Applied Soft Computing Journal 188 (2026) 114384

Contents lists available at ScienceDirect

Applied Soft Computing

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

Identification of Bitcoin volatility drivers using statistical and machine learning methods

Piotr Fiszeder [a,b], Witold Orzeszko [a], Radosław Pietrzyk [c], Grzegorz Dudek [d,e,f,*]

[a] Faculty of Economic Sciences and Management, Nicolaus Copernicus University in Torun, ul. Gagarina 13a, Torun 87-100, Poland
[b] Faculty of Finance and Accounting, Prague University of Economics and Business, nám. W. Churchilla 1938/4, Prague 3, 130 67, Czech Republic
[c] Faculty of Economics and Finance, Wroclaw University of Economics and Business, ul. Komandorska 118-120, Wroclaw 53-345, Poland
[d] Faculty of Electrical Engineering, Czestochowa University of Technology, Al. AK 17, Czestochowa 42-201, Poland
[e] Faculty of Mathematics and Computer Science, University of Łódź, ul. Banacha 22, Łódź 90-238, Poland
[f] CAMINO – Centre for Data Analysis, Modelling and Computational Sciences, University of Łódź, ul. Narutowicza 68, Łódź 90-136, Poland

HIGHLIGHTS

- Analysis of 62 variables reveals key drivers of Bitcoin volatility.
- BMA, LASSO, and RF assess exogenous variable importance in forecasting.
- Selected exogenous variables significantly enhance Bitcoin variance predictions.
- Outlier replacement and standardization boost daily forecast accuracy.
- Top factors: lagged variances, trading volume, Google search intensity.

ARTICLE INFO

Keywords:
Cryptocurrency volatility
Market and policy uncertainty
Random forest
LASSO
Bayesian model averaging
HAR
Google trends attention
Trading volume

ABSTRACT

This study advances the understanding of Bitcoin volatility forecasting by analysing an extensive set of 62 explanatory variables, including cryptocurrency market behaviour, Google search trends, financial indices, and economic indicators. We employ Bayesian Model Averaging (BMA), Least Absolute Shrinkage and Selection Operator (LASSO), and Random Forest (RF) methods to assess variable importance and forecast accuracy. Our research demonstrates that LASSO and RF models incorporating exogenous variables significantly improve both daily and weekly Bitcoin variance forecasts compared to models using only lagged Bitcoin volatilities. Key factors influencing Bitcoin volatility include lagged realised variances, trading volume, and Google search intensity. The study reveals that the impact of these variables on Bitcoin volatility is time-varying, reflecting its evolving relationship with broader economic indicators and market sentiment. Our findings contribute to the literature by providing a comprehensive analysis of Bitcoin volatility drivers, evaluating the effectiveness of variable transformations, and comparing the performance of advanced forecasting methods in handling the cryptocurrency's extreme volatility. These insights are valuable for researchers, investors, portfolio managers, and policymakers navigating the dynamic cryptocurrency market.

1. Introduction

Since the emergence of Bitcoin (BTC) in 2009, the crypto assets market has grown dynamically, reaching a market capitalization of over USD 3.5 trillion.[1] Another stimulus for the development of this market has been the approval by the U.S. Securities and Exchange Commission on January 10, 2024, of the introduction of BTC exchange-traded funds, which has made it easier for small investors to invest in BTC. However, investing in cryptocurrencies involves high investment risk, which is a result of their high volatility. For this reason, effective volatility

* Corresponding author at: Faculty of Electrical Engineering, Czestochowa University of Technology, Al. AK 17, Czestochowa 42-201, Poland.
E-mail address: grzegorz.dudek@pcz.pl (G. Dudek).

[1] https://www.forbes.com.

https://doi.org/10.1016/j.asoc.2025.114384
Received 26 February 2025; Received in revised form 20 August 2025; Accepted 26 November 2025
Available online 13 December 2025

forecasting becomes crucial in the context of investment and hedging strategies, and for facilitating cryptocurrency payments. As a result, both in academia and among practitioners, there is a growing interest in various methods of volatility forecasting. In recent years, machine learning (ML) has been gaining increasing importance in the literature on this subject. The growing popularity of ML techniques in the context of financial market forecasting is, among other things, a result of their effectiveness in applying to nonlinear, non-stationary, and noisy data. In addition, they demonstrate strong robustness to missing data and can produce reliable results even when the number of observations is relatively small compared to the number of variables in the model (see, e.g., [1,2]). A review of the extensive literature on the application of these methods in various areas of finance can be found, for example in [3–9]. Recent advances in ML-based forecasting for Bitcoin markets are exemplified by the works [10–13].

Volatility forecasting models can generally be classified into two groups based on the nature of their regressors: those that rely solely on historical volatility data and those that incorporate additional exogenous variables. A prominent example from the first group is the linear heterogeneous autoregressive model (HAR) [14], which employs as regressors three lagged volatility components – daily, weekly, and monthly. Previous studies have shown that among the most commonly used statistical and ML models, no single method consistently provides the most accurate volatility forecasts based solely on historical volatility [15]. However, the accuracy of these models can be enhanced by integrating additional determinants that affect cryptocurrency variance (see Section 2). Peng, Prentice, Shams, and Sarker [16] present a systematic literature review of studies aimed at identifying factors influencing cryptocurrency pricing. It is worth noting that nearly all of these studies focus on the prices and returns of cryptocurrencies rather than on their volatility. Conversely, Gunnarsson, Isern, Kaloudis, Risstad, Vigdel, and Westgaard [17] review the literature on forecasting volatility using ML methods. However, this review encompasses all financial time series and cites only three studies specifically addressing BTC volatility.

As a starting point, we provide a comprehensive review of existing studies on the determinants of Bitcoin volatility (see Section 2). This review summarizes the explanatory variables and methodological approaches employed in prior research, thereby helping to contextualize our work and position it within the current state of the literature. The determinants of BTC volatility have been analysed in numerous studies. Some of these studies are reviewed by Kyriazis [18], Wang, Ma, Bouri, and Guo [19], and Benhamed, Messai, and El Montasser [20]. The most commonly examined factors include stock indices, fiat currencies, oil, gold, attention and sentiment measures, uncertainty indices, and trading volume. However, many of these studies do not demonstrate that the variables considered lead to improved ex-post volatility forecasts. Since investors and risk managers are primarily concerned with out-of-sample forecasting performance, such analyses are of critical importance. Based on the literature review presented in Section 2, we identify a research gap that we aim to address in this study. Existing work lacks a comprehensive approach to out-of-sample forecasting daily and weekly BTC volatility using a broad set of factors, including stock indices, currencies, commodities, bonds, volatility indices, risk and uncertainty indices, BTC-specific variables, and sentiment indicators, analysed with both linear and non-linear statistical and ML models.

The main goal of this work is to identify the primary drivers of Bitcoin volatility using various predictive models and to assess their effectiveness in improving ex-post volatility forecasts. The study employs a variety of statistical and ML models, which are compared in terms of both their ability to identify key volatility drivers and their forecasting accuracy for daily and weekly BTC volatility. Additionally, an analysis was conducted to determine whether data preprocessing, namely outlier replacing and standardisation, improves forecast accuracy. Our paper makes four significant contributions to the field of Bitcoin volatility analysis.

1) Unprecedented Range of Volatility Drivers: We develop the most extensive framework to date for out-of-sample forecasting of daily and weekly Bitcoin volatility, incorporating not only lagged BTC variances but also 33 diverse exogenous factors. These include BTC-specific market variables, broader financial market indicators, policy and uncertainty indices, and Google search trends. This scope substantially exceeds that of prior studies, which typically focused on a narrow set of predictors.
2) Enhanced Variable Representation and Preprocessing: We propose a systematic approach to representing each factor in multiple transformed forms (e.g., logarithms and differences, variances and returns) and evaluate the impact of two preprocessing techniques – standardisation and winsorisation – on forecasting accuracy. This process yields a total of 62 explanatory variables, offering a more comprehensive representation of potential drivers of BTC volatility. The combination of dual representation and preprocessing assessment, rarely explored in prior research, reveals transformation effects that materially influence predictive performance.
3) Multi-Method Variable Importance Assessment: We employ three complementary and methodologically distinct approaches – Bayesian model averaging (BMA), LASSO regression, and random forests (RF) – to identify the most influential determinants of Bitcoin volatility. This multi-method analysis captures both linear and nonlinear dependencies, providing a richer and more robust understanding than studies relying on a single modelling paradigm.
4) Empirical Benchmarking of Predictive Models: We present the first direct out-of-sample comparison of HAR, BMA, LASSO, and RF models for BTC volatility forecasting using an extensive predictor set. Our results demonstrate that LASSO and RF, by effectively managing high-dimensional predictor spaces and capturing extreme volatility, deliver significantly more accurate forecasts. These findings underscore the critical role of variable selection and modelling techniques in capturing the complex dynamics of Bitcoin volatility.

Our approach is distinctive in that, unlike previous studies which typically consider only a single time series for each explanatory factor and apply a single selected transformation, we systematically evaluate different representations of each variable. This is important, as it is not evident a priori which transformation of a given series most strongly influences Bitcoin volatility. Our results show that the choice of representation can materially impact the analysis and forecasting performance.

Furthermore, prior research has typically relied on a single method to assess variable importance, potentially biasing the identification of key predictors. An exception is the recent study by Feng, Qi, and Lucey [21], who employed multiple, albeit methodologically similar, regularization techniques, including LASSO, ridge regression, and elastic net. In contrast, our study incorporates three fundamentally different methods: BMA, LASSO, and RF, enabling a more comprehensive and robust assessment of volatility drivers.

Through these contributions, our paper aims to provide a more nuanced and thorough understanding of the factors driving Bitcoin volatility, potentially improving forecast accuracy and informing both academic research and practical applications in cryptocurrency markets.

