.79 | 0.83 | 0.71 | 0.79 | 0.79 | 0.95 | 0.96 | 0.97 | 0.93 | 0.96 |
| Medium | 0.58 | 0.73 | 0.73 | 0.79 | 0.58 | 0.73 | 0.74 | 0.98 | 0.98 | 0.99 | 0.97 | 0.98 |
| Large | 0.66 | 0.75 | 0.76 | 0.82 | 0.62 | 0.78 | 0.70 | 0.95 | 0.97 | 0.96 | 0.88 | 0.97 |
| All | 0.60 | 0.72 | 0.72 | 0.78 | 0.61 | 0.72 | 0.70 | 0.95 | 0.96 | 0.96 | 0.93 | 0.95 |
| | $\alpha = 1\%$ | | | | | | | | | | | |
| Small | 0.57 | 0.60 | 0.53 | 0.54 | 0.59 | 0.53 | 0.73 | 0.91 | 0.95 | 0.92 | 0.91 | 0.96 |
| Medium | 0.59 | 0.59 | 0.53 | 0.61 | 0.51 | 0.53 | 0.68 | 0.91 | 0.94 | 0.89 | 0.95 | 0.93 |
| Large | 0.61 | 0.52 | 0.50 | 0.59 | 0.52 | 0.50 | 0.56 | 0.81 | 0.86 | 0.84 | 0.83 | 0.87 |
| All | 0.60 | 0.57 | 0.54 | 0.59 | 0.53 | 0.54 | 0.65 | 0.87 | 0.90 | 0.88 | 0.91 | 0.90 |

**Table 8**
This table reports the average VaR absolute deviations (AD mean) and capital requirements (CR) with multiplicative factor 3, as in McAleer and Da Veiga (2008). The results are reported for each specification, $\mathcal{M}_i$, $0 = 1, \ldots, 5$, at the $\alpha = 1\%$ and $\alpha = 5\%$ risk levels.

| | AD | | | | | | CR | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\mathcal{M}_0$ | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ | $\mathcal{M}_0$ | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ |
| | $\alpha = 5\%$ | | | | | | | | | | | |
| Small | 14.76 | 14.10 | 13.69 | 14.08 | 15.88 | 14.04 | 92.31 | 95.86 | 95.54 | 93.23 | 107.65 | 95.54 |
| Medium | 9.81 | 9.71 | 9.71 | 9.60 | 10.57 | 9.71 | 66.85 | 67.99 | 66.37 | 67.27 | 72.11 | 66.37 |
| Large | 4.78 | 4.79 | 4.76 | 4.59 | 4.78 | 4.75 | 37.10 | 37.10 | 36.56 | 36.26 | 40.40 | 36.56 |
| All | 9.18 | 8.75 | 8.72 | 8.64 | 9.59 | 8.72 | 64.08 | 65.48 | 64.47 | 64.05 | 68.94 | 64.47 |
| | $\alpha = 1\%$ | | | | | | | | | | | |
| Small | 15.80 | 15.48 | 15.62 | 15.30 | 16.29 | 15.76 | 169.82 | 154.81 | 152.70 | 150.46 | 162.66 | 152.70 |
| Medium | 12.61 | 11.82 | 12.06 | 11.70 | 12.07 | 12.06 | 120.81 | 109.53 | 106.15 | 107.55 | 109.51 | 106.15 |
| Large | 4.82 | 4.75 | 4.68 | 4.61 | 4.52 | 4.68 | 63.71 | 60.14 | 58.98 | 57.80 | 61.25 | 58.98 |
| All | 10.49 | 9.90 | 9.60 | 9.76 | 11.06 | 9.61 | 114.30 | 104.19 | 100.29 | 100.90 | 104.61 | 100.29 |

Specifically, AD is computed as $\text{AD}(\alpha) = \left(\sum_{s=1}^{S} i_{Q+s}\right)^{-1} \left(\sum_{s=1}^{S} i_{Q+s} |y_{Q+s} - \text{VaR}_{Q+s|Q+s-1}(\alpha)|\right)$ where $i_{Q+s} = \mathbb{1}(y_{Q+s} < \text{VaR}_{Q+s|Q+s-1}(\alpha))$, and where $S$ and $Q$ are the length of the in-sample and out-of-sample periods, respectively. The average daily capital requirement is computed as $\text{CR}(\alpha) = -kS^{-1} \sum_{s=1}^{S} \min\left(\text{VaR}_{Q+s|Q+s-1}(\alpha), \overline{\text{VaR}}_{Q+s}(\alpha)\right)$, where $\overline{\text{VaR}}_{Q+s}(\alpha)$ is the average daily VaR over the past 60 days, and $k = 3$.

