Forecasting Bitcoin risk measures: A robust approach

Carlos Trucios *

São Paulo School of Economics - FGV, Brazil.

June 2, 2018

Abstract

In the last few years, Bitcoin and other cryptocurrencies have attracted the interest of many investors, practitioners and researchers. However, little attention has been paid to the predictability of the risk measures of Bitcoin. In this paper, we compare the one-step-ahead volatility forecast of Bitcoin using several GARCH-type models and also evaluate the performance of several procedures when estimating the Value-at-Risk. We also take into account the presence of outliers and estimate the volatility and Value-at-Risk in a robust fashion. Our results show that robust procedures outperform the non-robust ones to forecast the volatility as well as to estimate the Value-at-Risk. These results suggest that the presence of outliers play an important role in modelling and forecasting Bitcoin risk measures.

Keywords: Cryptocurrency, GARCH, Model confidence set, Outliers, Realised volatility, Value-at-Risk.

JEL Classification: C51, C52, C53

2010 Mathematics Subject Classification: 62M10, 62F40, 62G35, 62P20, 91B84

*E-mail: ctrucios@gmail.com

The author acknowledges financial support from São Paulo Research Foundation (FAPESP) grant 2016/18599-4 and is also grateful by the support of the Centre of Quantitative Studies in Economics and Finance (CEQEF) and Center of Applied Research on Econometrics, Finance and Statistics (CAREFS). The author thanks helpful comments of Luiz K. Hotta and João H. G. Mazzeu.
1 Introduction

Since its creation in 2008, Bitcoin has attracted the interest of many investors, practitioners and researchers. This interest has grown quickly in the last years, probably due to the decentralised nature of Bitcoin and by its large profits; see Nakamoto (2008).

Previous studies, such as Sapuric and Kokkinaki (2014), Baek and Elbeck (2015) and Brière et al. (2015), have observed that Bitcoin is highly volatile, thus forecasting the volatility as well as estimating the Value-at-Risk (VaR) as better as possible is crucial for better decisions of investors and practitioners.

Bitcoin daily volatility has been previously studied by Dyhrberg (2016), Balcilar et al. (2017), Chu et al. (2017), Katsiampa (2017), Liu et al. (2017), Pichl and Kaizoji (2017), Naimy and Hayek (2018) and Catania et al. (2018) among others. However, most of the studies have been focused on the in-sample framework and comparisons have been made based on information criteria. In this context, Katsiampa (2017) estimates the volatility of Bitcoin using several GARCH-type models assuming Gaussian errors and concludes that the best model to estimate the volatility is the AR(1)-CGARCH(1,1), Chu et al. (2017) analyse the seven most popular cryptocurrencies using GARCH-type models with different error distributions and conclude that the best model to estimate the Bitcoin volatility is the IGARCH(1,1). Liu et al. (2017) compare the GARCH model assuming the normal reciprocal inverse Gaussian (NRIG) distribution against the Gaussian and Student-t error distributions and conclude that the GARCH model with Student-t errors estimates better the volatility. Charles and Darné (2018) replicate the study of Katsiampa (2017) and additionally take into account the presence of extreme observations. They find that, using the jump-filtered returns as in Laurent et al. (2016), the AR(1)-GARCH(1,1) model reports smaller information criteria than the models considered in Katsiampa (2017).

In an out-of-sample point of view, Naimy and Hayek (2018) compare the one-step-ahead volatility forecasts estimated by GARCH and EGARCH models with Gaussian, Student-t and generalised error distributions. The authors compare the predicted volatility with the realised volatility using the root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) and they conclude that EGARCH models present the best performance. Catania et al. (2018) compare the Gaussian GARCH model with
the generalised autoregressive score (GAS) models of Creal et al. (2013) and Harvey (2013). This comparison is based on the QLIKE measure and the predicted volatility is compared with the squared observed returns. They conclude that the GARCH model is outperformed by the GAS models. Peng et al. (2018) compare the volatility forecast obtained by GARCH, EGARCH and GJR models assuming symmetric and asymmetric Gaussian and Student-t errors against the Support Vector Regression GARCH model of Bezerra and Albuquerque (2017) and they find that the latter yields more accurate forecasts. Although out-of-sample comparisons are available in the literature, most of them are restrictive since they do not consider models which presented better performance in previous studies, and, additionally, they leave out several error distributions and models of the GARCH family. In this sense, a comprehensive out-of-sample comparison is needed.