For the selection of BTC volatility drivers, we deliberately chose models that share a key characteristic – a built-in mechanism for feature selection – while differing fundamentally in their methodological foundations. BMA offers a principled statistical approach to model averaging, explicitly accounting for model uncertainty while maintaining interpretability, which is particularly valuable for identifying key predictors in high-dimensional settings. LASSO provides embedded variable selection, effectively handling large sets of correlated regressors and mitigating overfitting, a common challenge in volatility forecasting. RF captures complex nonlinear relationships and interactions between predictors without imposing strong parametric assumptions. At each tree split, RF selects the variable that most effectively reduces impurity,

Table 1
Studies analysing factors of cryptocurrencies volatility.

Authors	Explanatory variables	Methods	Out-of-sample forecasts
Aalborg, Molnár, de Vries [22]	Trading volume, transaction volume, number of BTC addresses, VIX, Google searches for Bitcoin	Linear regression model	No
Al Guindy [23]	Investor attention from Twitter (X platform)	Panel regression, HAR, VAR	No
Alam, Amendola, Candila, Jabarabadi [24]	Monthly monetary aggregate M3 for U.S. and South Africa	GARCH, GJR, IGARCH, GARCH-MIDAS, structural break GARCH-MIDAS	No
Aslanidis, Bariviera, López [25]	Google Trends uncertainty index, Google Trends cryptocurrency index	Shannon and Rényi's transfer entropy	No
Aysan, Demir, Gozgor, Lau [26]	GPR	Bayesian graphical VAR, OLS and quantile-on-quantile regressions	No
Babalos, Bouri, Gupta [27]	Introduction of Spot Bitcoin ETFs, volatility of Grayscale Bitcoin Trust ETF	GARCH, wavelet coherence	No
Bakas, Magkonis, Oh [28]	Trading volume, market capitalisation, number of addresses, total circulation, miners revenue, Google searches for Bitcoin, consumer sentiment, US consumer confidence, MSCI World Index, S&P 500, Commodity Price Index, gold, oil, world economic activity, US industrial production, OECD inflation rate, USD index, US long and short term interest rate, global and US EPU, geopolitical risk, VIX, iTraxx Europe Swap Index	Dynamic Bayesian model averaging	No
Balcilar, Bouri, Gupta, Roubaud [29]	Trading volume	Granger causality, nonparametric causality in quantiles test	No
Benhamed, Messai, El Montasser [20]	Trading volume	General-to-specific modelling, asymmetric log ARCH specification	No
Będowska-Sójka, Górka, Hemmings, Zaremba [30]	EPU, GPR, VIX, OVX, GVZ, Emerging Markets ETF Volatility Index, EFA ETF Volatility Index, 20+ Year Treasury Bond ETF Volatility Index, USD index	DCC-GARCH, TVP-VAR, network analyses	No
Blau [31]	Speculative trading volume	Pearson and Spearman correlation, linear regression model, probit model	No
Bleher, Dimpfl [32]	Google searches for specific cryptocurrency names	Granger causality, VAR	Yes
Bourghelle, Jawadi, Rozin [33]	Crypto Fear and Greed Index, trading volume	Granger causality, linear and threshold VAR	No
Bourghelle, Jawadi, Rozin [34]	Crypto Fear and Greed Index, number of COVID-19 cases and deaths in the U.S., trading volume	Granger causality, linear and threshold VAR	No
Bouri, Gkillas, Gupta, Pierdzioch [35]	US–China trade tensions from Google searches	HAR, RF	Yes
Bouri, Kristoufek, Azoury [36]	S&P 500, EPU	GARCH with skewness and kurtosis, wavelet coherence	No
Bouri, Lau, Lucey, Roubaud [37]	Trading volume	Copula-Granger-causality in distribution test	No
Brauneis, Sahiner [38]	Sentiment from crypto market news data	HAR, MLP, LSTM, XGBoost, LightGBM, CNNBiLSTM	Yes
Candila [39]	Google searches for specific cryptocurrency names	Double asymmetric GARCH-MIDAS	Yes
Conlon, Corbet, McGee [40]	Expected and unexpected trading volume, VIX, EPU	Regression models	No
Conrad, Custovic, Ghysels [41]	Trading volume, volatility for S&P Global Luxury Index and S&P 500; VIX, Variance Risk Premium, Google searches for Bitcoin, Baltic exchange dry index, Chinese Yuan	GARCH-MIDAS	No
Corbet, McHugh, Meegan [42]	Monetary policy announcements	GARCH	No
Di, Xu [43]	BTC implied volatility, VIX, aggregate implied volatility indices for emerging (VXEEM) and developing markets (VXEFA), GVZ, OVX, USD index	VAR, generalised forecast error variance decomposition	No
Dias, Fernando, Fernando [44]	Google searches, Wikipedia page views, daily number of merits shared on Bitcointalk.org, positivity and negativity from news headlines, Twitter happiness index, VIX, trading volume, returns of Dow Jones Index, EUR/USD, gold, hash rate	PCA, Granger causality, moments quantile regression	Yes
Ding, Wu, Cui, Goodell, Du [45]	Climate policy uncertainty	GARCH, genetic programming	Yes
Dyhrberg [46]	Federal funds rate, gold cash, gold futures, FTSE, USD/EUR, USD/GBP	GARCH, EGARCH	No
Eom, Kaizoji, Kang, Pichl [47]	Google searches for Bitcoin	AR-X	No
Elsayed, Gozgor, Lau [48]	Returns of oil, gold, S&P 500, S&P 500 bond index, USD index, US EPU, TEU, VIX	TVP-VAR, dynamic connectedness approaches, network analyses, causality-in-variance LM test	No
Elsayed, Gozgor, Yarovaya [49]	Cryptocurrencies CRIX index, policy UCRY, price UCRY, EPU, gold, VIX	TVP-VAR	No
Fang, Bouri, Gupta, Roubaud [50]	EPU	GARCH-MIDAS, DCC-MIDAS	No
Fang, Su, Yin [51]	News-based implied volatility index NVIX, global EPU, financial uncertainty FU	GARCH-MIDAS, DCC-MIDAS	No
Feng, Qi, Lucey [21]	Global and US EPU, GPR, US CPI and PPI, OVX, GVZ, federal funds rate, USD index, WTI and Brent oil, Google searches for Bitcoin, number of blocks, average transaction fee, difficulty, block size, hash rate, mining profitability, trading volume	HAR, LASSO, RR, EN	Yes
Figá-Talamanca, Patacca [52]	Trading volume, Google searches for Bitcoin	GARCH, EGARCH	No

(continued on next page)

Table 1 (*continued*)

Authors	Explanatory variables	Methods	Out-of-sample forecasts
Gbadebo, Adekunle, Adedokun, Lukman, Akande [53]	Trading volume, BTC market capitalisation, Morgan Stanley Capital International All Country World Index, Google searches for Bitcoin	ARDL	No
Ghani, Ghani, Ali, Mustafa, Kosar [54]	EPU, GPR, TEU, Twitter market uncertainty, trade policy uncertainty	GARCH-MIDAS	Yes
Gkillas, Tantoula, Tzagarakis [55]	Number of transactions	Hybrid model of HAR and RF	Yes
Güler [56]	Trading volume, Crypto Fear & Greed Index, American Association of Individual Investors Index	GARCH, CGARCH, EGARCH, GJR-GARCH, and AP-ARCH	No
Kristjanpoller, Minutolo [57]	Seven technical analysis indicators	Hybrid MLP-GARCH model, GARCH, EGARCH, APGARCH	Yes
Kristoufek [58]	Trading volume, Google searches for Bitcoin and BTC, on-chain transfers volume in BTC, number of active addresses, market capitalisation of the largest fiat-backed stablecoins, amount of emitted BTC, average amount of hashes being solved, prices of S&P 500 futures and gold futures, VIX	Linear regression model	No
Kufo, Gjeci, Pilkati [59]	Trading volume, returns of MSCI All Country World Index, Google searches for specific cryptocurrency names, USD/EUR	GARCH	No
Kyriazis, Papadamou, Tzeremes, Corbet [60]	Four TEU and four TMU indices	Nonlinear quantile causality in volatility	No
Lehrer, Xie, Yi [61]	Sentiment index from Twitter	LASSO, regression tree, boosted trees, bagged trees, RF, SVR, least squares SVR, HAR, HAR-J, HAR-CJ, HAR-RS	Yes
Liang, Zhang, Li, Ma [62]	GVZ, VIX, Google searches for Bitcoin, EPU, GPR	GARCH-MIDAS	Yes
Long, Chatziantoniou, Gabauer, Lucey [63]	Investor sentiment from Reddit	Asymmetric TVP-VAR	No
Long, Xie, Zhou, Lucey, Urquhart [64]	Investor sentiment from Reddit, Twitter	Granger causality, TVP-VAR,	No
López-Cabarcos, Pérez-Pico, Piñeiro-Chousa, Šević [65]	Returns of S&P 500, VIX, investor sentiment from Stocktwits messages	GARCH, EGARCH	No
Lyócsa, Molnár, Plíhal, Širaňová [66]	News and sentiment about cryptocurrency regulation, the hacking of cryptocurrency exchanges, scheduled macroeconomic news announcements	HAR, quantile regression HAR	No
Maghyereh, Abdoh [67]	RV of future contracts for S&P 500, USD/EUR, and US 10-year T-note futures, gold and oil.	Granger causality connectedness measure, wavelet coherence, dynamic frequency-domain connectedness	No
Mandaci, Cagli [68]	Herding intensity measures	Granger causality with a Fourier approximation	No
Mokni [69]	EPU	Symmetric and asymmetric causality in quantiles tests	No
Omura, Cheung, Su [70]	RV of natural gas futures	HAR	Yes
Nguyen, Nguyen, Nguyen, Pham, Nguyen [71]	Federal funds rate, Chinese interbank rate, USD/EUR and USD/GBP exchange rates, returns of FTSE, futures and spot gold returns	GARCH, EGARCH, fixed effects models	No
Nouir, Hamida [72]	Trading volume, US and China EPU and GPR, oil, returns of S&P 500, EUR/USD	ARDL, quantile regression	No
Papadamou, Kyriazis, Tzeremes [73]	EPU, gold	Granger causality, nonparametric causality in quantiles test	No
Sabah [74]	Venues that accept cryptocurrencies as a payment method, type of business, continent of location, market capitalisation, returns, VIX	Pearson correlation, linear regression model, Granger causality, impulse response function	No
Said, Somasuntharam, Yaakub, Sarmidi [75]	Twitter data, Google searches for specific cryptocurrency names	Granger causality, impulse response function, VAR	No
Salisu, Ndako, Vo [76]	Brent and WTI oil prices	Feasible quasi-generalised least square model	Yes
Salisu, Ogbonna [77]	Google searches for cryptocurrency	GARCH-MIDAS	Yes
Sapkota [78]	Newspaper-based sentiment from the LexisNexis database, Google searches for Bitcoin	HAR	Yes
Seo, Kim [79]	Google searches for Bitcoin, VIX	GARCH, EGARCH, GJR, hybrid models of GARCH and ANN, HONN	Yes
Shen, Urquhart, Wang [80]	Tweets for Bitcoin	Linear and nonlinear Granger causality, VAR	No
Smales [81]	Google searches for specific cryptocurrency names, EPU, VIX, UCRY, Aruoba-Diebold-Scotti business conditions index, term premium computed from 2- and 10-year yields, Treasury bill – Eurodollar spread	Panel regression models, panel causality, Granger causality	No
Teterin, Peresetsky [82]	Google searches for Bitcoin and keywords related to Bitcoin, uncertainty and risk	HAR	Yes
Tzeng, Su [83]	S&P 500, P/B and D/Y ratios, 3-months and 10-year yields, federal funds rate, commercial paper-Treasury bill spread, term spread, VIX, credit risk spread, U.S. EPU, TED spread, macroeconomic uncertainty index, M1 money supply, consumer confidence index, leading economic index, industrial production and its volatility, capacity utilization rate, Purchasing Managers' Indices in the manufacturing and services sectors, nonfarm payrolls, unemployment rate, CPI, volatility of PPI, U.S. exports and imports, trade balance, retail sales, CRB index, oil, gold, MSCI All Country World Index, global EPU, Baltic exchange dry index, OECD export, import values, OECD CPI and M1 growth, OECD G20 leading indicator, consumer opinion surveys	Linear regression model	Yes
Urquhart [84]	Google searches for Bitcoin	Granger causality, VAR	No
Uzonwanne [85]	FTSE 100, S&P 500, CAC 40, DAX 30, Nikkei 225	VARMA-AGARCH	No