Table 8 reports the average AD and CR computed for different levels of market capitalization and for the whole sample of cryptocurrencies. Smaller values indicate better performance and are highlighted in gray. The results again show that the score-driven GHSKT specifications generally outperform the benchmark model. Among the several extensions of the score-driven GHSKT, $\mathcal{M}_3$ reports best results, followed by $\mathcal{M}_2$, indicating that two volatility components and the presence of the leverage effect help in reducing the magnitude of VaR violations and the daily capital requirement.

*4.4.3. Density prediction results*

We now analyze the results for density prediction. Let $p_{i,j,t+h|t}(y_{t+h})$ be the $h$-step-ahead predicted density evaluated in $y_{t+h}$ for model $i$ and series $j$. Model comparison is done according to the continuous ranked probability score (CRPS) scoring rule (Matheson & Winkler, 1976) and its weighed version (wCRPS) introduced by Gneiting and Ranjan (2011). The CRPS integrates over the Brier score, which measures the (quadratic) distance between the predicted and empirical distribution function proxied by a step function in $y_{t+h}$. Formally, we define:

$$\text{CRPS}_{i,j,t,|t+h} = \int_{\Re} \left(F_{i,j,t+h|t}(z) - \mathbb{1}_{\{y_{t+h} \le z\}}\right)^2 \text{d}z,$$

where $F_{i,j,t+h|t}(z) = \int_{-\infty}^{z} p_{i,j,t+h|t}(x)\text{d}x$ is the $h$-step-ahead predictive CDF of model $i$ for series $j$ evaluated in $z$. The wCRPS is obtained by weighting the Brier score as follows:

$$\text{wCRPS}_{i,j,t,|t+h} = \int_{\Re} \omega(z) \left(F_{i,j,t+h|t}(z) - \mathbb{1}_{\{y_{t+h} \le z\}}\right)^2 \text{d}z,$$

where $\omega : \Re \to \Re^+$ is a continuous weight function, which emphasizes regions of interest of the predictive distribution, such as the tails or the center. We report results for the left tail of the distribution (the losses) and follow Gneiting and Ranjan (2011) by defining the decreasing weight function $\omega(z) = 1 - \Phi(z/c)$, where $\Phi$ is the CDF of a standard Gaussian distribution, and $c = 10$ is set to match the scale of the data. This way, discrepancies in the left tail of the return distribution are weighed more than those in the right tail. Note that, when $\omega(z) = 1$, the wCRPS and CRPS are equivalent. Alternative weight specifications, focusing on the right tail and on the

center of the predictive distribution, lead to similar conclusions and are reported in the supplementary material accompanying this paper.[14]

Table 9 reports the results for the CRPS and left tail wCRPS measures. The results are similar to the risk measure prediction when $\alpha = 5\%$ and indicate that for $h = 1$ the ratios of outperformance of the score-driven GHSKT specifications are higher than the ratios of underperformance with respect to the Beta-Skew-$t$-EGARCH model. For multi-step-ahead prediction, $h > 1$, the results change and indicate that the Beta-Skew-$t$-EGARCH model performs better. An exception is again the specification with time-varying skewness, $\mathcal{M}_4$, which outperforms the benchmark for all forecast horizon. Thus, our indication for density prediction is to implement the $\mathcal{M}_4$ model. Note that evidence of underperformance of the score-driven GHSKT with respect to Beta-Skew-$t$-EGARCH for multi-step-ahead density predictions is an interesting result per se. Indeed, while the volatility filter implied by the GHSKT assumption leads to better volatility predictions than the one implied by the skewed Student's $t$ of Fernández and Steel (1998) (see Table 4), there is evidence that density predictions are better when the skewed Student's $t$ distributional assumption is made.

