Forecasting Bitcoin and Ethereum risk measures through MSGARCH models: Does the specification matter?

Luiz K. Hotta[†, a]
Carlos Trucíos[‡, a]
Pedro L. Valls Pereira[*, b]
Mauricio Zevallos[**, a]

[a]Department of Statistics, Universidade Estadual de Campinas (UNICAMP), Brazil.
[b]Sao Paulo School of Economics - FGV, Brazil.

Abstract Recent studies have suggested that more complex models than GARCH are better suited for forecasting cryptocurrency risk measures, such as Value-at-Risk and Expected Shortfall. Among these studies, some highlight the advantages of MS-GARCH models over traditional GARCH models. While improvements over single-regime GARCH models have been observed by using MSGARCH, the literature has only focused on the MSGARCH specification proposed by Haas, Mittnik and Paolella (Journal of Financial Econometrics, 2004) overlooking several other well-established MSGARCH specification alternatives. In this paper, we illustrate that exploring alternative MSGARCH specifications can lead to improvements in risk measure performance, emphasizing the potential benefits of using several specifications.

Keywords: Cryptocurrencies; Expected Shortfall; Maximum likelihood; Model misspecification; Value-at-Risk.

JEL codes: C13, C18, C52, C53, G17.

1. Introduction

In finance, forecasting risk measures such as Value-at-Risk (VaR) and Expected Shortfall (ES) is fundamental for risk management, the development of (algorithmic) trading systems, the formulation of order limit strategies and option pricing models, among other applications.

How to cite: Hotta, L. K., Trucíos, C., Valls Pereira, P. L. & Zevallos, M. (2025). Forecasting Bitcoin and Ethereum risk measures through MSGARCH models: Does the specification matter?. *Revista Brasileira de Finanças, 23*, e202503. https://doi.org/10.12660/rbfin.v23n1.2025.92777
Submitted on January 15, 2025. Published online on March 10, 2025. Editor in charge: Bruno Giovannetti.

[†]hotta@unicamp.br
[‡]ctrucios@unicamp.br
[*]pedro.valls@fgv.br
[**]amadeus@unicamp.br

©2025 Sociedade Brasileira de Finanças, under a Creative Commons Attribution 4.0 International license.

In the cryptocurrency market, forecasting VaR and ES is particularly challenging due to its inherent characteristics, such as high volatility, extreme price movements, and periods of extreme turbulent behavior (see Ardia, Bluteau and Rüede (2019), Troster et al. (2019), Trucíos (2019), Trucíos and Taylor (2023) and de Prince et al. (2025)). To overcome these difficulties in the modeling process, more complex models than the single-regime GARCH are necessary (see Ardia et al. (2018), Caporale and Zekokh (2019), Troster et al. (2019), Trucíos (2019) and Maciel (2021)). In particular, the aforementioned studies showed that the robust GARCH of Trucíos et al. (2017), the MSGARCH of Haas et al. (2004) and the GAS of Harvey (2013) and Creal et al. (2013) are among the best alternatives for forecasting risk measures.

Several studies highlight the superiority of certain complex models over single-regime GARCH-type models in terms of risk measures forecasting. However, almost all of these studies compare the proposed complex models only against various single-regime GARCH-type models. In this respect, Trucíos and Taylor (2023) provides a comprehensive comparison of a wide class of models and concludes that the GAS model is hard to outperform in terms of risk measures predictability and that MSGARCH models are better than single-regime GARCH alternatives. However, in Trucíos and Taylor (2023), as well as in similar studies comparing MSGARCH models over single-regime GARCH-type models, the only used MSGARCH specification is the proposal of Haas et al. (2004), neglecting other well-established regime-switching specifications, such as those proposed by Gray (1996) or Klaassen (2002).

The purpose of the paper is to enhance the class of MSGARCH alternatives in the comparison of forecasting risk forecasting ability on criptocurrencies. Thus, in this paper, we apply two other well-established MSGARCH specifications proposed in Gray (1996) and Klaassen (2002), to Bitcoin and Ethereum, the major cryptocurrencies, and show that these alternative specifications can outperform the model proposed by Haas et al. (2004)[1]. This underscores the importance for practitioners of exploring other MSGARCH specifications when looking for accurate risk forecasts.