(*continued on next page*)

Table 1 (*continued*)

Authors	Explanatory variables	Methods	Out-of-sample forecasts
Walther, Klein, Bouri [86]	Returns and RV for S&P 500, MSCI Emerging Markets 50, Dow Jones Precious Metals, oil; VIX, global financial stress index, global and Chinese EPU, global real economic activity, USD index	GARCH-MIDAS	Yes
Wan, Song, Zhang, Yin [87]	Oil, diesel, heating oil, jet fuel, natural gas, propane, conventional gasoline, regular gasoline	Generalised forecast error variance decomposition based on volatility impulse response function from block DECO-GARCH	No
Wang, Andreeva, Martin-Barragan [88]	Close, open, high, low prices, trading volume, trading count, block size, transactions per block, payments per block, confirmation time, hash rate, difficulty, price and trading volume for oil, gold, silver, DJI, S&P 500, Nasdaq, Russell 2000, CNY/USD, USD/Euro, US EPU, Google searches for Bitcoin, blockchain and cryptocurrency, Crypto Fear & Greed Index	GARCH, RF, LSTM	Yes
Wang, Ma, Bouri, Guo [19]	Returns and RV of: S&P 500, emerging market index, global commodity index, USD index; VIX, OVX, global and China EPU, global real economic activity index, GPR, geopolitical threats index, geopolitical acts index, Google searches for Bitcoin, technical indicators: moving-average, momentum, on-balance volume.	AR-X, PCA, PLS, LASSO, EN forecast combination	Yes
Wu, Ho, Wu [89]	Global EPU, 14 developed and 7 emerging markets EPU	GARCH-MIDAS	No
Wu, Hossain, Zhang [90]	EPU, oil, Nasdaq, gold, number of active addresses, block size, mining difficulty, hash rate, trading volume, market value	Liner regression model	No
Wu, Yin, Umar, Iqbal [91]	Oil	GARCH, CARR, VS-CARR, VS-ACARR	Yes
Xia, Sang, He, Wang [92]	EPU, UCRY	GARCH-MIDAS, asymmetric GARCH-MIDAS	Yes
Yen, Cheng [93]	Chine, US, Japan, Korea EPU	Linear regression model	No
Yin, Nie, Han [94]	Returns of oil, RV of oil, oil realised skewness, oil supply shocks, global aggregate demand for industrial commodities, oil market demand shocks, returns of gold, VIX, macroeconomic uncertainty index, financial market uncertainty index, EPU	GARCH-MIDAS	No
Yousaf, Ali, Marei, Gubareva [95]	Gold, silver, palladium, platinum, aluminium, copper, lead, nickel, tin, zinc	DCC-GARCH, TVP-VAR	No
Yu [96]	EPU	HAR, HAR-CJ, HAR-CJ with leverage effect	Yes
Zhang, Wang [97]	Google searches for specific cryptocurrency names	Linear and nonlinear Granger causality, quantile regression	No
Zhou [98]	Trading volume, number of transactions, hash rate, miners' revenue, number of news from Reuters	EGARCH, TGARCH, NGARCH, AVGARCH, APARCH	No
Zhou, Xie, Wang, Gong, Zhu [99]	EPU, GPR, VIX, OVX, RV for: S&P 500, U.S. dollar index, U.S. 30-year T-bond, gold, oil, natural gas	AR, ARIMA, HAR, GPR, RF, LightGBM, LSTM, LSTNet, MTGNN, EMGNN	Yes
Zhu, Zhang, Wu, Zheng, Zhang [100]	Google searches for Bitcoin	Granger causality, VAR, VAR with higher moments and asymmetrical effects	Yes

RV - realised variance, VIX – CBOE Volatility Index, GVZ - CBOE Gold ETF Volatility Index, OVX - Crude Oil ETF Volatility Index, MSCI - Morgan Stanley Capital International, EPU - economic policy uncertainty, GPR - global geopolitical risk, UCRY - cryptocurrency uncertainty, TEU - Twitter economic uncertainty, TMU - Twitter market uncertainty, CPI - consumer price index, CRB - Commodity Research Bureau, TED - Treasury–Eurodollar, AR-X - autoregressive model with exogenous variables, ARDL - autoregressive distributed lag, VAR – vector autoregression, VARMA - Vector Autoregressive Moving Average, TVP-VAR - time-varying parameter vector autoregression, GARCH - generalised autoregressive conditional heteroscedasticity, EGARCH - exponential GARCH, TGARCH - threshold GARCH, GJR – Glosten Jagannathan Runkle GARCH, AGARCH - asymmetric GARCH, CGARCH - component GARCH, APARCH - asymmetric power ARCH, NGARCH - nonlinear GARCH, AVGARCH - absolute value GARCH, DCC - dynamic conditional correlation, DECO-GARCH - dynamic equicorrelation GARCH, GARCH-MIDAS - GARCH mixed-data sampling, VS-ACARR - volatility-spillover-asymmetric conditional autoregressive range, HAR - heterogeneous autoregressive, HAR-J - HAR with jump component, HAR-CJ - HAR with continuous jump component, HAR-RS – HAR with semi-variance component, PCA - principal component, PLS - partial least squares, LASSO - least absolute shrinkage and selection operator, RR – ridge regression, EN - elastic net, GPR - Gaussian process regression, RF – random forest, ANN - artificial neural network, HONN - higher order neural network, MLP - Multilayer Perceptron, LSTM - Long Short-Term Memory, MGNN - Multivariate Graph Neural Network EMGNN - Evolving Multiscale Graph Neural Network, LSTNet - Long- and Short-Term Time-Series Network, XGBoost - Extreme Gradient Boosting, LightGBM - Light Gradient Boosted Machine, CNNBiLSTM – hybrid CNNbidirectional LSTM.

thereby ranking predictors by importance in a data-driven manner. By combining these three models, we are able to: (1) benchmark predictive performance across fundamentally different modelling paradigms, (2) identify consistent volatility drivers that are robust across linear, regularized, and nonlinear approaches, and (3) evaluate whether nonlinear ML techniques can outperform or complement traditional statistical models in terms of forecasting accuracy. This integration of diverse modelling strategies remains relatively unexplored in the Bitcoin volatility literature, making the comparative analysis itself a novel and valuable contribution of our study.

The remainder of the paper is organised as follows: Section 2 reviews the literature on the drivers of cryptocurrency volatility. Section 3 presents the methodologies employed and describes our research procedure. Section 4 provides a description of the data. Section 5 presents and discusses the results of empirical research, and Section 5.4 concludes the paper.

2. Literature review

We review studies that aim to explain the volatility of cryptocurrencies, excluding those that focus solely on cryptocurrency returns. Table 1 summarises 82 papers, detailing the explanatory variables and methodologies employed. The most commonly analysed potential drivers of volatility include stock indices, fiat currencies, oil, gold, attention and sentiment measures, uncertainty indices, and trading volume. The predominant methods used are the generalised autoregressive conditional heteroscedasticity (GARCH) models, the HAR model, linear regression model, and Granger causality analysis. Notably, most of these studies (55 papers) assess the impact of explanatory variables on cryptocurrency volatility using in-sample analysis, without considering out-of-sample forecasts. However, all of these papers employ only statistical methods. For out-of-sample forecasting, nonliner ML models are used only in 9 studies, while 18 studies rely solely on statistical methods.

Table 2
Description of applied variables.