### 4.5. Individual results for bitcoin, ethereum, ripple, and litecoin

Detailed results for the four main cryptocurrencies are omitted to save space and reported in the supplementary material accompanying this paper. Overall, we do not find significant differences between the five score-driven GHSKT specifications $\mathcal{M}_i$, $i = 1, \ldots, 5$. Below we summarize the results.

#### 4.5.1. Volatility prediction

(i) For BTC and LTC, we observe outperformance of the score-driven GHSKT specifications with respect to the benchmark for all forecast horizons. When the loss function is MSE, gains are in the range of 7%–18% and the null hypothesis of equal predictive ability is rejected. When the loss is QLike, gains are in the range of 1%–3% and the null hypothesis of equal predictive ability cannot be rejected.

(ii) The results for ETH and XRP are similar among all specifications.

#### 4.5.2. Risk measure prediction

(i) For the QL loss function, the results are similar across the considered specifications, and the null hypothesis of equal predictive ability cannot be rejected.

(ii) For the FZ loss when $\alpha = 1\%$, the score-driven GHSKT specifications outperform Beta-Skew-$t$-EGARCH with gains in the range of 20%–30% for all four cryptocurrencies and all forecast horizons.

(iii) When $\alpha = 5\%$, BTC and LTC outperformance gains are in the range of 1%–6%. For ETH and XRP, the score-driven GHSKT specifications underperform Beta-Skew-$t$-EGARCH with losses in the range of 1%–10%.

#### 4.5.3. Density prediction

The results for density prediction indicate that the performance of the score-driven GHSKT and Beta-Skew-$t$-EGARCH specifications is similar. In this case, the null hypothesis of equal predictive ability is rarely rejected.

## 5. Conclusion

In this paper, we analyzed and forecasted cryptocurrencies' financial time series, of which BTC is the most prominent example. The dynamics of these series are quite complex and they display extreme observations, asymmetries, and several nonlinear characteristics that are difficult to model. To accommodate these features, we developed a model that can incorporate long memory, leverage effects, and time-varying higher-order moments. The finite sample properties of the MLE were investigated via a Monte Carlo experiment reported in the supplementary material accompanying this paper. The results show that the MLE performs well in term of the bias and RMSE, and that it performs well with small samples. In the empirical application, we applied this model to 606 cryptocurrencies, concentrating on the main four: BTC, ETH, XPR, and LTC. We found that a robust filter for the volatility of cryptocurrencies' time series is strongly required by the data. Our forecast analysis indicated that incorporating time-varying skewness significantly improves volatility, risk measures, and density predictions. The inclusion of additional features, like multiple-component volatility and a "leverage" term in the variance recursion, also improves the results. We found that time-varying kurtosis does not play a predominant role for cryptocurrency time series. The results were robust to different levels of market capitalization and forecast horizons. Overall, our suggestion for investors and risk managers is to implement a robust specification (like a score-driven specification with a fat-tailed conditional distribution) with the inclusion of time-varying skewness, especially if the goal is to analyze a large number of cryptocurrencies.

## Acknowledgments

Leopoldo Catania acknowledges support from the Aarhus Universitets Forskningsfond (AUFF) NOVA under the grant "The Dynamic Shape of Randomness" and from CREATES, the Center for Research in Econometric Analysis of Time Series (DNRF78) at Aarhus University, Denmark, funded by the Danish National Research Foundation.

Stefano Grassi gratefully acknowledges financial support from the University of Rome "Tor Vergata", Italy under the grant "Beyond Borders" (CUP: E84I20000900005).

[14] We follow the implementation of Gneiting and Ranjan (2011) and compute wCRPS with the following approximation:

$$\text{wCRPS}_{i,j,t,|t+h} \approx \frac{z_u - z_l}{M-1} \sum_{m=1}^{M} w(z_m) \left(F_{i,j,t+h|t}(z) - \mathbb{1}_{\{y_{t+h} \le z\}}\right)^2,$$

where $z_m = z_l + m \times (z_u - z_l)/M$, and $z_u$ and $z_l$ are the upper and lower values, respectively, which define the range of integration. The accuracy of the approximation can be increased to any desired level by $M$. Setting $z_l = -100$, $z_u = 100$ and $M = 1000$ provides an accurate approximation when working with returns in percentage points.

