proceedings

MDPI

Proceedings

Using Entropy to Forecast Bitcoin's Daily Conditional Value at Risk [†]

Hellinton H. Takada [1,*], Sylvio X. Azevedo [2], Julio M. Stern [2] and Celma O. Ribeiro [1]

1 Polytechnic School, University of São Paulo, São Paulo 05508-010, Brazil; celma@usp.br
2 Institute of Mathematics and Statistics, University of São Paulo, São Paulo 05508-090, Brazil; sylvioazevedo@gmail.com (S.X.A.); jstern@ime.usp.br (J.M.S.)
* Correspondence: hellinton@gmail.com

† Presented at the 39th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Garching, Germany, 30 June–5 July 2019.

Published: 21 November 2019

Abstract: Conditional value at risk (CVaR), or expected shortfall, is a risk measure for investments according to Rockafellar and Uryasev. Yamai and Yoshiba define CVaR as the conditional expectation of loss given that the loss is beyond the value at risk (VaR) level. The VaR is a risk measure that represents how much an investment might lose during usual market conditions with a given probability in a time interval. In particular, Rockafellar and Uryasev show that CVaR is superior to VaR in applications related to investment portfolio optimization. On the other hand, the Shannon entropy has been used as an uncertainty measure in investments and, in particular, to forecast the Bitcoin's daily VaR. In this paper, we estimate the entropy of intraday distribution of Bitcoin's logreturns through the symbolic time series analysis (STSA) and we forecast Bitcoin's daily CVaR using the estimated entropy. We find that the entropy is positively correlated to the likelihood of extreme values of Bitcoin's daily logreturns using a logistic regression model based on CVaR and the use of entropy to forecast the Bitcoin's daily CVaR of the next day performs better than the naive use of the historical CVaR.

Keywords: entropy; conditional value at risk; cryptocurrency

1. Introduction

In finance, risk management is the activity of identifying, analyzing, estimating and controlling the risk of losing money. For our purposes, risk management is a procedure for shaping a loss distribution of an investment. The value at risk (VaR) is the most popular risk measure and it represents how much an investment might lose during usual market conditions with a given probability in a time interval. In other words, VaR is a percentile of a loss distribution. Another very popular risk measure is the conditional value at risk (CVaR), or the expected shortfall. CVaR is a risk measure for investments reintroduced in the literature by Rockafellar and Uryasev [1], for a former reference see Love et al. [2]. According to Sarykalin et al. [3], it approximately (or exactly, under certain conditions) equals the average of some percentage of the worst-case loss scenarios.

Relative to the definitions, there is a near correspondence between VaR and CVaR. For instance, Yamai and Yoshiba [4] defined CVaR as the conditional expectation of loss given that the loss is beyond the VaR level. Consequently, considering the same confidence level, VaR is a lower bound for CVaR. In particular, Rockafellar and Uryasev [1,5] showed that CVaR is superior to VaR in applications related to investment portfolio optimization. In practice, the choice between VaR and CVaR rests on the differences in mathematical properties, stability of statistical estimation, simplicity of optimization procedures, acceptance by regulators, and so on [3]. For instance, in terms of mathematical properties, the CVaR of a portfolio is a continuous and convex function with respect to positions in instruments, whereas the VaR may be even a discontinuous function.

The volatility is the standard deviation of the distribution of logreturns and a very simple and earlier measure of financial risk. The corresponding variance is a natural measure of the statistical uncertainty but it just captures a small portion of the informational content of the distribution of the logreturns. On the other hand, the entropy is a more general measure of uncertainty than the variance because it may be related to higher-order moments of a distribution [6–8]. According to Dionisio et al. [8], the variance measures the concentration around the mean while the entropy measures the dispersion of the density irrespective of the location of the concentration. Finally, for Pele et al. [9], the entropy of a distribution function is strongly related to its tails and this feature is more important for distributions with heavy tails or with an infinite second-order moment for which the variance does not make sense.