On the other hand, most of the papers available in the literature do not consider the presence of outliers, which, as mentioned by, for instance, Carnero et al. (2012), Boudt et al. (2013) and Trucios and Hotta (2016), affect drastically the volatility forecast and VaR estimation. As far as we know, only Catania and Grassi (2017), Charles and Darné (2018) and Catania et al. (2018) take into account the presence of outliers to estimate the Bitcoin volatility, being the last one the only work in an out-of-sample context. However, Catania et al. (2018) only consider the Gaussian GARCH model against the GAS models.

In a VaR context, Chu et al. (2017), Chan et al. (2017), Osterrieder and Lorenz (2017), Stavroyiannis (2018) and Gkillas and Katsiampa (2018) estimate the VaR of Bitcoin. However, none of these works consider the presence of outliers in a context of conditional heteroscedastic.

The contribution of this paper is threefold. First, we carry out an extensive out-of-sample comparison of the Bitcoin volatility forecast using several GARCH-type models with different error distributions, filling, then, a gap in the literature and also summarising the GARCH-type results found in the previous works. Second, we address the presence of outliers and show that the better volatility forecast is obtained using a robust procedure, suggesting that outliers play a crucial role when modelling and forecasting the volatility of Bitcoin. Finally, we evaluate the performance of the VaR estimation using of the best non-robust procedures to forecast the volatility, chosen by the Model Confidence Set (MCS), and
the robust procedure proposed by Trucíos et al. (2017) showing that non-robust procedures underestimate the VaR while the robust procedure estimates the VaR with good results. These results are useful for academics, as an important reference for future research, as well as for practitioners, providing information about how to better measure the risk.

The rest of the paper is organized as follows: Section 2 describes the models, realised measures and loss functions used in the out-of-sample comparison. Section 3 describes the data and reports the main findings. Finally, Section 4 presents the conclusions and future researches.

2 Methodology

In this section, we briefly describe the GARCH-type models used to forecast the volatility of Bitcoin as well as the realised volatility measures used as volatility proxies. At the end of this section, we also describe the loss functions used to evaluate the out-of-sample performance and the robust bootstrap procedure of Trucíos et al. (2017) to estimate the VaR.

2.1 GARCH models

The GARCH model is commonly used for modelling and forecasting the second-order moments of return in economic and financial time series. Since its introduction by Bollerslev (1986), several extensions have been proposed in the literature. These extensions differ to each other in how the volatility equation is defined.

Let \( r_t \) be the observed returns at time \( t \) and \( \epsilon_t \) the error term follows a white noise process. The GARCH(1,1) model is defined by

\[
\begin{align*}
    r_t &= \sigma_t \epsilon_t, \\
    \sigma_t^2 &= \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2,
\end{align*}
\]

with \( \sigma_t^2 \) being the conditional variance (or squared volatility) at time \( t \) and \( \omega, \alpha \) and \( \beta \) parameters satisfying some stationary conditions. Sufficient conditions for stationarity are given by \( \omega > 0, \alpha, \beta \geq 0 \) and \( \alpha + \beta < 1 \).

As mentioned above, different extensions of the GARCH model involve different spec-
ifications of the volatility equation ($\sigma_t^2$). Table 1 describe the volatility equations of the GARCH-type models used in the comparison. For good reviews of univariate GARCH-type models, see, for instance, Teräsvirta (2009) and Rodríguez and Ruiz (2012).

In all non-robust cases, several error distributions are assumed, namely, Normal, Skew Normal, Student-t, Skew Student-t, GED, Skew GED, Normal Inverse Gaussian, Generalized Hyperbolic and the Johnson’s reparametrized SU innovation distribution; see Ghalanos (2018) for details of the parametrization distributions and GARCH-type models used in this paper.