The remaining part of the paper is organized as follows. Section 2 introduces the used MSGARCH models and their estimation method. Section 3 describes the data, the model comparison procedure and reports the main

[1]The MSGARCH model proposed by Haas et al. (2004) is nowadays the most widely used model in empirical applications (partly because of the friendly and versatile MSGARCH R package implemented by Ardia, Bluteau, Boudt, Catania and Trottier (2019)), and Gray (1996) and Klaassen (2002) are prominent in the literature given their strong empirical focus and use. These facts motivated our models choice.

results. Finally, Section 4 concludes.

2. Markov Switching GARCH Models

Consider an unobserved state variable S_t that assumes K possible regimes (states): $S_t \in \mathbb{M}$ with $\mathbb{M} = \{1,\ldots,K\}$. Suppose that the $\{S_t\}$ process follows a first-order Markov chain process with transition probability matrix Π, where the elements $\pi_{ij} = \mathbb{P}(S_t = j|S_{t-1} = i)$ represent the probability of the Markov chain moving from state i at time $t-1$ to state j at time t.

In this paper, three different MSGARCH models are compared. All models are specified with two regimes and an order of (1,1) for each regime, as these are the most common choices in empirical applications. The first specification was introduced by Gray (1996). In this model, returns are generated by $y_t = \{h(S_t,\mathscr{F}_{t-1})\}^{1/2} z_t$ where $\mathscr{F}_{t-1}$ is the information up to time $t-1$ and does not include S_t or lagged values of S_t. For each regime, the volatility is given by $h_t^{(k)} = h(S_t = k,\mathscr{F}_{t-1}) = \omega^{(k)} + \alpha^{(k)} y_{t-1}^2 + \beta^{(k)} h_{t-1}$, for $k = 1,2$ where h_t is the volatility at time t given the information available up to time $t-1$ and is expressed as $h_t = \mathbb{V}(y_t|\mathscr{F}_{t-1}) = p_{1,t} h_t^{(1)} + (1-p_{1,t}) h_t^{(2)}$, with $p_{i,t} = \mathbb{P}(S_t = i|\mathscr{F}_{t-1})$. In this and the subsequent notations, the superscript $^{(k)}$ denotes values of parameters and volatility for regime k; for instance, $\alpha^{(k)} = \alpha(S_t = k)$ and $h_t^{(k)}$ indicates the volatility at time t for regime k.

One of the main critiques of the model proposed by Gray (1996) is that it does not utilize all information available up to time t. To address this, Klaassen (2002) proposed a modification in which the model is given by a Markov process $\{S_t\}$ with two regimes, where the conditional variance is given by $\mathbb{V}(y_t|\mathscr{F}_{t-1},\tilde{S}_t) = \omega^{(k)} + \alpha^{(k)} y_{t-1}^2 + \beta^{(k)} \mathbb{E}_{t-1}(\mathbb{V}_{t-2}(y_{t-1}|\tilde{S}_{t-1})|S_t)$, where $\tilde{S}_t = \{S_t, S_{t-1}, \cdots\}$ denotes the regime path. $\mathbb{E}_t$ and $\mathbb{V}_t$ denote, respectively, the expectation and variance conditioned on the information $\mathscr{F}_t$, with the conditional distribution $y_t|\mathscr{F}_{t-1},\tilde{S}_t \sim WS(0,\mathbb{V}(y_t|\mathscr{F}_{t-1},S_t))$, where WS represents any distribution with zero mean and conditional variance $\mathbb{V}(y_t|\mathscr{F}_{t-1},S_t)$.