Variable	Abbreviation	Source	Economic justification
Realised variance of Bitcoin	RVBTC	api.kraken.com	Explained variable
BTC-specific factors			
Trading volume	Volume	finance.yahoo.com	Associated with increased information flow and investor activity
Number of transactions per day	transactions	data.nasdaq.com	
Average block size per day	Block	data.nasdaq.com	Reflects blockchain usage intensity and potential congestion effects on transaction costs
Market capitalisation	Capital	data.nasdaq.com	Influences on liquidity and stability and perceived market risk.
Average hash rate per day	Hash	data.nasdaq.com	Represents network security and miner confidence, affecting long-term trust in BTC
Number of unique addresses per day	addresses	data.nasdaq.com	Serves as a proxy for user base growth and diversification of market participants
Google searches for Bitcoin	Google	trends.google.com	Serve as a proxy for rising investor attention and uncertainty
Financial markets			
S&P 500	S&P	Refinitiv Eikon	Reflect changes in the most important capital markets
Nasdaq Composite	Nasdaq	Refinitiv Eikon	
Euro STOXX 50	STOXX	Refinitiv Eikon	
FTSE 100	FTSE	Refinitiv Eikon	
Nikkei 225	Nikkei	Refinitiv Eikon	
Shanghai Composite	SSE	www.investing.com	
EUR/USD	EUR	www.investing.com	Reflect shifts in macroeconomic fundamentals and investor sentiment across global regions
JPY/USD	JPY	www.investing.com	
CNY/USD	CNY	www.investing.com	
Nominal broad U.S. dollar index	USDX	fred.stlouisfed.org	
U.S. 2-years bond yield	US2Y	www.investing.com	Express market expectations regarding interest rates and the broader economic outlook
U.S. 10-years bond yield	US10Y	www.investing.com	
NYMEX light sweet WTI crude oil	Oil	Refinitiv Eikon	Reflect shifts in global demand, supply shocks, and inflationary pressures, all of which can alter risk perceptions.
Gold spot	Gold	www.investing.com	
NYMEX Henry Hub natural gas	Gas	Refinitiv Eikon	
Bloomberg Commodity Index	BCOM	Refinitiv Eikon	
Market and policy uncertainty			
CBOE Volatility Index	VIX	www.investing.com	Measure implied volatility in key equity and commodity markets, serving as benchmarks for investor fear
EURO STOXX 50 Volatility Index	VSTOXX	Refinitiv Eikon	
CBOE Gold ETF Volatility Index	GVZ	www.cboe.com	
Economic policy uncertainty index based on newspapers	EPU	www.policyuncertainty.com	Capture macroeconomic, geopolitical, and health-related uncertainty as well as shifts in investor sentiment and risk aversion
Twitter-based market uncertainty index	TMU-SCA	www.policyuncertainty.com	
Twitter-based economic uncertainty index	TEU-SCA	www.policyuncertainty.com	
Geopolitical risk index based on newspapers	GPR	www.policyuncertainty.com	
Risk aversion index based on financial variables	RA_BEX	www.nancyxu.net	
Uncertainty index based on financial variables	UNC_BEX	www.nancyxu.net	
Infectious disease equity market volatility tracker based on newspapers	infectious	www.policyuncertainty.com	

Research on the accuracy of out-of-sample volatility forecasts predominantly focuses on a single or a very narrow set of selected factors. The explanatory variables investigated include Google search trends [32,35,39,77–79,82,100]. Twitter-based sentiment index [61], newspaper-based sentiment derived from the LexisNexis database [78], sentiment from crypto market news data [38], VIX [79], economic policy uncertainty (EPU) [90,96], climate policy uncertainty [45], cryptocurrency uncertainty (UCRY) [90], the number of BTC transactions [55], RV of natural gas futures [70], and Brent and WTI oil prices [76,91], technical analysis indicators [57].

Studies that incorporate a wide range of factors influencing Bitcoin volatility forecasts are relatively rare. Walther, Klein, and Bouri [86] utilise factors such as the returns and realised variance of the S&P 500, MSCI Emerging Markets 50, Dow Jones Precious Metals, oil, the VIX Volatility Index, the global financial stress index, global and Chinese EPU, global real economic activity, and the USD index. In contrast, Liang, Zhang, Li, and Ma [62] consider variables including the CBOE Gold ETF Volatility Index (GVZ), VIX, Google searches for Bitcoin, EPU, and global geopolitical risk (GPR). Meanwhile, Ghani, Ghani, Ali, Mustafa, and Kosar [54] incorporate EPU, GPR, Twitter economic uncertainty (TEU), Twitter market uncertainty, and trade policy uncertainty as regressors. All three studies employ the generalised autoregressive conditional heteroscedasticity - mixed-data sampling (GARCH-MIDAS) model. However, rather than including all factors in a single model, each study builds separate models for each exogenous variable. Tzeng and Su [83] use a linear regression model and 28 U.S. and 12 global variables to predict the monthly volatility of BTC, Cardano, Dogecoin, Ethereum, Litecoin, and XRP. Similarly to the above analyses, they apply only one variable per model.

Research that jointly considers a wide range of potential drivers within a single model is scarce. Zhou, Xie, Wang, Gong, and Zhu [99] apply EPU, GPR, VIX, OVX, and RV for S&P 500, the U.S. dollar index, the U.S. 30-year T-bond, gold, oil, natural gas to predict daily volatility of BTC. They use the following methods: AR, ARIMA, HAR, GPR, RF, LightGBM, LSTM, LSTNet, MTGNN, and EMGNN. Wang, Ma, Bouri, and Guo [19] utilise 17 economic variables and 3 technical strategies to forecast the monthly volatility of BTC. They employ several methods, including the autoregressive model with exogenous variables (AR-X), the principal component analysis (PCA), the partial least squares (PLS) model, LASSO, EN, and forecast combination methods such as mean, median, trimmed mean, and discounted mean-squared prediction error (DMSPE). Feng, Qi, and Lucey [21] use 19 economic variables and indicators to forecast daily volatility for BTC and Ethereum, applying four methods: HAR, LASSO, RR, and EN. The most comprehensive study is conducted by Wang, Andreeva, and Martin-Barragán [88], who use 27 external and 6 internal determinants to forecast the daily, weekly, and monthly volatility of BTC, Ethereum, Litecoin, and Ripple. Their analysis is divided into two parts: one focusing on internal variables (using lagged and moving averages), and another on external drivers. They apply GARCH, RF, and long short-term memory (LSTM).

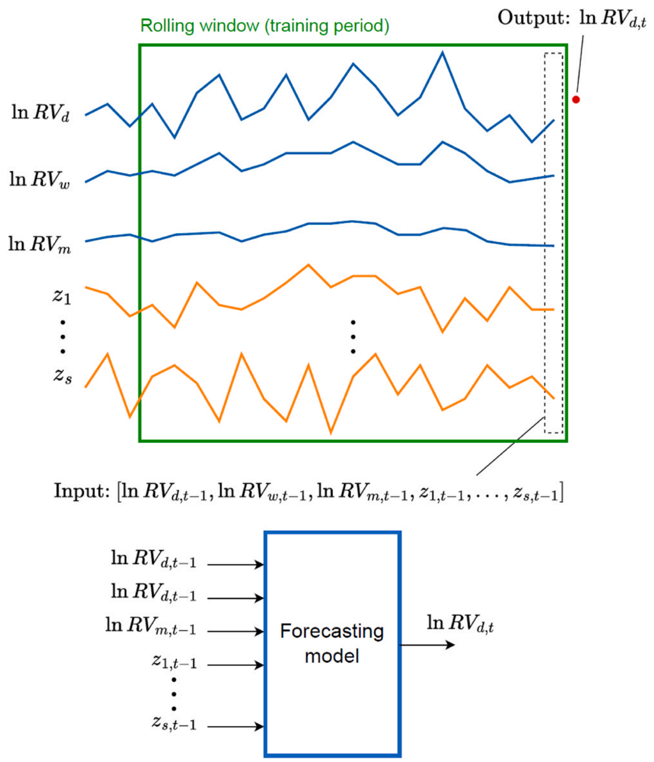

Fig. 1. The input and output data used in the forecasting models with exogenous variables.

3. Applied methods and forecasting procedure

3.1. Realised variance models

Our models rely on realised variance (RV), which is derived from intraday price data:

$$RV_{d,t} = \sum_{k=1}^{K} r_{k,t}^2, \tag{1}$$

where $r_{k,t}$ is the intraday return, K is the number of intraday observations during a day.

We use RV because models based on this measure consistently yield more accurate forecasts compared to those that rely solely on daily closing prices [15]. As explanatory variables, we also use weekly and monthly average realised variances defined as:

$$RV_{w,t} = \frac{RV_{d,t-4} + RV_{d,t-3} + RV_{d,t-2} + RV_{d,t-1} + RV_{d,t}}{5}, \tag{2}$$

$$RV_{m,t} = \frac{RV_{d,t-21} + RV_{d,t-20} + \ldots + RV_{d,t}}{22} \tag{3}$$

Our benchmark model is HAR introduced by Corsi [14]. It combines three volatility components which describe daily, weekly and monthly volatility. These components reflect the behaviour of investors with different time horizons. We use the log transformation of the model which guarantees positive predictions of volatility:

$$\ln RV_{d,t} = \gamma_0 + \gamma_1 \ln RV_{d,t-1} + \gamma_2 \ln RV_{w,t-1} + \gamma_3 \ln RV_{m,t-1} + \varepsilon_t \tag{4}$$

Despite its simplicity, the HAR model turns out to be very effective in forecasting variance of financial returns and is widely used in empirical research.

To investigate the influence of exogenous variables on RV predictions, we select three models with built-in mechanisms for variable selection, allowing us to evaluate variable importance: BMA, LASSO, and RF. These models predict $\ln RV_{d,t}$ at time t based on the lagged RV variables (1)-(3) and exogenous variables (see Table 2). The lag is $\tau = 1$ for daily horizon predictions, or $\tau > 1$, for predictions with horizons longer than one day. In addition to prediction, these models allow us to determine the relative importance of variables. The general form of these models can be expressed as:

$$\ln RV_{d,t} = f\left(\ln RV_{d,t-\tau}, \ln RV_{w,t-\tau}, \ln RV_{m,t-\tau}, z_{1,t-\tau}, \ldots, z_{s,t-\tau}\right), \tag{5}$$

where $z_{i,t-\tau}$ denotes the value of the i-th exogenous variable at time $t-\tau$, f represents the functional form specific to each model (BMA, LASSO, or RF), and s is the number of exogenous variables.