Table 1: GARCH-type models

<table>
<thead>
<tr>
<th>Model</th>
<th>Volatility equation</th>
<th>Proposed by</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH</td>
<td>$\sigma_{t+1}^2 = \omega + \alpha r_t^2 + \beta \sigma_t^2$</td>
<td>Bollerslev (1986)</td>
</tr>
<tr>
<td>IGARCH</td>
<td>$\sigma_{t+1}^2 = \omega + \alpha r_t^2 + (1 - \alpha) \sigma_t^2$</td>
<td>Engle and Bollerslev (1986)</td>
</tr>
<tr>
<td>EGARCH</td>
<td>$\log(\sigma_{t+1}^2) = \omega + \alpha z_t^2 + \gamma (</td>
<td>z_t</td>
</tr>
<tr>
<td>GJR</td>
<td>$\sigma_{t+1}^2 = \omega + \alpha r_t^2 + \gamma I(r_t &lt; 0) r_t^2 + \beta \sigma_t^2$</td>
<td>Glosten et al. (1993)</td>
</tr>
<tr>
<td>APARCH</td>
<td>$\sigma_{t+1}^2 = \omega + \alpha (</td>
<td>r_t</td>
</tr>
<tr>
<td>CGARCH</td>
<td>$\sigma_{t+1}^2 = \omega + \alpha (r_t^2 - q_t) + \beta (\sigma_t^2 - q_t)$</td>
<td>Lee and Engle (1999)</td>
</tr>
<tr>
<td>TGARCH</td>
<td>$\sigma_{t+1}^2 = \omega + \alpha \sigma_t (</td>
<td>z_t</td>
</tr>
<tr>
<td>AVGARCH</td>
<td>$\sigma_{t+1}^2 = \omega + \alpha \sigma_t (</td>
<td>z_t - \eta_2</td>
</tr>
<tr>
<td>NGARCH</td>
<td>$\sigma_{t+1}^2 = \omega + \alpha \sigma_t^2 (</td>
<td>z_t - \eta_2</td>
</tr>
<tr>
<td>NAGARCH</td>
<td>$\sigma_{t+1}^2 = \omega + \alpha \sigma_t^2 (</td>
<td>z_t - \eta_2</td>
</tr>
<tr>
<td>FGARCH</td>
<td>$\sigma_{t+1}^2 = \omega + \alpha \sigma_t^2 (</td>
<td>z_t - \eta_2</td>
</tr>
<tr>
<td>Robust</td>
<td>$\sigma_{t+1}^2 = \omega + \gamma_c \alpha \rho \left( \frac{r_t^2}{\sigma_t^2} \right) + \beta \sigma_t^2$</td>
<td>Boudt et al. (2013)</td>
</tr>
<tr>
<td>GARCH</td>
<td>with $\rho(x) = \begin{cases} 1, &amp; \text{if } x &gt; c, \ x, &amp; \text{if } x \leq c, \end{cases}$ and $\gamma_c$ being a constant to ensure consistency</td>
<td></td>
</tr>
</tbody>
</table>

2.2 Realised measures

Evaluate the predictive ability of different approaches to forecast the volatility is challenging since the volatility is a latent variable and consequently is not directly observable. A good proxy for the conditional variance is the realised variance (Andersen et al., 2003) which use intra-day data. In this paper we use the realised variance - RV and some alternative realised measures which are robust to microstructure noise. Specifically, we use the Bipower

---

1In Table 1, $z_t$ corresponds to the devolatilised return $z_t = \frac{r_t}{\sigma_t}$. 

---
variation - BV (Barndorff-Nielsen and Shephard, 2004), MinRV (Andersen et al., 2012) and MedRV (Andersen et al., 2012) as proxies of the true conditional variance. We prefer to use realised measures instead of square observed daily returns as in Catania et al. (2018) because realised measures have showed to be a better proxy (Alizadeh et al., 2002; McAleer and Medeiros, 2008; Patton, 2011) and are most widely used nowadays. Table 2 presents the realised measures used in this paper.

Table 2: Realised measures.

<table>
<thead>
<tr>
<th>Realised measure</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>RV ( \sum_{i=1}^{N} r_i^2 )</td>
<td>BV ( \frac{\pi}{2} \left( \frac{N}{N-1} \right) \sum_{i=1}^{N-1}</td>
</tr>
<tr>
<td>MinRV ( \frac{\pi}{\pi - 2} \left( \frac{N}{N-1} \right) \sum_{i=1}^{N-1} \min(</td>
<td>r_i</td>
</tr>
</tbody>
</table>

For more details about realised measures as well as for asymptotic properties see McAleer and Medeiros (2008), Barndorff-Nielsen and Shephard (2004) and Andersen et al. (2012).