Haas et al. (2004) and Haas and Paolella (2012) proposed a model that also circumvents the path dependence estimation problem and, according to them, offers better interpretability than previous MSGARCH models. In their model, which can be viewed as a direct generalization of the single-regime GARCH model specification, the observation equation is written as $y_t = \{h_t^{(k)}\}^{1/2} z_t$, where z_t is an IID innovation process, and each regime-specific conditional variance depends only on its own lags. That is, $h_t^{(k)} = \omega^{(k)} + \alpha^{(k)} y_{t-1}^2 + \beta^{(k)} h_{t-1}^{(k)}$, with $\omega^{(k)} > 0$, $\alpha^{(k)}, \beta^{(k)} \geqslant 0$, and the transition process $\{S_t\}$ is defined as before. The key difference compared to previous

models is that the volatility in each regime depends on the past volatility in the same regime, whereas in the proposals of Gray (1996) and Klaassen (2002), the volatility at time t depends on the volatility at the previous time, independently of the current and previous regimes.

In this paper, the estimation was performed by maximum likelihood as follows. First, 5000 random parameter vectors in the parameter space were randomly generated (random grid) and their corresponding likelihood evaluated. Second, the three vectors with largest likelihood values were used as initial values in the optimization routine and solutions were recorded. Third, among the three solutions obtained, we selected the final estimates as those with the largest likelihood.

Optimization was carried out using the Nelder-Mead algorithm (Nelder and Mead, 1965) with transformed parameters, and for the ES calculation, numerical integration was performed by the adaptive Gauss-Kronrod quadrature method. Computations were performed using our own programs written in Julia (Bezanson et al., 2017).

3. Empirical Applications

3.1 Data

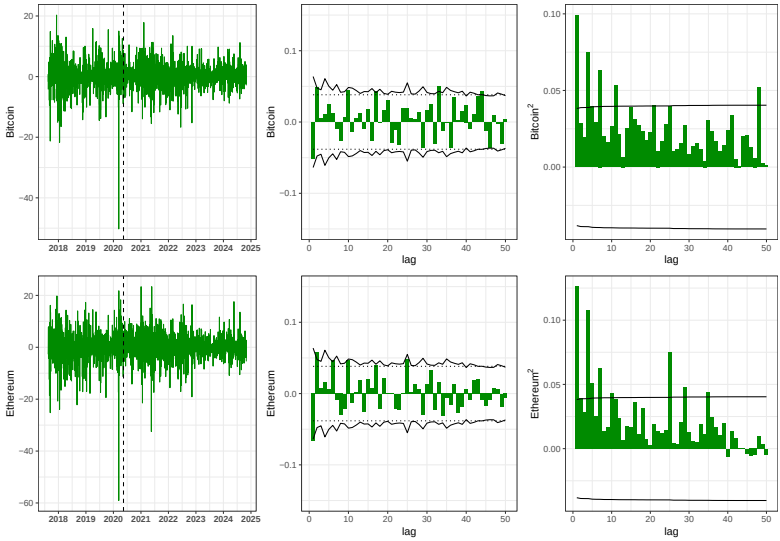
Daily prices of Bitcoin (BTC/USDT) and Ethereum (ETH/USDT) spanning from August 17, 2017 to October 31, 2024 (2633 observations), were freely downloaded from Binance, one of the largest crypto-exchange worldwide. As usual, prices were converted into log returns (in percentage) as $r_t = 100 \times \log(P_t/P_{t-1})$, where P_t is the closing price at day t. Since cryptocurrencies are traded 24/7, the last price traded on the day is considered as closing price.

Some descriptive statistics over the full sample period are reported in Table 1 and reveals that returns exhibit heavy tails, positive mean and negative skewness. Bitcoin and Ethereum reported, approximately, an annualized volatility of 72% and 92%, respectively. Additionally, Figure 1 reports the daily returns and sample auto-correlation function of returns and squared returns. The 95% confidence bands were computed using the generalized non-parametric Bartlett's formula (Francq and Zakoïan, 2009) for returns and the Bartlett's formula (Bartlett, 1946) for squared returns, respectively. The Figure reveals no or very weak ARMA effect was observed, so returns were simply centered.