In the literature, there are empirical papers showing that entropy has good predictive power for risk. For instance, Billio et al. [10] showed that entropy has the ability to forecast and predict banking crises using directly the entropy of systemic risk measures. In addition, Pele et al. [9] showed that entropy of the intraday distribution of logreturns is a strong predictor of daily VaR, performing better than the classical GARCH models, for a time series of EUR/JPY exchange rates. Similarly, Pele and Mazurencu-Marinescu-Pele [11], instead of using the entropy of the intraday distribution of logreturns, defined the entropy using symbolic time series analysis (STSA) showing that their entropy is a strong predictor of daily VaR, performing better than the classical GARCH models, using high-frequency data for Bitcoin.

There is a recent interest in the statistical properties and risk behavior of cryptocurrencies [12–14] and, in particular, Bitcoin [15]. Consequently, in this paper, we estimate the entropy of the symbolic intraday distribution of Bitcoin's logreturns through the STSA [11] and we model and forecast the Bitcoin's daily CVaR using the estimated entropy. The main contribution of this paper is the extension of the study performed by Pele and Mazurencu-Marinescu-Pele [11] to include the CVaR. The rest of the paper is organized as follows: in Section 2, we present the details of the methodology; in Section 3, we present our empirical study describing the dataset, the results and the corresponding comments; finally, in Section 4, we conclude the paper.

2. Methodology

In this section, we review the methodology to estimate the entropy of the symbolic intraday distribution of logreturns through the STSA, a logistic model connecting the daily VaR and the entropy, and a forecasting model for the daily VaR using the entropy based on a quantile regression published by Pele and Mazurencu-Marinescu-Pele [11]. In addition, we introduce the two main contributions of this paper: a logistic model connecting the daily CVaR and the entropy, and a forecasting model for the daily CVaR using the entropy based on a modified quantile regression model. It is also important to mention that the Bitcoin exchange rate is hereinafter referred to as Bitcoin price.

2.1. Entropy of Symbolic Intraday Logreturns

In the intraday context, it is usual to consider a set of days $d \in \{1, \ldots, D\}$ and each day equally partitioned in M time bins. Consequently, for a day d and a time bin $m \in \{1, \ldots, M\}$, we associate a price $P_{d,m}$ and a logprice $p_{d,m} = \ln P_{d,m}$. Then, the intraday logreturn of an asset is defined as follows:

$$r_{d,m} = p_{d,m} - p_{d,m-1}; d = 1, \ldots, D; m = 2, \ldots M. \tag{1}$$

For the empirical study of this paper, it is possible to define $r_{d,1} = p_{d,1} - p_{d-1,M}; d = 2, \ldots, D$ because the Bitcoin is continuously traded. However, it is important to point that for other kind of assets, it would be better to ignore the logreturn $r_{d,1}$. In addition, $r_{1,1}$ is not defined.

The intraday logreturns is usually very noisy. The idea behind the STSA technique [16] to produce low-resolution data from high-resolution data. In particular, STSA is a transformation of a real number sequence to a binary sequence. In our case, the STSA transformation is applied to the intraday logreturn to obtain the symbolic intraday logreturn. The symbolic intraday logreturn is defined as follows:

$$s_{d,m} = \begin{cases} 1, & r_{d,m} \leq 0 \\ 0, & r_{d,m} > 0 \end{cases}. \tag{2}$$

Basically, the symbolic intraday logreturn is a binary sequence of 0s representing increasing prices and 1s representing decreasing prices.

Based on the Shannon entropy definition [17], the entropy of the symbolic intraday logreturns is defined as follows:

$$h_d = -\pi_d \log_2 \pi_d - (1 - \pi_d) \log_2 (1 - \pi_d), \tag{3}$$

where $\pi_d = \Pr(s_{d,m} = 1)$ and $1 - \pi_d = \Pr(s_{d,m} = 0)$. It is possible to notice that the entropy of the symbolic intraday logreturns is a daily entropy. In addition, we estimate $\pi_d, d = 2, \ldots, D$ using the sample frequency $\sum_{m=1}^{M} s_{d,m} / M, d = 2, \ldots, D$ and π_1 using the sample frequency $\sum_{m=2}^{M} s_{1,m} / (M - 1)$.