Fig. 1 illustrates the models with exogenous variables for the daily horizon prediction scenario.

In the empirical part of the work, we also consider BMA, LASSO and RF models without exogenous variables to assess the overall impact of these variables on the results.

In the subsequent three subsections, we provide detailed descriptions of the models.[2] For clarity, we use the following notation: $\mathbf{x}$ represents the vector of explanatory variables (containing only RV or RV and exogenous variables depending on the forecasting scenario), and y represents the scalar dependent variable ($\ln RV_{d,t}$).

3.2. BMA method

Standard regression models often fail to incorporate the uncertainty surrounding the model structure and variable selection. Bayesian Model Averaging (BMA) addresses this limitation by estimating a model for all possible combinations of explanatory variables and presenting the final model as a weighted average of these models. The concept, first introduced by Leamer [101], and further developed by Mitchell and Beauchamp [102], Raftery, Madigan, and Hoeting [103], Hoeting, Madigan, Raftery, and Volinsky [104], highlights the advantage of accounting for model uncertainty instead of relying on a single "best" model.

By considering a wide range of model specifications and weighting them according to their posterior probabilities, BMA improves robustness and reduces the risk of omitting relevant factors or overemphasizing less important ones. For a linear regression model:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \ldots + \beta_k x_{k,t} + \varepsilon_t, \quad \varepsilon_t \sim N\left(0, \sigma^2 \mathbf{I}\right) \tag{6}$$

where β_0, β_1,..., β_k, are the model coefficients, k is the number of potential variables $x_{1,t}$, $x_{2,t}$, ..., $x_{k,t}$ and ε_t is normal IID error term with variance σ^2; there are 2^k possible sets of variables (models), denoted as M_j, for $j = 1, 2, \ldots, 2^k$.

The BMA uses posterior probability $P(M_j|D)$ as weights, with D representing the set of all data inputs. It is given by the formula:

$$P(M_j|D) = \frac{P(D|M_j)P(M_j)}{P(D)} = \frac{P(D|M_j)P(M_j)}{\sum_{i=1}^{2^k} P(D|M_i)P(M_i)} \tag{7}$$

It is, therefore, the ratio of its marginal probability $P(D|M_j)$ multiplied by the prior model probability $P(M_j)$, for which M_j is the relevant model, to the sum of the marginal probabilities over the entire model space. The prior on parameters is commonly set using Zellner's g-prior [105], with coefficients drawn from a normal distribution with mean 0 and variance proportional to g.

The main output of BMA is the posterior inclusion probability (PIP) for each variable, defined as the sum of posterior model probabilities across all models including that variable. PIPs indicate the relative importance of variables in prediction, while their total sum reflects the average model size.

[2] The code of our models is available at the following address: https://github.com/GMDudek/Bitcoin_drivers.

3.3. LASSO model

LASSO [106] is a regularisation method for linear regression that enhances the model's predictability and interpretability. Like other regularisation techniques, such as ridge regression [107,108], it works by shrinking the coefficient estimates towards zero, which can considerably reduce their variance [109,110]. This process reduces model complexity and helps avoid overfitting or collinearity. Unlike ridge regression, however, LASSO can also perform variable selection by forcing some coefficients to be exactly zero. LASSO regression achieves this by adding an L_1 penalty term $\sum_{j=1}^{k} |\beta_j|$ to the residual sum of squares (RSS). Thus, the LASSO coefficients minimise the quantity:

$$RSS_{L_1} = \sum_{t=1}^{N} (y_t - \hat{y}_t)^2 + \lambda \sum_{j=1}^{k} |\beta_j| \tag{8}$$

where N is the number of observations and $\hat{y}_t$ are the fitted values.

The hyperparameter $\lambda \geq 0$ controls the extent of shrinkage: when $\lambda = 0$, LASSO produces the same coefficients as ordinary least squares (OLS) regression, and when λ is very large, all coefficients are shrunk to zero. When λ is sufficiently large, the L_1 penalty results in some coefficient estimates being exactly equal to zero. This feature allows the LASSO model to select the most important predictors that influence the response variable.

LASSO has demonstrated good empirical performance across various contexts. When the true model is sparse (i.e. contains only a few significant terms) and the number of predictors (k) is large relative to the number of observations (N), LASSO often outperforms criteria like AIC and BIC, as well as stepwise methods and ridge regression, in predictive metrics such as mean squared error [111].

3.4. Random Forest

The RF model [112] is a robust and widely used ML algorithm for regression and classification tasks due to its ability to model nonlinear relationships and complex feature interactions. It constructs an ensemble of decision trees, each trained on a different bootstrap sample and using a random subset of predictors (random subspace method). This combination of bagging and feature randomness mitigates overfitting and enhances model generalisation.

The output of the RF is the weighted average of predictions from all trees, producing a reliable estimate of the conditional mean even under noisy and uncertain conditions:

$$f(\mathbf{x}) = \sum_{t=1}^{N} w_t(\mathbf{x}) y_t, \tag{9}$$

$$w_t(\mathbf{x}) = \frac{1}{p} \sum_{j=1}^{p} \frac{\mathbb{I}\{\mathbf{x}_t \in \ell_j(\mathbf{x})\}}{\sum_{\tau=1}^{N} \mathbb{I}\{\mathbf{x}_\tau \in \ell_j(\mathbf{x})\}}, \tag{10}$$

where p denotes the total number of trees, $\ell_j(\mathbf{x})$ denotes the leaf that is obtained when dropping $\mathbf{x}$ down the j-th tree, and $\mathbb{I}$ denotes the indicator function.

Trees are built by selecting optimal split predictors and thresholds to minimise a criterion such as mean squared error (MSE). The key hyperparameters include: number of trees in the forest, p, minimum leaf size (or equivalent parameter determining the tree size), represented as q, and the number of predictors randomly chosen for each split, indicated as r. These control the bias-variance trade-off.

RF provides two methods for estimating predictor importance. The first method, denoted as RF1, involves using out-of-bag observations (observations not included in the bootstrap sample) and permuted predictors. For each predictor, the importance is determined by measuring the increase in prediction error (MSE) when the values of that predictor are randomly permuted across the out-of-bag observations. This calculation is performed for every trained tree, then averaged across the ensemble and divided by the standard deviation of the entire ensemble.

The second method, denoted as RF2, calculates importance based on the improvement in the split criterion (MSE) at each split in each tree. This measure is then averaged over all trees in the forest separately for each predictor. Consequently, predictors that effectively decrease data variability post-splits receive a higher importance score.

3.5. Forecasting procedure

In this section, we apply the predictive models discussed in Sections 3.2–3.4, i.e. BMA, LASSO, and RF. For each model, we consider two general specifications (without and with exogenous variables):

$$\ln RV_{d,t} = f(\ln RV_{d,t-\tau}, \ln RV_{w,t-\tau}, \ln RV_{m,t-\tau}), \tag{11}$$

and

$$\ln RV_{d,t} = f(\ln RV_{d,t-\tau}, \ln RV_{w,t-\tau}, \ln RV_{m,t-\tau}, z_{1,t-\tau}, \ldots, z_{s,t-\tau}), \tag{12}$$

where $\tau = 1, 2, \ldots, 5$ is the prediction horizon and $z_{1,t-\tau}, \ \ldots, \ z_{s,t-\tau}$ denote $s = 59$ explanatory variables described in Section 4.

As a baseline model, we use HAR (described in Section 3.1) due to its relative simplicity and widespread popularity in financial studies for forecasting volatility. The HAR model is of the form (11), where f is the simplest, i.e. linear function. The parameters of this model are estimated by ordinary least squares.

To evaluate the effectiveness of the models, we generate out-of-sample forecasts for both one-day ahead and one-week ahead volatility. Based on models (12), we examine whether explanatory variables improve the accuracy of volatility forecasts. Moreover, we identify which of these variables are the most important determinants of Bitcoin volatility.

Furthermore, we investigate the effects of data preprocessing on the forecasting accuracy of our models. Specifically, we consider two types of data transformations: the replacement of outliers and standardisation. Both transformations are commonly used in data science to help enhance the reliability and performance of ML models. The outliers, defined as data points lower than $Q_1 - 1.5 \times IQR$ or higher than $Q_3 + 1.5 \times IQR$ are replaced by these lower and upper threshold values, respectively, where Q_1 and Q_3 are quartiles and IQR is the interquartile range. Standardisation is a well-known statistical technique involving subtracting the mean and dividing by the standard deviation. We also apply a combination of both transformations: first, we replace outliers, and then we standardise the resulting data. All the aforementioned transformations are applied to the exogenous variables.[3]

Finally, we consider four models without exogenous variables of the form (11):

- HAR, BMA, LASSO and RF,
  and twelve models with exogenous variables of the form (12), denoted as:
- BMA-X, LASSO-X and RF-X (models without any additional transformations)
- BMA-X_out, LASSO-X_out and RF-X_out (models with outliers replaced)
- BMA-X_st, LASSO-X_st and RF-X_st (models with standardised data)
- BMA-X_out_st, LASSO-X_out_st and RF-X_out_st (models with both outliers replaced and standardised data).

For the daily forecasts (i.e. for $\tau = 1$), we utilise a rolling window

[3] In addition, we also tried applying the aforementioned transformations to RV variables as well, but this did not lead to an improvement in forecast accuracy.

approach. Our initial training sample covers the period from August 2, 2017, to December 31, 2019 (August 1, 2017 is not included in this range because in case of first-differenced data first observation comes from August 2, 2017). We train the models using this training sample and then generate one-day-ahead forecasts for January 1, 2020. To keep the training sample up to date, we add one new observation and remove the oldest one, thereby maintaining a rolling window. With the updated training sample, we retrain the models and generate forecasts for January 2, 2020. This procedure is repeated, providing daily forecasts from January 1, 2020, to March 31, 2023. This results in a total of 848 daily forecasts for each method.