Table 1
Descriptive statistics of daily returns

	Min	Q1	Median	Mean	Q3	Max	Sd	Skew	Kurtosis
BTC	−50.26	−1.45	0.09	0.11	1.72	20.30	3.79	−0.99	17.83
ETH	−59.05	−1.97	0.07	0.08	2.30	23.38	4.79	−1.00	15.27

Figure 1
Daily returns, sample auto-correlation function of returns, and sample auto-correlation function of squared returns



Daily returns (left panel), sample auto-correlation function of returns (middle panel), and sample auto-correlation function of squared returns (right panel) for Bitcoin and Ethereum. Full sample period. Vertical dashed lines (May 14, 2020) split the data into in-sample and out-of-sample periods.

3.2 Out-of-sample comparison

The three MSGARCH models specified in Section 2 (Gray, 1996; Klaassen, 2002; Haas et al., 2004), were compared in terms of their risk forecasting ability using a rolling windows scheme with window size equal to 1000 days, which implies in a out-of-sample period for evaluation equal to 1633 days. For each window, we computed the one-step-ahead VaR and ES at the 1% and 2.5% risk levels using the aforementioned models with standard normal and Student-t innovation densities. Henceforth, the Gray (1996) models with normal and

Student-t distributions are denoted as Gray-N and Gray-T, respectively, and the other models are denoted similarly. To compare the six estimated models, we calculated a set of well known test statistics.

Specifically, we used the following calibration tests: the unconditional coverage test (UC) of Kupiec (1995); the conditional coverage test (CC) of Christoffersen (1998); the dynamic quantile test (DQ) of Engle and Manganelli (2004); the VaR quantile test (VQ) of Gaglianone et al. (2011); the exceedance residual test (ER) of McNeil and Frey (2000); the two-sided conditional calibration test (CoC) of Nolde and Ziegel (2017); and the one-sided intercept expected shortfall regression test (ESR) of Bayer and Dimitriadis (2022). In addition, we used the following average scoring functions: the average quantile loss function of González-Rivera et al. (2004), denoted AQL; the loss function of Fissler et al. (2016), denoted AFZG; the loss function of Nolde and Ziegel (2017), denoted ANZ; and the loss function of Taylor (2019), denoted AAL.

The results for hits (percentage of returns smaller than VaR), calibration tests, and average scoring functions of both cryptocurrencies are presented in Table 2. In that table, p-values of calibration tests smaller than 0.05 and procedures with average scoring values that do not belong to the 75% Model Confidence Set (MCS) of Hansen et al. (2011) are marked in bold.

Based on the results, we selected what we considered the best models for each series. A model was considered to have good performance when it passed all the calibration tests and simultaneously was inside the MCS. Additionally, among those methods, we highlight the one(s) that exhibited the best performances in terms of hits. In the tables, the selected models for each case are marked in bold.

For the ETH, at 1% risk level, Gray-N, Klaassen-N, Klaassen-T and Haas-T passed the calibration tests and were inside the MCS. Among these, Gray-N, Klaassen-T and Haas-T present the best (same) percentage of hits. At 2.5% risk level only Gray-N and Klaassen-N passed the calibration tests and were inside the MCS, the latest being selected in terms of hits.

For the BTC, at both risk levels, none of the methods passed the calibration tests and were inside the MCS. Most of the methods fail with the CoC test and to a lesser extent with the VQ test. Regarding those tests, simulation studies performed by Hotta et al. (2024) on GARCH models indicate that the empirical size of those calibration tests is bigger than the nominal size. Then, removing these tests from the analysis, we found that at 1% risk level, only Gray-N and Klaassen-T passed the (remaining) calibration tests and are inside the MCS, being the later method the best in terms of hits. In addition, at 2.5% risk level, Gray-T, Klaassen-N, Klaassen-T and Haas-T methods pass the tests and are

inside the MCS. Among these methods, the best performance (in terms of hits) is shown by the Haas-T, followed by the Klaassen-T method.

In summary, Gray's model was selected only once, while Haas's model was chosen twice and Klaassen's model was selected three times. Models with Gaussian distributions were selected two times (Gray-N and Klaassen-N), while those with Student-t distribution 4 times (Haas-T and Klaassen-T, two times each).