**Table 9**
This table reports the percentage of times each specification, $\mathcal{M}_i$, $i = 1, \ldots, 5$, outperforms or underperforms the benchmark Beta-Skew-$t$-EGARCH model for density prediction. The results are reported for the continuous ranked probability score (CRPS) and its weighted version (wCRPS) with emphasis on the left tail for the three forecast horizons: (i) one day ($h = 1$), (ii) one week ($h = 7$), and (iii) two weeks ($h = 14$). Grey cells indicate when the percentage of outperformance is higher than the corresponding percentage of underperformance. Outperformance (Underperformance) is evaluated as a positive (negative) value of the Diebold–Mariano test statistics evaluated with HAC standard errors associated with a $p$-value lower than 5%.

| | Outperformance | | | | | | | | | | Underperformance | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | CRPS | | | | | Left tail wCRPS | | | | | CRPS | | | | | Left tail wCRPS | | | | |
| | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ |
| | $h = 1$ | | | | | | | | | | | | | | | | | | | |
| Small | 0.13 | 0.15 | 0.14 | 0.21 | 0.10 | 0.07 | 0.09 | 0.10 | 0.07 | 0.03 | 0.03 | 0.03 | 0.02 | 0.00 | 0.44 | 0.12 | 0.07 | 0.08 | 0.04 | 0.38 |
| Medium | 0.14 | 0.13 | 0.15 | 0.21 | 0.10 | 0.07 | 0.08 | 0.09 | 0.08 | 0.06 | 0.02 | 0.03 | 0.03 | 0.00 | 0.32 | 0.07 | 0.06 | 0.05 | 0.02 | 0.35 |
| Large | 0.02 | 0.02 | 0.02 | 0.10 | 0.09 | 0.02 | 0.02 | 0.02 | 0.04 | 0.02 | 0.02 | 0.03 | 0.02 | 0.01 | 0.45 | 0.06 | 0.07 | 0.07 | 0.02 | 0.36 |
| All | 0.09 | 0.09 | 0.10 | 0.16 | 0.09 | 0.06 | 0.07 | 0.07 | 0.07 | 0.04 | 0.03 | 0.04 | 0.03 | 0.01 | 0.37 | 0.07 | 0.06 | 0.06 | 0.03 | 0.32 |
| | $h = 7$ | | | | | | | | | | | | | | | | | | | |
| Small | 0.03 | 0.03 | 0.04 | 0.11 | 0.07 | 0.02 | 0.01 | 0.01 | 0.04 | 0.02 | 0.23 | 0.28 | 0.29 | 0.06 | 0.36 | 0.21 | 0.17 | 0.20 | 0.04 | 0.31 |
| Medium | 0.07 | 0.04 | 0.04 | 0.18 | 0.06 | 0.02 | 0.02 | 0.02 | 0.10 | 0.02 | 0.09 | 0.12 | 0.12 | 0.02 | 0.30 | 0.15 | 0.12 | 0.14 | 0.03 | 0.27 |
| Large | 0.02 | 0.02 | 0.02 | 0.19 | 0.12 | 0.00 | 0.01 | 0.01 | 0.04 | 0.02 | 0.17 | 0.21 | 0.21 | 0.05 | 0.44 | 0.23 | 0.20 | 0.17 | 0.07 | 0.37 |
| All | 0.03 | 0.03 | 0.03 | 0.15 | 0.07 | 0.01 | 0.01 | 0.01 | 0.05 | 0.02 | 0.17 | 0.20 | 0.20 | 0.04 | 0.34 | 0.19 | 0.16 | 0.16 | 0.04 | 0.30 |
| | $h = 14$ | | | | | | | | | | | | | | | | | | | |
| Small | 0.04 | 0.03 | 0.02 | 0.13 | 0.04 | 0.02 | 0.02 | 0.02 | 0.05 | 0.02 | 0.36 | 0.38 | 0.35 | 0.10 | 0.40 | 0.26 | 0.27 | 0.26 | 0.08 | 0.35 |
| Medium | 0.06 | 0.03 | 0.03 | 0.26 | 0.10 | 0.01 | 0.02 | 0.02 | 0.12 | 0.02 | 0.17 | 0.17 | 0.15 | 0.01 | 0.29 | 0.19 | 0.17 | 0.18 | 0.03 | 0.30 |
| Large | 0.04 | 0.02 | 0.02 | 0.23 | 0.10 | 0.02 | 0.01 | 0.01 | 0.03 | 0.03 | 0.31 | 0.31 | 0.31 | 0.07 | 0.44 | 0.27 | 0.23 | 0.27 | 0.08 | 0.43 |
| All | 0.03 | 0.03 | 0.02 | 0.18 | 0.07 | 0.01 | 0.02 | 0.02 | 0.05 | 0.03 | 0.25 | 0.27 | 0.26 | 0.06 | 0.36 | 0.23 | 0.23 | 0.23 | 0.05 | 0.33 |