2.2. Entropy and Daily VaR and CVaR

Intuitively, the entropy of the symbolic intraday logreturns is higher at the presence of higher uncertainty in the returns and lower at the presence of lower uncertainty in the returns. Consequently, the likelihood of extreme negative daily logreturns is explained by higher values of entropy. In [11], it was verified that the entropy is positively correlated to the likelihood of extreme negative daily logreturns and the relation between VaR and entropy was modeled using the following logistic regression model:

$$\Pr(y_d = 1) = \frac{e^{b_0 + b_1 h_d}}{1 + e^{b_0 + b_1 h_d}}, \tag{4}$$

where b_0 and b_1 are constants to be estimated;

$$y_d = \begin{cases} 1, & r_d \leq -\mathrm{VaR}_\alpha \\ 0, & r_d > -\mathrm{VaR}_\alpha \end{cases}, d = 2, \ldots, D \tag{5}$$

are the indicators of the lower tails of the daily logreturns; $r_d = \ln P_d - \ln P_{d-1}, d = 2, \ldots, D$ are the daily logreturns; P_d is the closing price of day d; and VaR_α is the daily value at risk at the significance level $\alpha \in \,]0, 1[$ defined by

$$\Pr(r_d \leq -\mathrm{VaR}_\alpha) = \alpha \tag{6}$$

or, alternatively,

$$\mathrm{VaR}_\alpha = -\inf\{z | F(z) \geq \alpha\}, \tag{7}$$

where $F(\cdot)$ is the cumulative distribution function of the daily logreturns.

In this paper, the hypothesis is also that the entropy is positively correlated to the likelihood of extreme negative daily logreturns and we model the relation between CVaR and entropy using the following logistic regression model:

$$\Pr\left(u_d = 1\right) = \frac{e^{c_0 + c_1 h_d}}{1 + e^{c_0 + c_1 h_d}}, \tag{8}$$

where c_0 and c_1 are constants to be estimated;

$$u_d = \begin{cases} 1, & r_d \leq -\mathrm{CVaR}_\alpha \\ 0, & r_d > -\mathrm{CVaR}_\alpha \end{cases}, d = 2, \ldots, D \tag{9}$$

are the indicators of the lower tails of the daily logreturns; and CVaR_α is the daily conditional value at risk at the significance level $\alpha \in]0, 1[$ defined by

$$\mathrm{CVaR}_\alpha = -\frac{1}{\alpha} \int_{-\infty}^{-\mathrm{VaR}_\alpha} z f\left(z\right) dz, \tag{10}$$

where $f\left(\cdot\right)$ is the continuous probability density function of the daily logreturns.

2.3. Forecasting Model for Daily VaR and CVaR

Pele et al. 2017 and Pele et al. 2019 [9,11] considered a quantile regression model to forecast the daily VaR using the entropy as the explanatory variable. The forecasting model for the daily VaR_α at day $k + w + 1$ using the entropy of the day $k + w$ is given by:

$$\hat{\mathrm{VaR}}_{\alpha, k+w+1} = -\hat{b}_0^k - \hat{b}_1^k h_{k+w}, \tag{11}$$

where $\hat{b}_0^k$ and $\hat{b}_1^k$ are estimated using a quantile regression model between the dependent variable r_d and the independent variable h_{d-1} for $d \in \mathcal{W}_w\left(k\right)$;

$$\mathcal{W}_w\left(k\right) = \begin{cases} \left\{2, \ldots, w\right\}, & k = 0 \\ \left\{k+1, \ldots, k+w\right\}, & k = 1, 2, \ldots \end{cases}. \tag{12}$$

Based on Koenker and Bassett [18], we consider the following optimization problem for the quantile regression estimation:

$$\left\{\hat{b}_0^k, \hat{b}_1^k\right\} = \arg\min \sum_{d \in \mathcal{W}_w(k)} \rho_\alpha\left(r_d - b_0^k - b_1^k h_{d-1}\right), \tag{13}$$

where

$$\rho_\alpha\left(z\right) = z\left(\alpha - \mathbb{I}_{\Re_{<0}}\left(z\right)\right) \tag{14}$$

is the asymmetric absolute loss function and

$$\mathbb{I}_{\mathcal{A}}\left(z\right) = \begin{cases} 1, & z \in \mathcal{A} \\ 0, & z \notin \mathcal{A} \end{cases}. \tag{15}$$

is the indicator function.