2.3 Robust loss function

The evaluation of the volatility forecast will be made using volatility proxies such as described in the previous section. However, these proxies are imperfect since they are estimates of the integrated variance. To avoid that imperfect volatility proxies lead to misleading results in the predictability comparison of the volatility forecast, Patton (2011) propose the use of robust loss functions, these functions are robust to the microstructure noise in the volatility proxy. The general class of the robust loss function is defined by Patton (2011) and given by

\[
L(\hat{\sigma}^2, h, b) = \begin{cases} 
  h - \hat{\sigma}^2 + \hat{\sigma}^2 \log(\frac{\hat{\sigma}^2}{h}), & \text{for } b = -1, \\
  \frac{\hat{\sigma}^2}{h} - \log(\frac{\hat{\sigma}^2}{h}) - 1, & \text{for } b = -2, \\
  \frac{(\hat{\sigma}^{2b+4} - h^{2b+4})}{(b+1)(b+2)} - \frac{h^{b+1}(\hat{\sigma}^2 - h)}{b+1}, & \text{otherwise},
\end{cases}
\]
where $\hat{\sigma}^2$ is the squared forecasts volatility and $h$ the squared volatility proxy. In this paper we consider the robust loss function of Patton (2011) with three different values of $b$: $b = -2$ (QLIKE), $b = -1$ (MSE) and $b = 0$ (hereafter denoted by RLF).

2.4 VaR estimation

Assuming that returns are zero mean, the one-step-ahead VaR in the non-robust models is estimated as usual by $\alpha\%VaR = Q_{\alpha\%}\hat{\sigma}_{T+1|T}$ where $Q_{\alpha\%}$ is the $\alpha\%$ quantile of the assumed error distribution (scaled to have unit variance) and $\hat{\sigma}_{T+1|T}$ is the one-step-ahead volatility forecast.

To estimate the VaR in a robust way, we use the robust bootstrap procedure recently proposed by Trucios et al. (2017). The procedure is based on a residual-based bootstrap scheme combining with a robust estimator and robust filters for the volatility. The procedure can be summarized in the following steps:

- **Step 1:** Estimate the parameters $\omega$, $\alpha$ and $\beta$ in a robust way and obtain the standardized residuals $\hat{e}_t = \frac{r_t}{\hat{\sigma}_t}$. Denote by $\hat{F}_e$ the empirical distribution of these centred standardized residuals.

- **Step 2:** Using $\epsilon^*_{t}$ (bootstrap extractions from $\hat{F}_e$) generate bootstrap series through the following recursion.

$$r^*_t = \sigma^*_t \epsilon^*_t,$$
$$\sigma^2_{t+1} = \hat{\omega} + \hat{\alpha} \sigma^2_{t} c r_c \left( \frac{r^2_{t+1}}{\sigma^2_{t}} \right) + \hat{\beta} \sigma^2_{t},$$

where $\sigma^2_{t} = \sigma^2_{t}$ and the filter $r_c(\cdot)$ is defined in Table 3. Using the bootstrap series estimate the parameters $\hat{\omega}^*$, $\hat{\alpha}^*$ and $\hat{\beta}^*$ using the same estimator in Step 1.

- **Step 3:** Obtain $h$-steps-ahead forecast for returns as

$$\hat{r}^*_{T+h|T} = \epsilon^*_{T+h} \hat{\sigma}^*_{T+h|T},$$
$$\hat{\sigma}^2_{T+h|T} = \hat{\omega}^* + \hat{\alpha}^* \hat{\sigma}^2_{T+h|T} c r_c \left( \frac{r^2_{T+h|T}}{\hat{\sigma}^2_{T+h|T}} \right) + \hat{\beta}^* \hat{\sigma}^2_{T+h|T},$$