## Appendix A. Details on the GHSKT parameterization

The first two central moments of the GHSKT distribution detailed in Aas and Haff (2006) are given by:

$$\mathbb{E}(Y) = \mu + \frac{\beta\delta^2}{\nu-2}, \qquad \text{Var}(Y) = \frac{2\beta^2\delta^4}{(\nu-2)^2(\nu-4)} + \frac{\delta^2}{\nu-2}, \tag{A.1}$$

where we note that the scale enters the mean of $Y$. This implies that constructing a filter for $\delta$ (i.e., $\delta_t = \psi(\delta_{t-1}, y_{t-1})$) would prevent a clear interpretation in terms of volatility estimates. Therefore, the implementation of the GHSKT distribution detailed in Aas and Haff (2006) is not immediate in time-varying volatility contexts.[15] To this end, we impose the two moment conditions $\mathbb{E}(Y) = 0$ and $\text{Var}(Y) = 1$, which are fulfilled by:

$$\mu = -\frac{\beta\delta^2}{\nu-2}, \qquad \delta = \left(\frac{2\bar{\beta}^2}{(\nu-2)^2(\nu-4)} + \frac{1}{\nu-2}\right)^{-1/2}, \tag{A.2}$$

where $\bar{\beta} = \beta\delta$. Now we can build the map $(\bar{\beta}, \nu) \to (\mu, \delta, \beta, \nu)$, or equivalently $\boldsymbol{\theta} \to \boldsymbol{\theta}^*$, such that for any value of $\bar{\beta} \in \Re$ and $\nu > 4$, the new density $f(y; \boldsymbol{\theta})$ satisfies the two moment conditions we require. Note that the mapping $\boldsymbol{\theta} \to \boldsymbol{\theta}^*$ is bijective, and the associated $4 \times 2$ Jacobian matrix, $\mathcal{J}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \partial\boldsymbol{\theta}^*/\partial\boldsymbol{\theta}$, has elements:

$$\begin{aligned}
\frac{\partial\delta}{\partial\bar{\beta}} &= -\frac{2\bar{\beta}}{(\nu-2)^2(\nu-4)}\xi, \\
\frac{\partial\delta}{\partial\nu} &= -\frac{1}{2}\left[\frac{-2\bar{\beta}^2[2(\nu-2)(\nu-4)+(\nu-2)^2]}{(\nu-2)^2(\nu-4)} - \frac{1}{(\nu-2)^2}\right]\xi,
\end{aligned} \tag{A.3}$$

where $\partial\beta/\partial\bar{\beta} = \left(\delta - \bar{\beta}\frac{\partial\delta}{\partial\bar{\beta}}\right)/\delta^2$, $\partial\beta/\partial\nu = -(\bar{\beta}/\delta^2)(\partial\delta/\partial\nu)$, $\partial\mu/\partial\bar{\beta} = -\left(\delta + \bar{\beta}\frac{\partial\delta}{\partial\bar{\beta}}\right)/(\nu-2)$, $\partial\mu/\partial\nu = -\bar{\beta}\left[\left(\frac{\partial\beta}{\partial\nu}(\nu-2) - \delta\right)/(\nu-2)^2\right]$, $\partial\nu/\partial\bar{\beta} = 0$, $\partial\nu/\partial\nu = 1$, and $\xi = \left(\frac{2\bar{\beta}^2}{(\nu-2)^2(\nu-4)} + \frac{1}{\nu-2}\right)^{-3/2}$. The Jacobian matrix, $\mathcal{J}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \partial\boldsymbol{\theta}^*/\partial\boldsymbol{\theta}$, is necessary for the derivation of the model extensions in Section 3. We also report the score of the GHSKT distribution with respect to the $\boldsymbol{\theta}^*$ parameterization, i.e., $\nabla_{\boldsymbol{\theta}^*}(y; \boldsymbol{\theta}^*) = \partial \log p(y; \boldsymbol{\theta}^*)/\partial\boldsymbol{\theta}^*$:

$$\frac{\partial \log p(y;\boldsymbol{\theta}^*)}{\partial\mu} = -\beta(\beta(y-\mu)\zeta + 1) + \frac{(\nu+1)(y-\mu)}{2(\delta^2+(y-\mu)^2)}, \tag{A.4}$$

$$\frac{\partial \log p(y;\boldsymbol{\theta}^*)}{\partial\delta} = \frac{\nu}{\delta} + \beta^2\delta\zeta - \frac{\delta(\nu+1)}{2(\delta^2+(y-\mu)^2)}, \tag{A.5}$$

$$\frac{\partial \log p(y;\boldsymbol{\theta}^*)}{\partial\beta} = \frac{\nu+1}{2\beta}\text{sgn}(\beta) + \zeta\beta(\delta^2+(y-\mu)^2) + (y-\mu), \tag{A.6}$$

$$\frac{\partial \log p(y;\boldsymbol{\theta}^*)}{\partial\nu} = \frac{\log 2}{2} + \log\delta + \frac{\log|\beta|}{2} + \frac{1}{2}\zeta - \frac{1}{2}\Psi\left(\frac{\nu}{2}\right) - \frac{\log(\delta^2+(y-\mu)^2)}{4}, \tag{A.7}$$

[15] See also Mencía and Sentana (2004) for similar arguments applied to the multivariate generalized hyperbolic case.

where sgn(·) denotes the sign function, $\Psi(\cdot)$ denotes the digamma function, and

$$\zeta = \frac{\dot{K}^{(\nu)}_{(\nu+1)/2}(z)}{K_{(\nu+1)/2}(z)z}, \tag{A.8}$$

with $\dot{K}^{(\nu)}_{(\nu+1)/2}(z)$ being the derivative of the modified Bessel function of the third kind evaluated in $z = \sqrt{\beta^2(\delta^2 + (y-\mu)^2)}$ with respect to its order.[16] It follows that the score with respect to the $\boldsymbol{\theta}$ parametrization, $\nabla_{\boldsymbol{\theta}}(y;\boldsymbol{\theta}) = \partial \log p(y;\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, is obtained as:

$$\nabla_{\boldsymbol{\theta}}(y;\boldsymbol{\theta}) = \mathcal{J}_{\boldsymbol{\theta}}(\boldsymbol{\theta})' \nabla^*_{\boldsymbol{\theta}}(y;\boldsymbol{\theta}^*). \tag{A.9}$$

The quantity $\nabla_{\boldsymbol{\theta}}(y;\boldsymbol{\theta})$ is used to model the higher-order moments of the cryptocurrencies' financial time series.

## Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2021.06.005.