Our forecasting model for the daily CVaR_α at day $k + w + 1$ using the entropy of the day $k + w$ is given by:

$$\hat{\mathrm{CVaR}}_{\alpha, k+w+1} = -\hat{c}_0^k - \hat{c}_1^k h_{k+w}, \tag{16}$$

where $\hat{c}_0^k$ and $\hat{c}_1^k$ are estimated using a quantile regression model between the dependent variable r_d and the independent variable h_{d-1} for $d \in \mathcal{W}_w\left(k\right)$. We consider the following optimization problem for the quantile regression estimation:

$$\left\{\hat{c}_0^k, \hat{c}_1^k\right\} = \arg\min \sum_{d \in \mathcal{W}_k} \rho_{\alpha^\star}\left(r_d - c_0^k - c_1^k h_{d-1}\right), \tag{17}$$

where

$$\alpha^\star = \hat{F}_{w,k}\left(-\frac{1}{\alpha}\int_{-\infty}^{\inf\left\{x|\hat{F}_{w,k}(x)\geq\alpha\right\}} z\hat{f}_{w,k}(z)\,dz\right) \tag{18}$$

is the significance level, $\hat{F}_{w,k}(\cdot)$ is the empirical cumulative distribution function of the logreturns estimated using the time window $\mathcal{W}_w(k)$ and $\hat{f}_{w,k}(\cdot)$ is the empirical density function of the logreturns estimated using the time window $\mathcal{W}_w(k)$.

3. Empirical Study

3.1. Bitcoin

There are several time series prices for Bitcoin depeding on the digital currency exchange and the currency used in the trading process. In order to compare our results to that obtained by Pele and Mazurencu-Marinescu-Pele [11], we adopt the BTC/USD exchange rate from Gemini Trust Company, LLC (Gemini). Gemini is a digital currency exchange and custodian that allows customers to buy, sell, and store digital assets. In particular, we consider the intraday closing prices of the minute-by-minute time bins and the time period from 8 October 2015 until 29 May 2019. According to Feng et al. [19] apud Pele and Mazurencu-Marinescu-Pele [11], the market capitalization, the daily transaction volume and the liquididy of Bitcoin before 2015 was not good.

For illustration purposes, in Figure 1, we present the Bitcoin's daily closing prices; in Figure 2, we present the Bitcoin's daily close-to-close logreturns; and, finally, in Figure 3, we present the empirical probability density and cumulative distribution functions of the Bitcoin's daily close-to-close logreturns. It is possible to notice the huge increase in the Bitcoin's prices until the end of 2017, the high volatility of the Bitcoin's logreturns and the change over time of the volatility pattern. In addition, it is also possible to notice the existence of extreme values in the distribution of Bitcoin's logreturns. In the following sections, we present the entropy of the symbolic intraday distribution of Bitcoin's logreturns, the logistic model connecting the daily CVaR and the entropy, and a forecasting model for the daily CVaR using the entropy based on a modified quantile regression model.

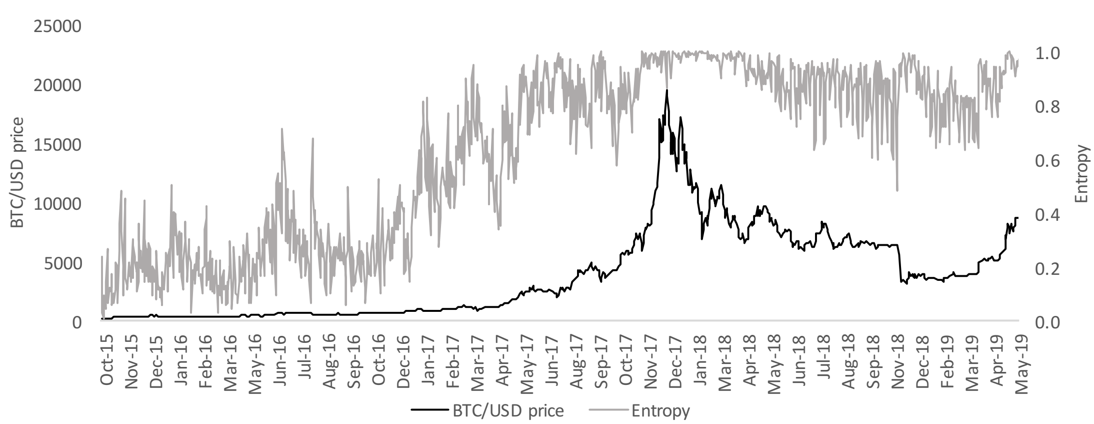

Figure 1. Bitcoin's daily closing prices and entropies.