for $h = 1, ..., H$, and where $\hat{r}^*_{T|T} = r_T$, $\epsilon^*_{T+h}$ are bootstrap extractions from $\hat{F}_e$ and
\( \hat{\sigma}^2_{t|T} \) is obtained through the recursion
\[
\hat{\sigma}^2_{t|T} = \hat{\omega} + \hat{\alpha} \hat{\sigma}^2_{t-1|T} + \hat{\beta} \hat{\sigma}^2_{t-1|T} c T_r c(x) \left( \frac{r^2_{t-1|T}}{\hat{\sigma}^2_{t-1|T}} \right) + \hat{\beta} \hat{\sigma}^2_{t-1|T}
\]
for \( t = 2, ..., T \), with \( \hat{\sigma}^2_{1|T} = \sigma^2_1 \), \( r_c(x) \) equal to \( x \) if \( x \leq c \) and \( \varepsilon^2_t \) (squared bootstrap extractions from \( \hat{\epsilon}_t \)) otherwise and \( c \) being a cut-off value defined a priori.

**Step 4:** Repeat steps 2 and 3 \( B \) times to obtain \( B \) bootstrap replicates \( (\hat{r}^*_T, ..., \hat{r}^*_B T) \), the \( \alpha \% \) VaR is estimated as the \( \alpha \% \) empirical quantile of the bootstrap replicates.

### 3 Data and results

We use daily Bitcoin closing prices (in US dollar) traded on Bitstamp from September 13, 2011 to December 31, 2017 (2280 observations). Tick-by-tick data were obtained from [bitcoincharts](https://api.bitcoincharts.com/v1/csv/) and the daily closing prices were constructed from the tick-by-tick data as the last price traded at each day.

Returns are calculated as \( r_t = \log(P_t) - \log(P_{t-1}) \) with \( P_t \) being the closing price at day \( t \). Table 3 reports descriptive statistics and Figure 1 shows the daily returns as well as the autocorrelation of returns and squared returns. We can observe that Bitcoin is highly volatile with an annualized standard deviation of 0.8350 \((0.0526 \times \sqrt{252})\), returns also present asymmetry and large Kurtosis. The large Kurtosis is probably explained by the presence of extreme returns, as observed in Figure 1. Because the returns series does not exhibit serial correlation no ARMA filter is applied to the data, the series is only centred to have zero mean.

<table>
<thead>
<tr>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Q1</th>
<th>Med.</th>
<th>Q3</th>
<th>Max</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0034</td>
<td>0.0526</td>
<td>-0.6639</td>
<td>-0.0111</td>
<td>0.0024</td>
<td>0.0200</td>
<td>0.4455</td>
<td>-1.4011</td>
<td>28.4802</td>
</tr>
</tbody>
</table>

To evaluate the out-of-sample performance we use a rolling windows scheme with windows size equal to 1000 days. In each window the one-step-ahead conditional variance is estimated. The one-step-ahead conditional variance is compared with the four realised measures described in Section 2.2 using five minutes high-frequency data (results using 10 minutes high-frequency data were also evaluated and the conclusions are similar). The non-robust GARCH-type models were estimated using the R package rugarch of Ghalanos (2017)

---

2[https://api.bitcoincharts.com/v1/csv/](https://api.bitcoincharts.com/v1/csv/)
and the realised measures were computed using the R package `highfrequency` of Boudt et al. (2017).

Differently of Charles and Darné (2018) which takes into account the presence of outliers using jump-filtered returns as in (Laurent et al., 2016), we estimate the conditional variance in a robust way using the robust estimator of Boudt et al. (2013) with the modification introduced in Trucios et al. (2017).

Table 4 reports the MSE, QLIKE and RLF between $\hat{\sigma}^2_{T+1}$ and the respective realised measure (RV, MinRV, MedRV, BP). The shadowed cells are the set of models with best out-of-sample performance obtained using the MCS approach (Hansen et al., 2011) at 75% significant level.\(^3\) In bold, the best model (first position in the MCS rank) in each case.

The MSE selects a large set of models reflecting a low power to distinguish between different models, indeed using the MSE just 15 of 100 models were left out (16 when considering the MinRV as a volatility proxy). Other works have also found that the MSE has a low power to distinguish between different models, see for instance Patton and Sheppard (2009) and Liu et al. (2015). On the other hand, observe that the QLIKE and RLF loss functions are better to choose a small set of models with the best performance.

In general, note that the MCS methodology select the same models by loss function regardless the realised measure used. An exception is observed when the QLIKE measure

\(^3\)We have used the R package MCS of Bernardi (2017)
is used in which case the CGARCH model belong to the group of models with the best performance only when the realised variance is used.