For the weekly forecasts, we adopt a similar approach but generate forecasts only once a week on Fridays. Every Friday, we forecast the volatility for the upcoming Monday through Friday (i.e. for τ from 1 to 5), and then calculate the weekly forecast as the sum of these five daily forecasts. This process results in a total of 169 weekly forecasts for each method.

The flowchart of our study is depicted in Fig. 2. The algorithm for generating forecasts with a given prediction horizon τ is presented in Algorithm 1.

Algorithm 1. Forecasting framework with prediction horizon τ.

Algorithm 1: Forecasting framework with prediction horizon τ.

Input:

- Data: $RV_{d,t}$, $RV_{w,t}$, $RV_{m,t}$, and $s = 59$ exogenous variables, $z_{1,t}, \dots, z_{s,t}$, from August 1, 2017 to March 31, 2023
- Models: $\mathcal{M}'$ – with lagged RV variables (11), $\mathcal{M}''$ – with lagged RV and exogenous variables (12)
- Test period: from T_1 (January 1, 2020) to T_2 (March 31, 2023)
- Prediction horizon τ

Output:

Forecasts of daily realised variances $\widehat{RV}_{d,t}$ for the test period with a prediction horizon τ

Procedure:

1. Select transformation method for exogenous variables:
- No transformation
- Outliers replacement
- Standardisation
- Combination of outliers replacement and standardisation

2. **For** $T = T_1$ to T_2 **do**

2.1. Construct training sets using a rolling window approach:
- For model $\mathcal{M}'$:
$\{[lnRV_{d,t}, lnRV_{w,t}, lnRV_{m,t}], lnRV_{d,t+\tau}\,; t = T - T_1 + 1, T - T_1 + 2, \dots, T - \tau\})$
- For model $\mathcal{M}''$:
$\{[lnRV_{d,t}, lnRV_{w,t}, lnRV_{m,t}, z_{1,t}, \dots, z_{s,t}], lnRV_{d,t+\tau}\,; t = T - T_1 + 1, T - T_1 + 2, \dots, T - \tau\}$

2.2. **If** $T == T_1$ or modulo$(T, 25) == 0$ **then**
Optimize hyperparameters for LASSO and RF

2.3. Train predictive models $\mathcal{M}'$ and $\mathcal{M}''$

2.4. Generate τ-day ahead forecast:
- Using model $\mathcal{M}'$:
$ln\, \widehat{RV}_{d,T} := f(ln\, RV_{d,T-\tau}, ln\, RV_{w,T-\tau}, ln\, RV_{m,T-\tau})$
- Using model $\mathcal{M}''$:
$ln\, RV_{d,T} := f(ln\, RV_{d,T-\tau}, ln\, RV_{w,T-\tau}, ln\, RV_{m,T-\tau}, z_{1,T-\tau}, \dots, z_{s,T-\tau})$

We employed distinct approaches to optimise hyperparameters for the LASSO and RF models. For LASSO, we selected the regularisation parameter λ using a 10-fold cross-validation procedure, which helps balance model complexity and performance. In the case of RF, we focused on optimising the minimum number of observations per leaf (q) by minimising the out-of-bag error. This approach ensures that each leaf contains a sufficient number of samples for reliable predictions without overfitting. While optimising q, we kept other RF hyperparameters constant. Specifically, we set the number of predictors randomly chosen for each split (r) to one-third of the total number of predictors, a common practice that promotes diversity in tree construction. Additionally, we fixed the number of trees (p) at 100, striking a balance between model complexity and computational efficiency. This careful tuning of hyperparameters for both LASSO and RF models aims to enhance their

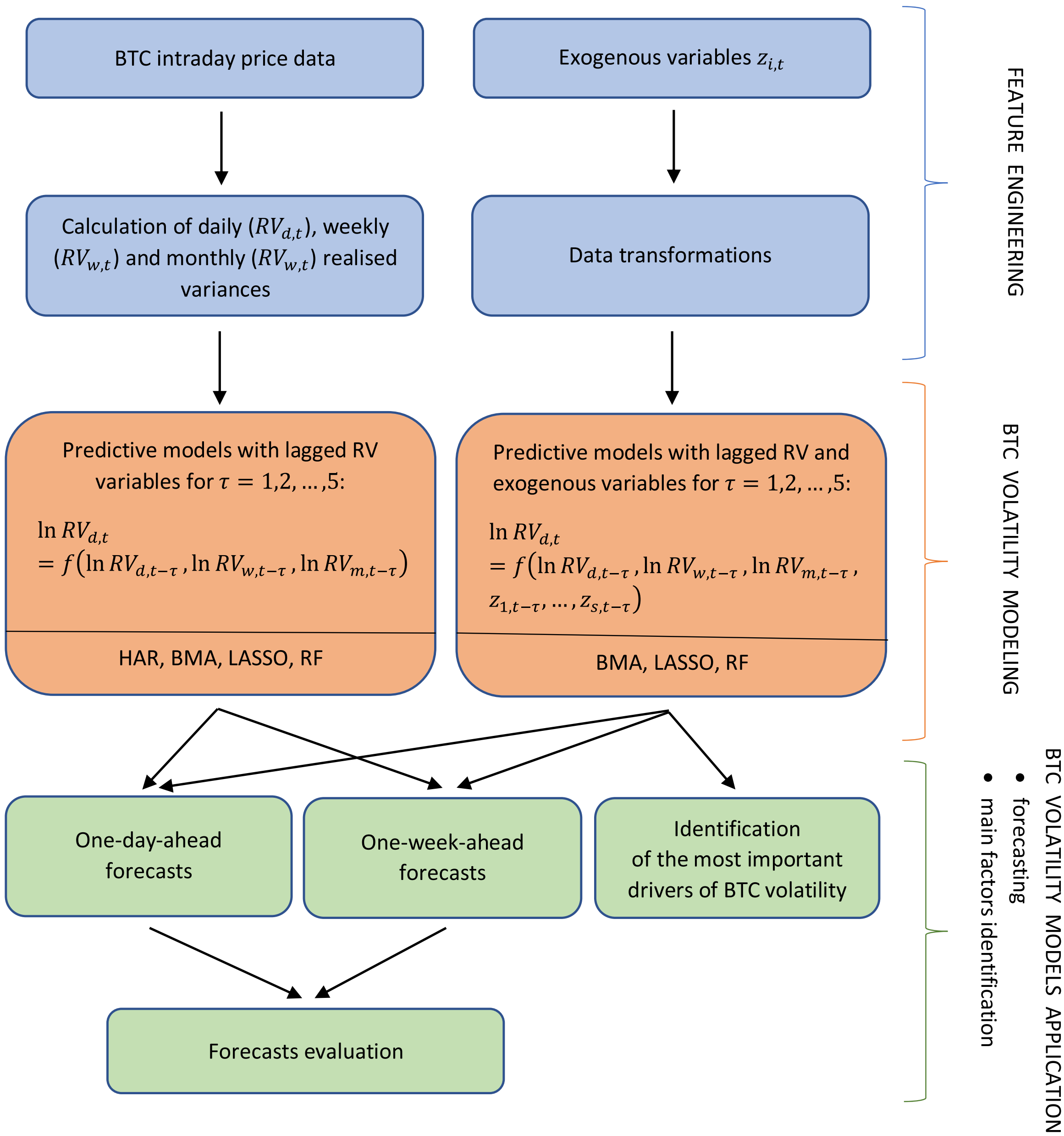

Fig. 2. The flowchart of the study.

predictive power while maintaining generalizability across different datasets. We perform this optimisation every 25 days (equivalent to five five-day weeks) on Fridays, starting from January 3, 2020.[4]

In this research, for the BMA analysis, we use the BMS package for R [113,114]. We set the g-prior to sample size (unit information prior) and the model prior (mprior) to the beta-binomial model prior (random). RF and LASSO forecasting models are implemented in Matlab 2023b using, respectively, TreeBagger and lasso functions.

To evaluate the accuracy of forecasts, we apply two fundamental measures: the mean squared error (MSE) and the mean absolute error (MAE). To further assess the predictive capabilities of our models, we implement the model confidence set (MCS) procedure, as proposed by Hansen, Lunde, and Nason [115]. This sophisticated method aims to identify the set of best-performing models from a larger pool. The MCS procedure operates iteratively, beginning with the full set of models and sequentially eliminating those found to be significantly inferior. This elimination process continues until the null hypothesis of equal forecast accuracy can no longer be rejected at a specified significance level. The resulting MCS represents the subset of models that are statistically indistinguishable from the best model, with a certain probability.

Additionally, we employ the Superior Predictive Ability (SPA) test [116] to complement our analysis. This test provides a robust method for comparing the performance of multiple forecasting models against a benchmark model, further validating our findings.

By combining these diverse evaluation techniques, we aim to provide a robust and nuanced assessment of our models' forecasting performance, ensuring that our conclusions are well-supported by both traditional metrics and advanced statistical methodologies.

4. Description of data

We analyse data obtained directly from the cryptocurrency exchange Kraken. Data sourced from coin-ranking websites may be unreliable due to the presence of non-traded prices, inaccurate timestamps, the use of

[4] For the first forecasts, i.e. for the period from January 1, 2020, to January 3, 2020, we perform the optimization on December 31, 2019.

Table 3
Evaluation of daily variance forecasts using MSE, MAE and the MCS test.

Method	MSE	Rank	P-value	MAE	Rank	P-value
HAR	0.237	9	0.091	0.998	7	0.030
BMA	0.237	8	0.091	0.998	8	0.020
BMA-X	0.380	15	0.078	1.194	16	0.001
BMA-X_out	0.233	5	0.091	1.020	11	0.007
BMA-X_st	0.273	13	0.091	1.130	15	0.000
BMA-X_out_st	0.233	6	0.091	1.019	10	0.014
LASSO	0.251	12	0.091	1.031	13	0.004
LASSO-X	0.525	16	0.078	1.063	14	0.064
LASSO-X_out	0.203	2	0.559*	0.861	2	0.888*
LASSO-X_st	0.362	14	0.078	1.013	9	0.086
LASSO-X_out_st	**0.195**	1	1.000*	**0.860**	1	1.000*
RF	0.246	11	0.091	1.027	12	0.003
RF-X	0.235	7	0.091	0.945	3	0.064
RF-X_out	0.231	3	0.091	0.948	4	0.086
RF-X_st	0.233	4	0.091	0.953	5	0.064
RF-X_out_st	0.238	10	0.091	0.962	6	0.064

Note: The values of MSE are multiplied by 10^4, the values of MAE are multiplied by 10^3, the lowest values of MSE and MAE are in bold, p-value is for the MCS test, * indicates that models belong to MCS with a confidence level of 0.90. The evaluation period is January 1, 2020 - March 31, 2023.