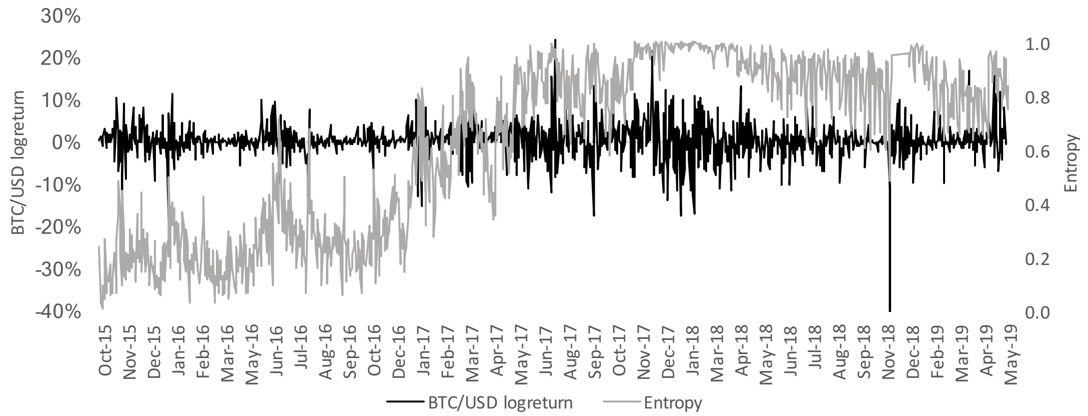

Figure 2. Bitcoin's daily close-to-close logreturns and entropies.

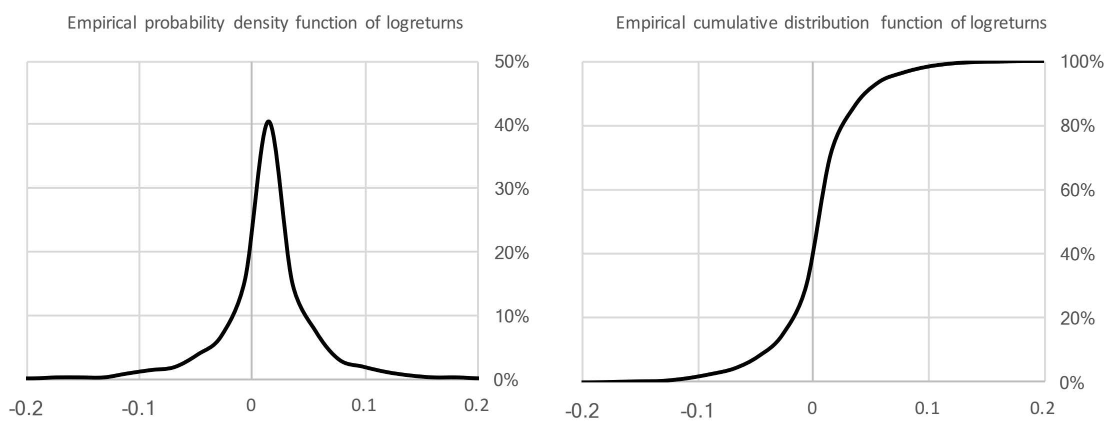

Figure 3. The empirical probability density function and the empirical cumulative distribution distribution function of the Bitcoin's daily close-to-close logreturns.

3.2. Entropy and Daily CVaR

In Figures 1 and 2, we also present the entropy of the symbolic intraday distribution of Bitcoin's logreturns. As it was mentioned in Section 2, the entropy of the symbolic intraday logreturns is higher at the presence of higher uncertainty in the returns and lower at the presence of lower uncertainty in the returns. In [9,11], they have tested the hypothesis that the daily logprice of Bitcoin is positively correlated to the entropy of the symbolic intraday distribution of Bitcoin's logreturns. However, we state that the logprice time series because of its level, non-stationarity and trend cause possible problems to the hypothesis verification. Consequently, we test the following model:

$$|r_d| = a_0 + a_1 h_d + \varepsilon_d, \tag{19}$$

where ε_d is the error term. Our hypothesis about Equation (19) is that the absolute value of daily logreturn of Bitcoin is positively correlated to the entropy of the symbolic intraday distribution of Bitcoin's logreturns. The estimation results of Equation (19) are shown in Table 1. It is possible to notice that the estimated coefficient a_1 of the entropy is positive and significant supporting our hypothesis.