4. Conclusions

Although the Haas-T model is the Markov-Switching GARCH model most widely used in empirical applications, the paper shows that there are other MSGARCH models that can perform statistically as well as, or in some cases, better than it. Therefore, when forecasting risk measures is the main goal, practitioners should consider several MSGARCH alternatives.

Acknowledgments All authors acknowledge financial support of the São Paulo Research Foundation (FAPESP), grant 2023/01728-0. The first and fourth authors also acknowledge financial support of the São Paulo Research Foundation (FAPESP), grants 2018/04654-9 and 2023/02538-0. Additionally, the first author acknowledges financial support from the National Council for Scientific and Technological Development (CNPq), grant 309328/2020-9. The second author also acknowledges financial support of the São Paulo Research Foundation (FAPESP), grants 2022/09122-0 and 2023/02538-0 as well as Programa de Incentivo a Novos Docentes da UNICAMP (PIND), grant 2525/23. The third author also acknowledges financial support from CNPq grant 311734/2021-9. The authors also thank the Centre for Applied Research on Econometrics, Finance, and Statistics (CAREFS), Brazil, for their computational support.

Conflict of interest The authors declare no conflict of interest.

Artificial Intelligence Some sections of this paper were supported by AI tools for language refinement and grammatical corrections. All AI-generated content was critically reviewed and validated by the authors to ensure accuracy and alignment with the scientific integrity of the study. The use of AI adhered to ethical guidelines, ensuring transparency and compliance with academic standards. The authors affirm that the AI tools did not compromise the originality or integrity of the work.

Table 2
One-step-ahead VaR and ES backtesting

	Risk level	Model	Hits	Calibration tests p-values							Avg. Scoring functions			
				UC	CC	DQ	VQ	ER	CoC	ESR	AQL	AFZG	ANZ	AAL
ETH	1%	**Gray-N**	0.9191	0.7392	0.8231	0.5550	0.7180	0.6742	0.8528	0.6089	0.1521	0.8439	3.8306	3.6712
		Gray-T	0.9804	0.9363	0.3472	0.5752	**0.0000**	0.8967	0.0714	0.8479	0.1469	0.8463	3.7823	3.6727
		Klaassen-N	0.7966	0.3918	0.1651	0.3863	0.5694	0.6355	0.6512	0.7651	0.1516	0.8436	3.8218	3.6669
		Klaassen-T	0.9191	0.7392	0.2936	0.5823	0.8004	0.9184	0.0956	0.8947	0.1470	0.8385	3.7540	3.6237
		Haas-N	0.7966	0.3918	0.1651	**0.0452**	0.4459	0.7715	0.6452	0.8198	0.1535	0.8451	3.8493	3.6799
		Haas-T	0.9191	0.7392	0.2936	0.5794	0.7879	0.8312	0.3314	0.8361	0.1472	0.8387	3.7532	3.6208
	2.5%	Gray-N	2.3284	0.6534	0.8974	0.3906	0.7297	0.6085	0.8835	0.6344	0.2947	0.9841	3.3785	3.4412
		Gray-T	2.8186	0.4189	0.2996	**0.0026**	0.1095	0.8606	0.0931	0.8407	0.2925	0.9896	3.3647	3.4439
		Klaassen-N	2.5123	0.9747	0.6657	0.6023	0.2037	0.7752	0.5707	0.7856	0.2928	0.9826	3.3580	3.4247
		Klaassen-T	2.4510	0.8987	0.2345	**0.0293**	0.3467	0.8266	0.2852	0.8278	0.2919	0.9808	3.3487	3.4143
		Haas-N	2.3284	0.6534	0.5219	**0.0159**	0.8501	0.8043	0.7791	0.7602	**0.2998**	0.9889	3.4043	3.4540
		Haas-T	2.6348	0.7295	0.3016	**0.0224**	0.2156	0.8177	0.3270	0.7788	0.2917	0.9805	3.3461	3.4117
BTC	1%	Gray-N	0.6127	0.0903	0.2240	0.6286	**0.0022**	0.9353	**0.0142**	0.9624	0.1150	0.8068	3.3545	3.4196
		Gray-T	0.9804	0.9363	0.3472	0.3357	0.1362	0.9926	**0.0000**	0.9937	**0.1191**	**0.8107**	**3.4420**	**3.4775**
		Klaassen-N	0.9191	0.7392	0.2936	0.5555	0.2789	0.9172	0.0656	0.8790	0.1185	**0.8109**	**3.4143**	**3.4584**
		Klaassen-T	0.9191	0.7392	0.2936	0.5326	0.9936	0.9995	**0.0000**	0.9972	0.1138	0.8052	3.3507	3.4173
		Haas-N	0.4902	**0.0216**	0.0688	0.4869	**0.0134**	0.5106	**0.0102**	0.8366	0.1218	**0.8141**	**3.4884**	**3.5202**
		Haas-T	1.0417	0.8666	0.3818	0.6043	0.3950	0.9991	**0.0000**	0.9952	0.1175	**0.8090**	**3.4006**	**3.4455**
	2.5%	Gray-N	1.5931	**0.0120**	**0.0314**	0.3903	**0.0150**	0.7822	**0.0139**	0.9912	0.2393	0.9291	3.0579	3.2500
		Gray-T	2.3284	0.6534	0.8974	0.9594	0.0700	0.9637	**0.0003**	0.9898	0.2316	0.9204	3.0166	3.2242
		Klaassen-N	2.9412	0.2665	0.2606	0.2068	0.1965	0.9774	**0.0235**	0.8687	0.2310	0.9213	3.0070	3.2175
		Klaassen-T	2.3897	0.7737	0.9571	0.4359	0.2607	0.9834	**0.0002**	0.9841	0.2298	0.9183	2.9929	3.2011
		Haas-N	1.6544	**0.0198**	0.0510	0.1530	0.1450	0.6540	**0.0248**	0.9702	**0.2393**	**0.9295**	**3.0756**	**3.2736**
		Haas-T	2.5735	0.8498	0.9790	0.2800	**0.0340**	0.9653	**0.0028**	0.9524	0.2328	0.9214	3.0158	3.2184