Table 4
Evaluation of daily variance forecasts based on the SPA test.

Compared methods	MSE P-value	MAE P-value
BMA vs. BMA-X	0.768	0.917
BMA vs. BMA-X_out	**0.019**	0.947
BMA vs. BMA-X_st	0.766	0.943
BMA vs. BMA-X_out_st	**0.017**	0.943
BMA-X_out vs. BMA-X_st	0.784	0.911
BMA-X_out vs. BMA-X_out_st	0.773	0.122
LASSO vs. LASSO-X	0.752	0.683
LASSO vs. LASSO-X_out	**0.048**	**0.000**
LASSO vs. LASSO-X_st	0.753	0.462
LASSO vs. LASSO-X_out_st	**0.073**	**0.000**
LASSO-X_out vs. LASSO-X_st	0.893	0.912
LASSO-X_out vs. LASSO-X_out_st	0.232	0.403
RF vs. RF-X	**0.073**	**0.000**
RF vs. RF-X_out	**0.072**	**0.000**
RF vs. RF-X_st	**0.082**	**0.000**
RF vs. RF-X_out_st	**0.045**	**0.000**
RF-X_out vs. RF-X_st	0.835	0.799
RF-X_out vs. RF-X_out_st	0.906	0.962

Note: The table presents p-values of the SPA test for pairs of models (displayed at the left). A p-value lower than a significance level means that the forecasts from the second model are more accurate than the forecasts from the first model, which is used as a benchmark model (the p-values lower than 0.1 are in bold). The evaluation period is January 1, 2020 - March 31, 2023.

non-fiat cross-rates, and potential wash trading (see [117]). We use 15-minute data to estimate RV of Bitcoin (BTC/USD). This data is utilised to calculate daily, weekly, and monthly realised variances for BTC (see formulas (1−3)). Based on the literature reviewed in Section 2 and our own experience, we select 33 exogenous variables, which are categorised into three main groups: (1) BTC-specific factors, (2) financial markets, and (3) market and policy uncertainty. The first group includes factors such as trading volume, the number of transactions per day, average block size per day, market capitalisation, the average hash rate per day, the number of unique addresses per day, and Google searches for Bitcoin. The financial markets group is the most extensive and includes stock indices (S&P 500, Nasdaq Composite, Euro STOXX 50, FTSE 100, Nikkei 225, Shanghai Composite), currency exchange rates (EUR/USD, JPY/USD, CNY/USD), the nominal broad U.S. dollar index, U.S. 2-year and 10-year bond yields, and commodities (NYMEX light sweet WTI crude oil, gold spot, NYMEX Henry Hub natural gas, and Bloomberg Commodity Index). The third group, market and policy uncertainty, includes the CBOE Volatility Index, EURO STOXX 50 Volatility Index, CBOE Gold ETF Volatility Index, an economic policy uncertainty index based on newspapers, a Twitter-based market uncertainty index, a Twitter-based economic uncertainty index, a geopolitical risk index based on newspapers, a risk aversion index based on financial variables, an uncertainty index based on financial variables, and an infectious disease equity market volatility tracker based on newspapers.

The data comes from various sources, which are detailed in Table 2. Since we utilise only daily data, we exclude variables with monthly or quarterly frequency, such as macroeconomic factors. Table 2 also includes a brief economic explanation of how each group of variables may affect BTC volatility. The data spans from August 1, 2017, to March 31, 2023.

The cryptocurrency market is active on weekends. However, we exclude weekend observations (i.e., Saturdays and Sundays) to ensure temporal alignment with traditional financial market data - particularly equities - which are used as explanatory variables in our models of BTC volatility.

For each exogenous factor presented in Table 2, we calculate two variables. For the variables in the first and third groups, i.e., BTC-specific factors and market and policy uncertainty, we apply[5]:

- the logarithm: $\ln x_t$,
- the first difference of logarithms: $\Delta x_t = \ln x_t - ln x_{t-1}$.

  The logarithmic transformation is commonly used in econometrics to stabilize variance, whereas the first difference of logarithms is often used to remove trends.

  For financial series, it is common to analyse returns and their volatility. Therefore, for the variables in the second group, two measures are calculated:
- the logarithmic return: $r_t = \ln(x_t/x_{t-1})$,
- the Parkinson estimator of variance[6] [118]: $\sigma^2_{Pt} = [\ln(H_t/L_t)]^2/(4\ln 2)$,

where H_t and L_t are the high and low prices of the day, respectively. The Parkinson estimator is utilised because it does not require intraday data and is more than 4.9 times as efficient as the variance estimator based on closing prices (see [118]).

We use the following prefixes to denote these transformations: ln, Δ, r, and v for the logarithm, the first difference of logarithms, the logarithmic return, and the Parkinson estimator of variance, respectively. Descriptive statistics for all analysed series are presented in Table A.1 in the Appendix. The data is available at the following address: https://doi.org/10.18150/SJHAHR.

Before the experimental studies, we tested the stationarity[7] of all exogenous variables using the Dickey-Fuller test.[8] For the following seven variables, the null hypothesis of a unit root could not be rejected: ln_volume, ln_transactions, ln_block, ln_capital, ln_addresses, ln_USDX, and ln_unc_bex. As a result, these series are excluded from the analysis, and the study is ultimately conducted on 59 stationary time series.

[5] The exception is the variable infectious disease equity market volatility tracker based on newspapers, for which due to numerous zero values, logarithms are not calculated.

[6] The exception is the variable nominal broad U.S. dollar index, for which due to the lack of daily low and high values, logarithms and logarithmic returns are calculated.

[7] The stationarity of variables is important for volatility modeling (see, e.g., [119]).

[8] These results are available from the authors upon request.

5. Empirical evaluation of forecasting models

5.1. Daily forecasts

In this section, we evaluate daily forecasts of BTC volatility. Firstly, we present results for all considered models (described in Section 5.1) jointly, based on the MCS test (see Table 3).

According to the both MSE and MAE measures, the most accurate forecast of variance are based on the LASSO-X_out_st method. However, the MCS test shows that in the set of best models, there are two kinds of LASSO models, namely LASSO-X_out_st and LASSO-X_out. Both models incorporate exogenous variables with outlier replaced, highlighting the effectiveness of this preprocessing step. When the LASSO method is used to the model with untransformed or only standardised regressors the forecasting accuracy is much lower.

The third place in the ranking takes RF-X_out for the MSE criterion and RF-X for the MAE measure. The results show that the RF model with exogenous variables predicts volatility of BTC much better than BMA.

Among models that do not incorporate exogenous variables, the HAR and BMA models demonstrate superior performance. This finding aligns with the conclusions of Dudek, Fiszeder, Kobus, and Orzeszko [15], who analysed models without exogenous variables. Their study revealed that simple linear models perform comparably to, if not better than, more complex alternatives in predicting volatility.

For robustness, we apply two additional evaluation measures: the quasi-likelihood loss function (QLIKE) and the coefficient of determination from the Mincer–Zarnowitz regression (R^2). The results of these measures are reported in Table A.2 in the Appendix. The conclusions drawn from QLIKE and R^2 are broadly consistent with those obtained using MSE and MAE. It is worth noting the relatively high values of the coefficient of determination for the LASSO-X_out_st and LASSO-X_out models, which indicate their strong forecasting performance.

To derive more precise conclusions about the performance of individual forecasting methods, we employ the SPA test. This test allows us