## References

Aas, K., & Haff, I. H. (2006). The generalized hyperbolic skew Student's t-distribution. *Journal of Financial Econometrics*, *4*, 275–309.
Abramowitz, M., & Stegun, I. A. (1964). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables, Vol. 55*. Courier Corporation.
Alexander, C., & Dakos, M. (2020). A critical investigation of cryptocurrency data and analysis. *Quantitative Finance*, *20*, 173–188.
Alizadeh, S., Brandt, M. W., & Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance*, *57*, 1047–1091.
Andersen, T. G., Bollerslev, T., Christoffersen, P. F., & Diebold, F. X. (2006). Volatility and correlation forecasting. *Handbook of Economic Forecasting*, *1*, 777–878.
Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, *96*, 42–55.
Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariant matrix estimation. *Econometrica*, *59*, 817–858.
Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *74*, 3–30.
Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 401–419.
Bayer, S., & Dimitriadis, T. (2020). Regression-based expected shortfall backtesting. *Journal of Financial Econometrics*.
Bernardi, M., & Catania, L. (2016). Comparison of value-at-risk models using the MCS approach. *Computational Statistics*, *31*, 579–608.
Bianchi, D. (2020). Cryptocurrencies as an asset class? An empirical assessment. *Journal of Alternative Investments*, *23*, 162–179.
Black, F. (1976). Studies of stock price volatility changes. In *Proceedings of the 1976 meeting of the business and economic statistics section* (pp. 177–181). Washington, D.C.: American Statistical Association.
Blæsild, P. (1981). The two–dimensional hyperbolic distribution and related distributions, with an application to Johannsen's bean data. *Biometrika*, *68*, 251–263.
Bloomberg (2017a). Japan's BITPoint to Add Bitcoin Payments to Retail Outlets. https://www.bloomberg.com/news/articles/2017-05-29/japan-s-bitpoint-to-add-bitcoin-payments-to-100-000s-of-outlets.
Bloomberg (2017b). Some central banks are exploring the use of cryptocurrencies. https://www.bloomberg.com/news/articles/2017-06-28/rise-of-digital-coins-has-central-banks-considering-e-versions.
Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*, 307–327.
Brooks, C., Burke, S. P., Heravi, S., & Persand, G. (2005). Autoregressive conditional kurtosis. *Journal of Financial Econometrics*, *3*, 399–421.
Caporale, G. M., & Zekokh, T. (2019). Modelling volatility of cryptocurrencies using Markov-switching GARCH models. *Research in International Business and Finance*, *48*, 143–155.
Catania, L., Grassi, S., & Ravazzolo, F. (2018a). Forecasting cryptocurrencies under model and parameter instability. *International Journal of Forecasting*, *35*, 485–501.
Catania, L., Grassi, S., & Ravazzolo, F. (2018b). Predicting the volatility of cryptocurrency time-series. In *Mathematical and statistical methods for actuarial sciences and finance* (pp. 203–207). Springer.
Charles, A., & Darné, O. (2019). Volatility estimation for bitcoin: Replication and robustness. *International Economics*, *157*, 23–32.
Chicago Mercantile Exchange (2017). CME Group announces launch of Bitcoin futures. http://www.cmegroup.com/media-room/press-releases/2017/10/31/cme_group_announceslaunchofbitcoinfutures.html.
Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics*, *10*, 407–432.
Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, *39*, 841–862.
Chu, J., Chan, S., Nadarajah, S., & Osterrieder, J. (2017). GARCH modelling of cryptocurrencies. *Journal of Risk and Financial Management*, *10*, 1–15.
Chu, J., Nadarajah, S., & Chan, S. (2015). Statistical analysis of the exchange rate of bitcoin. *PLoS One*, *10*, 1–27.
Cointelegraph (2017). South Korea officially legalizes bitcoin, huge market for traders. https://cointelegraph.com/news/south-korea-officially-legalizes-bitcoin-huge-market-for-traders.
Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*, 174–196.
Crane, B. (1959). *the sophisticated investor: A guide to stock-market profits*. Simon and Schuster.
Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, *28*, 777–795.
Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *20*, 134–144.
Dyhrberg, A. H. (2016). Bitcoin, gold and the dollar – A GARCH volatility analysis. *Finance Research Letters*, *16*, 85–92.
Engle, R. F., & Lee, G. G. (1999). A permanent and transitory component model of stock return volatility. In R. F. Engle, & H. L. White (Eds.), *Cointegration, causality, and forecasting: A festschrift in honor of Clive W. J. Granger*. New York: Oxford University Press.
Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, *22*, 367–381.
Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, *48*, 1749–1778.
Fernández, C., & Steel, M. F. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, *93*, 359–371.
Fissler, T., & Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *The Annals of Statistics*, *44*, 1680–1707.
Forbes (2017). Emerging applications for blockchain. https://www.forbes.com/sites/forbestechcouncil/2017/07/18/emerging-applications-for-blockchain.
Fortune (2018). Goldman sachs is said to add cryptocurrency trading desk—Sort of. http://fortune.com/2018/05/02/goldman-sachs-cryptocurrency-trading-desk/.
Gandal, N., Hamrick, J., Moore, T., & Oberman, T. (2018). Price manipulation in the bitcoin ecosystem. *Journal of Monetary Economics*, *95*, 86–96.
Gao, C. T., & Zhou, X. H. (2016). Forecasting var and ES using dynamic conditional score models and skew student distribution. *Economic Modelling*, *53*, 216–223.
Ghalanos, A., Rossi, E., & Urga, G. (2015). Independent factor autoregressive conditional density model. *Econometric Reviews*, *34*, 594–616.
Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378.

[16] See Mencía and Sentana (2004) for an analytical derivation of this quantity

Gneiting, T., & Ranjan, R. (2011). Comparing density forecasts using threshold-and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, *29*, 411–422.