Considering all loss functions and realised measures only five models are in the set of better models in all cases, namely, TARCH and AVGARCH assuming generalized error distribution (standard and skew version) and the robust GARCH model. The CGARCH model which was the best model in Katsiampa (2017) appear in the set when using the MSE and RLF loss functions and also using the QLIKE loss function and the realised variance as volatility proxy. The IGARCH model which was the best model in the in-sample analyse in Chu et al. (2017) only appear in the set when considering the MSE loss functions.

In all cases, the GARCH model estimated in a robust way reports the best results regardless the loss function and realised measure used. This result is extremely important since most of analyses available in the literature about Bitcoin do not take into account the presence of outliers and, as we can see in Table 3, results can be substantially improved using a robust approach. A simple model such as the GARCH model estimated considering the presence of outliers outperforms more sophisticated models such as TARCH or AVGARCH. These results are in concordance with Charles and Darné (2018) which also find, in an in-sample context, that sophisticated models are outperformed by a GARCH model when the presence of outliers is considered in the analyse.

The 1% VaR is also computed and backtesting procedures for the VaR are carried out. We compute the VaR for the models chosen by the MCS as the models with the best performance in all criteria. Results are shown in Table 5 and report the proportion of fails (returns smaller than the 1% VaR), \( p \)-values of the unconditional coverage - UC (Kupiec, 1995), conditional coverage - CC (Christoffersen, 1998), dynamic quantile -DQ (Engle and Manganelli, 2004) tests as well as the average quantile loss function of González-Rivera et al. (2004).

Results reveal that the non-robust procedures underestimate the VaR and all backtesting procedures reject the null hypothesis that the proportion of fails is equal to 0.01. Using the robust procedure of Trucios et al. (2017) all tests fail to reject the null hypothesis. Also, observe that the smallest average quantile loss function of González-Rivera et al. (2004) is obtained when the VaR is estimated in a robust way. Figure 2 reports the returns of the
<table>
<thead>
<tr>
<th>Distrbution</th>
<th>INNOV MSE ($\times 10^{-3}$)</th>
<th>QLIKE $\sigma^2$ ($\times 10^{-3}$)</th>
<th>RFL $\sigma^2$ ($\times 10^{-3}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RV</td>
<td>MinRV</td>
<td>MedRV</td>
</tr>
<tr>
<td>norm</td>
<td>0.620</td>
<td>0.6879</td>
<td>0.6755</td>
</tr>
<tr>
<td>std</td>
<td>0.630</td>
<td>0.6750</td>
<td>0.6656</td>
</tr>
<tr>
<td>gexp</td>
<td>0.622</td>
<td>0.6663</td>
<td>0.6541</td>
</tr>
<tr>
<td>sqed</td>
<td>0.619</td>
<td>0.6701</td>
<td>0.6598</td>
</tr>
<tr>
<td>sigma</td>
<td>0.629</td>
<td>0.6745</td>
<td>0.6622</td>
</tr>
<tr>
<td>ghyp</td>
<td>0.630</td>
<td>0.6748</td>
<td>0.6624</td>
</tr>
<tr>
<td></td>
<td>0.684</td>
<td>0.7331</td>
<td>0.7202</td>
</tr>
<tr>
<td></td>
<td>0.677</td>
<td>0.7255</td>
<td>0.7126</td>
</tr>
<tr>
<td>std</td>
<td>0.634</td>
<td>0.6799</td>
<td>0.6675</td>
</tr>
<tr>
<td>sqed</td>
<td>0.634</td>
<td>0.6770</td>
<td>0.6647</td>
</tr>
<tr>
<td>sigma</td>
<td>0.634</td>
<td>0.6770</td>
<td>0.6647</td>
</tr>
<tr>
<td>ghyp</td>
<td>0.634</td>
<td>0.6770</td>
<td>0.6647</td>
</tr>
<tr>
<td></td>
<td>0.669</td>
<td>0.7378</td>
<td>0.7205</td>
</tr>
<tr>
<td></td>
<td>0.669</td>
<td>0.7378</td>
<td>0.7205</td>
</tr>
<tr>
<td>sqed</td>
<td>0.634</td>
<td>0.6770</td>
<td>0.6647</td>
</tr>
<tr>
<td>sigma</td>
<td>0.634</td>
<td>0.6770</td>
<td>0.6647</td>
</tr>
<tr>
<td>ghyp</td>
<td>0.634</td>
<td>0.6770</td>
<td>0.6647</td>
</tr>
</tbody>
</table>

Table 4: Average MSE, QLIKE and RHF between $\hat{\sigma}_{T+1}^2$ and the realised measures.
out-of-sample period and the 1% VaR estimated in a robust way.