One-step-ahead VaR and ES backtesting for BTC/USDT (top panel) and ETH/UST (bottom panel) at the 1% and 2.5% risk levels. P-values of calibration tests smaller than 0.05 and procedures with average score not included in the 75% MCS are highlighted in bold. Window size equal to $n = 1000$ days and out-of-sample period spanning from May 14, 2020 to October 31, 2024 (1632 observations). Selected models are highlighted in bold.

References

Ardia, D., Bluteau, K., Boudt, K., Catania, L. and Trottier, D.-A. (2019). Markov-Switching GARCH models in R: The MSGARCH package, *Journal of Statistical Software*, *91*(4).

Ardia, D., Bluteau, K., Boudt, K. and Catania, L. (2018). Forecasting risk with Markov-switching GARCH models: A large-scale performance study, *International Journal of Forecasting*, *34*(4), 733–747.

Ardia, D., Bluteau, K. and Rüede, M. (2019). Regime changes in Bitcoin GARCH volatility dynamics, *Finance Research Letters*, *29*, 266–271.

Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time-series, *Supplement to the Journal of the Royal Statistical Society*, *8*(1), 27–41.

Bayer, S. and Dimitriadis, T. (2022). Regression-based expected shortfall backtesting, *Journal of Financial Econometrics*, *20*(3), 437–471.

Bezanson, J., Edelman, A., Karpinski, S. and Shah, V. B. (2017). Julia: A fresh approach to numerical computing, *SIAM Review*, *59*(1), 65–98.

Caporale, G. M. and Zekokh, T. (2019). Modelling volatility of cryptocurrencies using Markov-Switching GARCH models, *Research in International Business and Finance*, *48*, 143–155.

Christoffersen, P. F. (1998). Evaluating interval forecasts, *International Economic Review*, *39*(4), 841–862.