Table 1. Estimation results of Equation (19).

Parameter	Estimation	*p*-Value	Standard Error
a_0	0.006	0.000	0.005
a_1	0.032	0.000	0.008
R^2	0.132		

In this paper, we propose the study of the relation between entropy of the symbolic intraday distribution of Bitcoin's logreturns and the likelihood of extreme negative daily logreturns represented by the daily CVaR using Equation (8). The estimation results are shown in Tables 2 and 3 for $\alpha = 1\%$ and $\alpha = 5\%$, respectively. It is possible to notice that the estimated coefficients c_1 of the entropy for both $\alpha = 1\%$ and $\alpha = 5\%$ are positive and significant supporting the hypothesis that entropy is positively correlated to the likelihood of extreme values of daily logreturns.

Table 2. Estimation results of Equation (8) for $\alpha = 1\%$.

Parameter	Estimation	*p*-Value	Standard Error
c_0	−9.133	0.002	1.316
c_1	8.253	0.001	3.488

Table 3. Estimation results of Equation (8) for $\alpha = 5\%$.

Parameter	Estimation	*p*-Value	Standard Error
c_0	−6.961	0.001	0.592
c_1	7.800	0.001	1.605

3.3. Forecasting Daily CVaR

Let $\hat{\text{CVaR}}_{\alpha,\mathcal{W}_w(k)}$ be the historical daily CVaR at the significance level α calculated in the time window $\mathcal{W}_w(k)$. The forecasting model for the daily CVaR_α at day $k+w+1$ using $\hat{\text{CVaR}}_{\alpha,\mathcal{W}_w(k)}$ is given by:

$$\hat{\text{CVaR}}_{\alpha,k+w+1} = \hat{\text{CVaR}}_{\alpha,\mathcal{W}_w(k)}. \tag{20}$$

In order to study the forecasting performance of the daily CVaR_α, we estimate Equations (16) and (20) using a rolling window approach with a window length $w = 250$ trading days. For comparison purposes, in Tables 4 and 5, we present a backtest of models (16) and (20) for significance levels $\alpha = 1\%$ and 5%, respectively. The performance of the models is compared with the historical daily CVaR at the significance level α calculated in the time window $\mathcal{W}_w(k+1)$. In particular, we consider the mean absolute error (MAE) and the root mean squared error (RMSE) between $\hat{\text{CVaR}}_{\alpha,k+w+1}$ and $\hat{\text{CVaR}}_{\alpha,\mathcal{W}_w(k+1)}$. As it is possible to notice from our empirical results using Bitcoin, the use of entropy in the forecasting of the daily CVaR of the next day seems to be better than the naive use of the historical CVaR.

Table 4. Backtest results of daily CVaR at the significance level $\alpha = 1\%$.

Model	MAE	RMSE
Forecasting using entropy	5.26×10^{-5}	7.28×10^{-4}
Forecasting using historical CVaR	3.56×10^{-4}	4.52×10^{-3}

Table 5. Backtest results of daily CVaR at the significance level $\alpha = 5\%$.

Model	MAE	RMSE
Forecasting using entropy	1.04×10^{-4}	5.42×10^{-4}
Forecasting using historical CVaR	3.16×10^{-4}	1.51×10^{-3}

4. Conclusions

In this paper, we have two main contributions: a logistic model connecting the daily CVaR and the entropy, and a forecasting model for the daily CVaR using the entropy based on a modified quantile regression model. Basically, we extend the study performed by Pele and Mazurencu-Marinescu-Pele [11] to include the CVaR. In [9,11], they have tested the hypothesis that the daily logprice of Bitcoin is positively correlated to the entropy of the symbolic intraday distribution of Bitcoin's logreturns. However, since the logprice time series is in level and presents a non-stationarity behavior and a trend, the verification of their hypothesis becomes infeasible. Consequently, the hypothesis we verify is that the absolute value of daily logreturn of Bitcoin is positively correlated to the entropy. In addition, we also verify that entropy is positively correlated to the likelihood of extreme values of Bitcoin's daily logreturns using a logistic regression model based on CVaR and the use of entropy to forecast the Bitcoin's daily CVaR of the next day performs better than the naive use of the historical CVaR.