In general, our results show that the risk measures are better estimated when using robust procedures.

Table 5: Proportion of returns smaller than the 1% VaR, p-values of the UC, CC, DQ test and average quantile loss function (AQLF) (×10^3).

<table>
<thead>
<tr>
<th>Models</th>
<th>Prop. Fails</th>
<th>UC</th>
<th>CC</th>
<th>DQ</th>
<th>AQLF</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGARCH GED</td>
<td>0.1039</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>2.9755</td>
</tr>
<tr>
<td>TGARCH SGED</td>
<td>0.1039</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>2.9805</td>
</tr>
<tr>
<td>AVGARCH GED</td>
<td>0.1023</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>2.9776</td>
</tr>
<tr>
<td>AVGARCH SGED</td>
<td>0.1023</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>2.9710</td>
</tr>
<tr>
<td>Robust</td>
<td>0.0063</td>
<td>0.1475</td>
<td>0.3331</td>
<td>0.6602</td>
<td>1.4283</td>
</tr>
</tbody>
</table>

Figure 2: Return in the out-of-sample period (black solid line) and the estimated 1% VaR (red dashed line) obtained using the robust bootstrap procedure of Trucios et al. (2017).

4 Conclusions and future works

In this paper, we have made a comprehensive out-of-sample comparison in a context of the daily volatility forecast of Bitcoin using GARCH-type models with different error distributions. Additionally, we included a robust GARCH procedure and compare it with the non-robust models. Our results reveal that better results are obtained when the presence of outliers is not neglected and estimation of the volatility is made in a robust way.

Among the non-robust procedures, the models with better out-of-sample performance to forecast the volatility are the TARCH and AVGARCH models both considering generalized error distribution (standard and skew version). However, these models are outperformed by the GARCH model when estimated in a robust fashion. In particular, we use the robust estimator of Boudt et al. (2013) with the modification used in Trucios et al. (2017).

The VaR estimated in a non-robust way reports large values of the proportion of fails and all backtesting procedures reject the null hypothesis of that the proportion of fails is
equal to 0.01 giving a misleading picture of what can be expected in the future. On the other hand, the VaR estimated using the robust bootstrap procedure of Trucios et al. (2017) shown a good performance.

Additional comparisons considering different GARCH-type models estimated in a robust way as well as considering switch regime are interesting research topics. In the same spirit, to compare robust GARCH-type models with GAS models as the used in Catania et al. (2018) and some alternative procedures as the used in Peng et al. (2018) as well as analyse the performance of risk measures in a high-frequency context are interesting research topics.

Some papers such as Dyhrberg (2016), Balcilar et al. (2017) and Cermak (2017) have also consider explanatory variables to better estimate the volatility. Thus, one step further in this research is to analyse the performance of the out-of-sample volatility forecast considering explanatory variables and robust GARCH procedures.

This paper is in concordance with the results of Carnero et al. (2012), Trucios and Hotta (2016) and Trucios et al. (2017) which show the dramatic effect of additive outliers in the estimation and prediction of the volatility and VaR. For a good review about outliers in GARCH models, we refer to Hotta and Trucios (2018).

In a multivariate framework Boudt et al. (2013), Grané et al. (2014) and Trucios et al. (2018b) shown the dramatic effect of outliers in the estimation and prediction of the volatilities and co-volatilities. In this sense, it is important that future studies considering jointly Bitcoin with other cryptocurrencies take into account the presence of outliers. In this sense, the procedures proposed by, for instance, Croux et al. (2010), Boudt and Croux (2010), Boudt et al. (2013), Iqbal (2013), Trucios et al. (2018a) and Trucios et al. (2018a) could be useful.

References