Creal, D., Koopman, S. J. and Lucas, A. (2013). Generalized autoregressive score models with applications, *Journal of Applied Econometrics*, *28*(5), 777–795.

de Prince, D., Marçal, E. F. and Valls Pereira, P. L. (2025). Exploring co-explosive dynamics: Bitcoin price, attractiveness, and sentiment variables, *Economics Letters*, *246*, 112072.

Engle, R. F. and Manganelli, S. (2004). CAViaR: conditional autoregressive value at risk by regression quantiles, *Journal of Business & Economic Statistics*, *22*(4), 367–381.

Fissler, T., Ziegel, J. F. and Gneiting, T. (2016). Expected Shortfall is jointly elicitable with Value at Risk - Implications for backtesting, *Risk*, *5*(1), 8–16.

Francq, C. and Zakoïan, J.-M. (2009). Bartlett's formula for a general class of nonlinear processes, *Journal of Time Series Analysis*, *30*(4), 449–465.

Gaglianone, W. P., Lima, L. R., Linton, O. and Smith, D. R. (2011). Evaluating value-at-risk models via quantile regression, *Journal of Business & Economic Statistics*, *29*(1), 150–160.

González-Rivera, G., Lee, T.-H. and Mishra, S. (2004). Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood, *International Journal of Forecasting*, *20*(4), 629–645.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process, *Journal of Financial Economics*, *42*(1), 27–62.

Haas, M. and Paolella, M. S. (2012). Mixture and Regime-Switching GARCH Models, *in* L. Bauwens, C. M. Hafner and S. Laurent (eds), *Handbook of Volatility Models and Their Applications*, John Wiley & Sons, Hoboken, New Jersey, chapter 3, pp. 71–102.

Haas, M., Mittnik, S. and Paolella, M. S. (2004). A new approach to Markov-switching GARCH models, *Journal of Financial Econometrics*, *2*(4), 493–530.

Hansen, P. R., Lunde, A. and Nason, J. M. (2011). The model confidence set, *Econometrica*, *79*(2), 453–497.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*, Vol. 52, Cambridge University Press.

Hotta, L. K., Trucíos, C. and Zevallos, M. (2024). A note about calibration tests for VaR and ES, *in* C. Chiann, A. Pinheiro and C. M. Toloi (eds), *Time Series and Wavelet Analysis: Festschrift in Honor of Pedro A. Morettin*, Springer, pp. 59–80.

Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH, *in* J. D. Hamilton and B. Raj (eds), *Advances in Markov-switching Models*, Springer, pp. 223–254.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models, *The Journal of Derivatives*, *3*(2), 73–84.

Maciel, L. (2021). Cryptocurrencies value-at-risk and expected shortfall: Do regime-switching volatility models improve forecasting?, *International Journal of Finance & Economics*, *26*(3), 4840–4855.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach, *Journal of Empirical Finance*, *7*(3-4), 271–300.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization, *The Computer Journal*, *7*(4), 308–313.

Nolde, N. and Ziegel, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation, *The Annals of Applied Statistics*, *11*(4), 1833–1874.

Taylor, J. W. (2019). Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution, *Journal of Business & Economic Statistics*, *37*(1), 121–133.

Troster, V., Tiwari, A. K., Shahbaz, M. and Macedo, D. N. (2019). Bitcoin returns and risk: A general GARCH and GAS analysis, *Finance Research Letters*, *30*, 187–193.

Trucíos, C. (2019). Forecasting Bitcoin risk measures: A robust approach, *International Journal of Forecasting*, *35*(3), 836–847.

Trucíos, C., Hotta, L. K. and Ruiz, E. (2017). Robust bootstrap forecast densities for GARCH returns and volatilities, *Journal of Statistical Computation and Simulation*, *87*(16), 3152–3174.

Trucíos, C. and Taylor, J. W. (2023). A comparison of methods for forecasting value at risk and expected shortfall of cryptocurrencies, *Journal of Forecasting*, *42*(4), 989–1007.