Author Contributions: H.H.T. and S.X.A. developed the models, implemented the calculations and wrote the paper; J.M.S. and C.O.R. provided scientific supervision. All authors have read and approved the final manuscript.

Acknowledgments: The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper. The authors are grateful for the support of POLI-USP, the Polytechnic School of the University of São Paulo; IME-USP, the Institute of Mathematics and Statistics of the University of São Paulo; and FAPESP, the State of São Paulo Research Foundation, through the Research Center for Gas Innovation under the grant 2014/50279-4.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CVaR	Conditional Value at Risk
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
STSA	Symbolic Time Series Analysis
VaR	Value at Risk

References

1. Rockafellar, R.T.; Uryasev, S.P.O ptimization of conditional value-at-risk. *J. Risk* **2000**, *2*, 21–42.
2. Love, R.F.; Morris, J.G.; Wesolowsky, G.O. *Facilities Location: Models & Methods*; NY: North-Holland, The Netherlands, 1988.
3. Sarykalin, S.; Serraino, G.; Uryasev, S.P. Value-at-risk vs. conditional value-at-risk in risk management and optimization. *Tutorials Oper. Res.* **2008**, 270–294. doi:10.1287/educ.1080.0052.
4. Yamai, Y.; Yoshiba, T. Comparative analyses of expected shortfall and value-at-risk: their validity under market stress. *Monet. Econ. Stud.* **2002**, *20*, 181–237.
5. Rockafellar, R.T.; Uryasev, S.P. Conditional value-at-risk for general loss distributions. *J. Bank. Financ.* **2002**, *26*, 1443–1471.

6. Philippatos, G.C.; Wilson, C. Entropy, market risk and the selection of efficient portfolios. *Appl. Econ.* **1972**, *4*, 209–220.
7. Ebrahimi, N.; Maasoumi, E.; Soofi, E.S. Ordering univariate distributions by entropy and variance. *J. Econ.* **1999**, *90*, 317–336.
8. Dionisio, A.; Menezes, R.; Mendes, D.A. An econophysics approach to analyze uncertainty in financial markets: an application to the Portuguese stock market. *Eur. Phys. J. B* **2006**, *50*, 161–164.
9. Pele, D.T.; Lazar, E.; Dufour, A. Information entropy and measures of market risk. *Entropy* **2017**, *19*, 226.
10. Billio, M.; Casarin, R.; Costola, M.; Pasqualini, A. An entropy-based early warning indicator for systemic risk. *J. Int. Financ. Mark. Inst. Money* **2016**, *45*, 42–59.
11. Pele, D.T.; Mazurencu-Marinescu-Pele, M. Using high-frequency entropy to forecast bitcoins daily value at risk. *Entropy* **2019**, *21*, 102.
12. Zhang, W.; Wang, P.; Li, X.; Shen, D. Some stylized facts of the cryptocurrency market. *Appl. Econ.* **2018**, *50*, 5950–5965.
13. Hu, A.; Parlour, C.A.; Rajan, U. Cryptocurrencies: stylized facts on a new investible instrument. *Work. Pap.* **2018**.
14. Saa, O.T.; Stern, J.M. Auditable Blockchain Randomization Tool. *arXiv* **2019**, arXiv:1904.09500.
15. Colucci, S. On Estimating Bitcoin Value at Risk: A Comparative Analysis. *Work. Pap.* **2018**. doi:10.2139/ssrn.3236813.
16. Daw, C.; Finney, C.; Tracy, E. A review of symbolic analysis of experimental data. *Rev. Sci. Instrum.* **2003**, *74*, 915–930.
17. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.
18. Koenker, R.; Bassett, G. Regression quantiles. *Econometrica* **1978**, *46*, 33–50.
19. Feng, W.; Wang, Y.; Zhang, Z. Can cryptocurrencies be a safe haven: A tail risk perspective analysis. *Appl. Econ.* **2018**, *50*, 4745–4762.