Griffin, J. M., & Shams, A. (2020). Is bitcoin really un-tethered?. *The Journal of Finance*, *75*, 1913–1964.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, *35*, 705–730.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: With applications to financial and economic time series, Vol. 52*. Cambridge University Press.

Harvey, A., & Lange, R. J. (2017). Volatility modeling with a generalized t distribution. *Journal of Time Series Analysis*, *38*, 175–190.

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

Hencic, A., & Gourieroux, C. (2014). Noncausal autoregressive model in application to bitcoin/usd exchange rate. In *Proceedings of the 7th financial risks international forum* (pp. 1–25).

Hotz-Behofsits, C., Huber, F., & Zorner, T. O. (2018). Predicting crypto-currencies using sparse non–Gaussian state space models. *J. Forecast.*, *37*, 627–640.

Huber, P. J. (2004). *Robust statistics, Vol. 523*. John Wiley & Sons.

Janus, P., Koopman, S. J., & Lucas, A. (2014). Long memory dynamics for multivariate dependence under heavy tails. *Journal of Empirical Finance*, *29*, 187–206.

Janus, P., Lucas, A., & Opschoor, A. (2018). New HEAVY models for fat-tailed returns and realized covariance kernels. *Journal of Business & Economic Statistics*, *36*, 643–657.

Katsiampa, P. (2017). Volatility estimation for bitcoin: A comparison of GARCH models. *Economics Letters*, *158*, 3–6.

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*, 33–50.

Lucas, A., Schwaab, B., & Zhang, X. (2014). Conditional euro area sovereign default risk. *Journal of Business & Economic Statistics*, *32*, 271–284.

Lucas, A., Schwaab, B., & Zhang, X. (2017). Modeling financial sector joint tail risk in the euro area. *Journal of Applied Econometrics*, *32*, 171–191.

Lucas, A., & Zhang, X. (2016). Score-driven exponentially weighted moving averages and value-at-risk forecasting. *International Journal of Forecasting*, *32*, 293–302.

Matheson, J. E., & Winkler, R. L. (1976). Scoring rules for continuous probability distributions. *Management Science*, *22*, 1087–1096.

McAleer, M., & Da Veiga, B. (2008). Single-index and portfolio models for forecasting value-at-risk thresholds. *Journal of Forecasting*, *27*, 217–235.

McNeil, A. J., & Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance*, *7*, 271–300.

Mencía, J., & Sentana, E. (2004). *Estimation and testing of dynamic models with generalized hyperbolic innovations*: *CEMFI working paper no. 0411*.

Modigliani, F., & Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment. *The American Economic Review*, *48*, 261–297.

Nakamoto, S. (2009). Bitcoin: A peer-to-peer electronic cash system. https://Bitcoin.org/Bitcoin.pdf.

Nolde, N., & Ziegel, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation. *The Annals of Applied Statistics*, *11*, 1833–1874.

Núñez, J. A., Contreras-Valdez, M. I., & Franco Ruiz, C. A. (2019). Statistical analysis of bitcoin during explosive behavior periods. *PLoS One*, *14*, 1–22.

Palma, W. (2007). *Long-memory time series: Theory and methods*. John Wiley & Sons.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, *160*, 246–256.

Patton, A. J., Ziegel, J. F., & Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics*, *211*, 388–413.

Sapuric, S., & Kokkinaki, A. (2014). Bitcoin is volatile! isn't that right?. In *Business information systems workshops* (pp. 255–265). Lecture Notes in Business Information Processing.

Schilling, L., & Uhlig, H. (2019). Some simple bitcoin economics. *Journal of Monetary Economics*, *106*, 16–26.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. New York, USA: Chapman and Hall, ISBN: 0412246201.

The Economist (2019). What is libra?. https://www.economist.com/the-economist-explains/2019/07/12/what-is-libra.

Tibshirani, R. J., & Efron, B. (1993). *Monographs on statistics and applied probability*, *An introduction to the bootstrap*. New York: Chapman and Hall.

Trottier, D. A., & Ardia, D. (2016). Moments of standardized Fernandez–steel skewed distributions: Applications to the estimation of GARCH-type models. *Finance Research Letters*, *18*, 311–316.

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