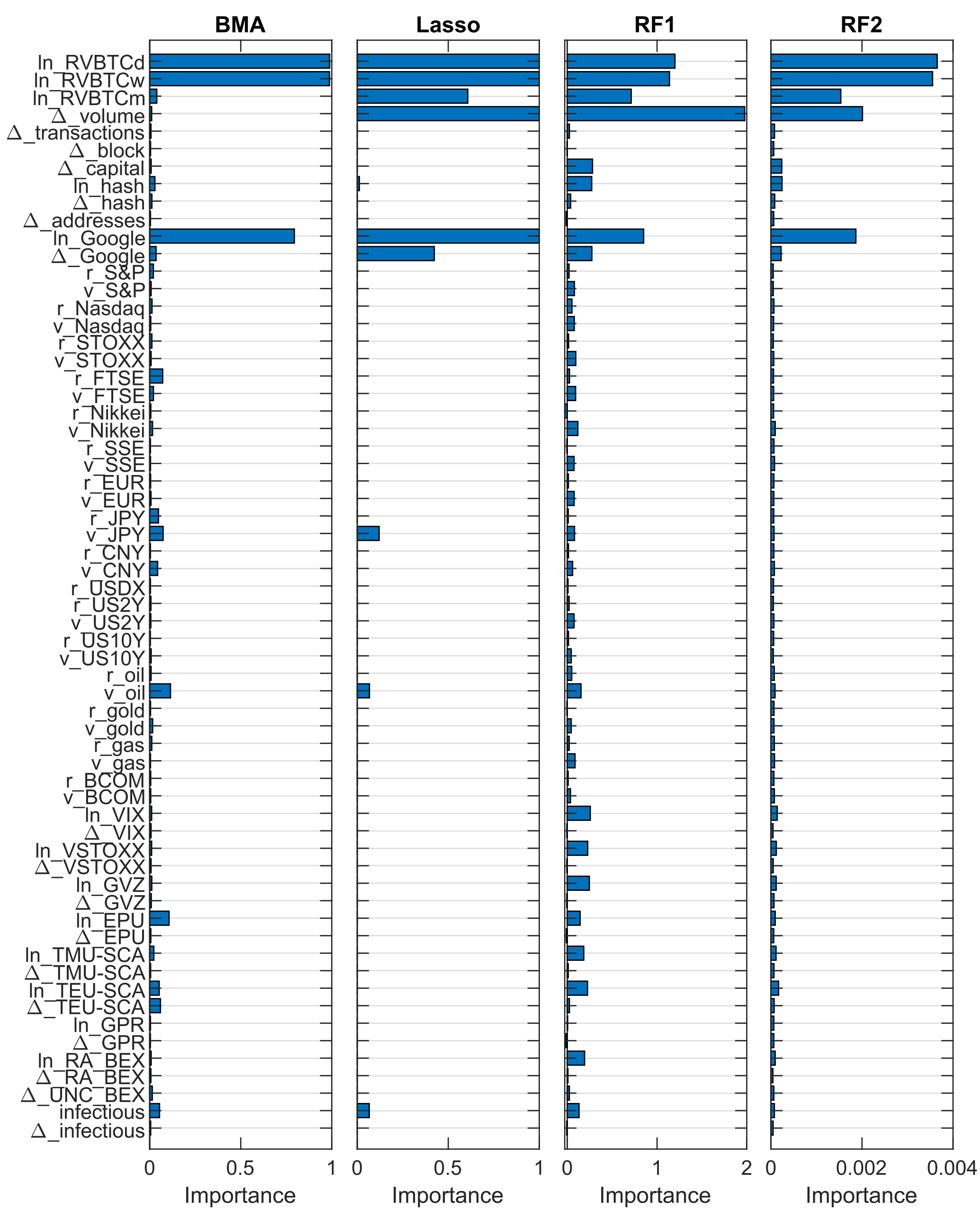

Fig. 3. Average importance of BTC volatility drivers.

to evaluate the models in pairwise comparisons, with the results presented in Table 4.

Adding untransformed regressors to the model increases the accuracy of variance forecasts only in the case of the RF model. It means that linear models like BMA and LASSO are unable to effectively use the information contained in the exogenous variables. In contrast, the RF model handles this quite well probably because it is well-suited to nonlinear relationships in the data and makes robust predictions.

Next, we check whether transformed exogenous variables influence the accuracy of predictions for each of the analysed methods separately. Firstly, we check the impact of the replacement of outliers in exogenous variables. For the BMA model such impact is ambiguous. The explanatory variables increase the accuracy of forecasts according to the MSE measure but have no influence under the MAE criterion. In contrast, for both LASSO and RF, such regressors improve forecasts of BTC variance.

The second transformation of exogenous variables, i.e. standardisation does not improve forecasts for the BMA and LASSO models but increases the forecasting accuracy for RF. Moreover, for all analysed methods, standardisation does not improve forecasts in comparison to replacing of outliers, regardless of whether it is used as the only

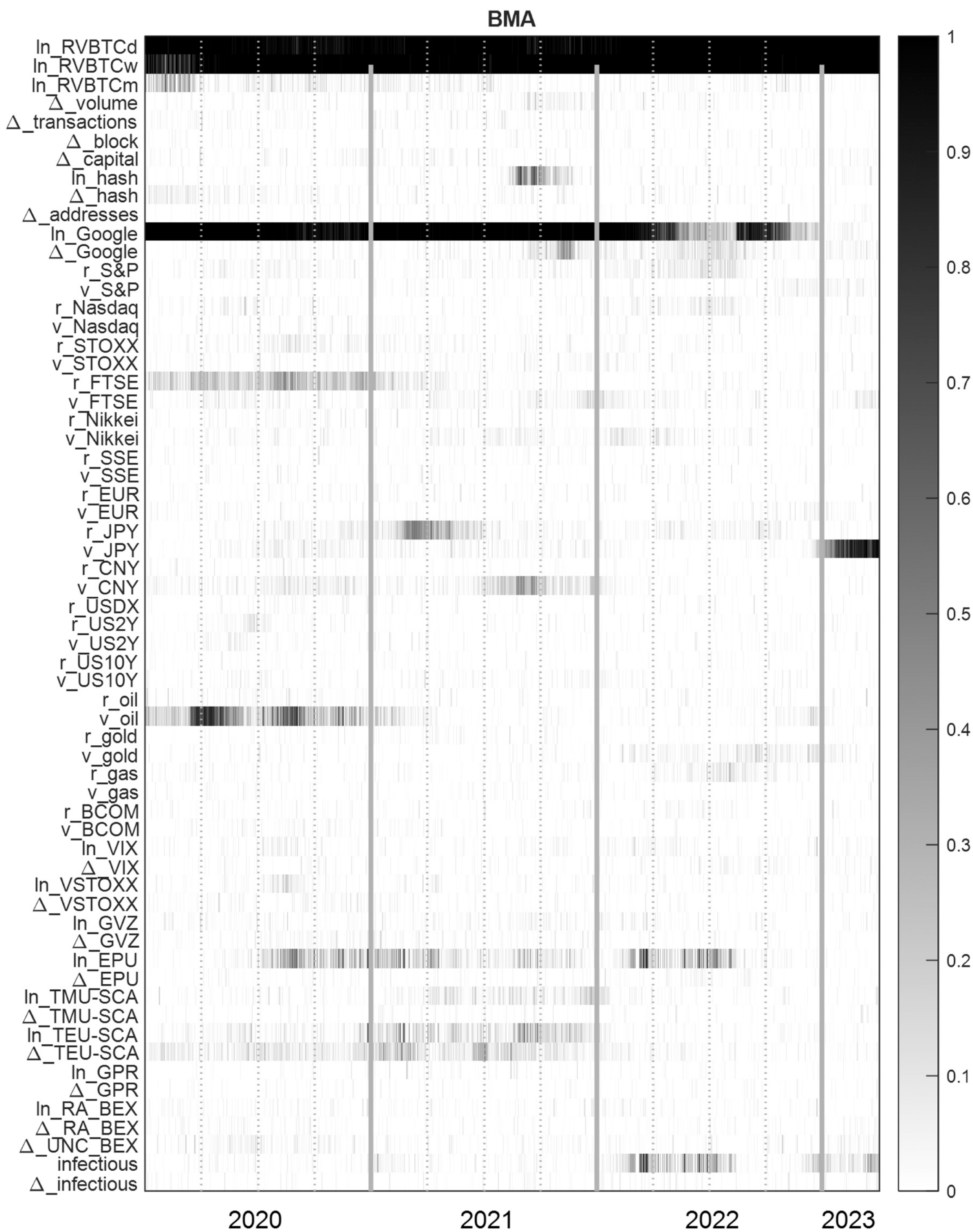

Fig. 4. Importance of BTC volatility drivers over time based on PIP in the BMA procedure.

transformation method or in combination with replacing of outliers. It means that standardisation is not necessary for our explanatory variables. However, it is important to note that our analysis is restricted to stationary variables with limited variability, achieved through the calculation of logarithms and their first differences, or alternatively, variances and returns. If someone analyses completely raw data, standardisation may be desirable to improve forecasting accuracy.

To identify the most influential factors explaining BTC variance, we employ four distinct methods: the posterior inclusion probability in BMA, the occurrence frequency of variable in LASSO models, and two RF techniques based on permutation across out-of-bag observations (RF1) and improvement in the split criterion (RF2). Our estimations utilise a rolling window, as detailed in Section 5.1. Fig. 3 presents the results of this analysis, which focuses on the BMA-X_out, LASSO-X_out, and RF-X_out models - variants incorporating exogenous variables with outliers replaced. We select this variant as it yields superior forecasting performance across all three types of models, providing a comprehensive and robust assessment of factor importance in BTC variance explanation.

The most important drivers of BTC variance are: $\ln RV_d$, $\ln RV_w$, $\ln RV_m$, Δ_volume, ln_Google and Δ_Google which denote, respectively, logarithms of daily, weekly and monthly realised variances, the first differences of logarithmised trading volume of BTC, logarithms of Google searches for Bitcoin and the first differences of logarithmised Google searches for Bitcoin. Indications of various methods are similar, however, there are some differences. The method based on PIP in the BMA model selects the smallest number of crucial factors. The $\ln RV_m$, Δ_volume and Δ_Google variables exhibit markedly less impact on Bitcoin variance when assessed through this approach, contrasting with the higher influence suggested by alternative methodologies.

Our analysis reveals a subset of explanatory factors that exhibit limited influence on Bitcoin volatility across certain methodologies. This group encompasses a diverse array of variables, including market capitalisation, the average hash rate per day, returns of the FTSE 100 index, variance of the JPY/USD currency pair, variance of WTI crude oil, the VIX Volatility Index, the EURO STOXX 50 Volatility Index, the CBOE Gold ETF Volatility Index, the economic policy uncertainty index, the Twitter-based market uncertainty index, the Twitter-based economic uncertainty index, the risk aversion index based on financial variables, the infectious disease equity market volatility tracker based on newspapers. While previous studies have attributed a significant impact to many of these factors in shaping BTC volatility (as detailed in Section 2), our comprehensive investigation suggests that their importance is considerably lower than previously thought, challenging established notions about the primary drivers of cryptocurrency market dynamics.

Based on the forecasts for successive time points, we also analyse the influence of the crucial explanatory variables on BTC variance over time. These results are given in Figs. 4–7.

The temporal analysis of explanatory variables' influence on Bitcoin variance yields method-dependent results, revealing nuanced dynamics in cryptocurrency market drivers. Notably, Bitcoin trading volume exhibited heightened impact in 2020, while the significance of Google searches for Bitcoin diminished during 2022–23 (illustrated in Figs. 4 and 6–7). Furthermore, PIP in the BMA procedure identifies additional crucial volatility drivers: WTI crude oil variance emerged as particularly influential in 2020, while the JPY/USD currency pair variance gained prominence in 2023. These temporal shifts in factor importance underscore the evolving nature of Bitcoin's relationship with broader economic indicators and market sentiment.

5.2. Weekly forecasts

We extend our analysis to weekly volatility forecasts, employing the MCS test to evaluate all models collectively. Table 5 presents these results, revealing distinct performance patterns across different evaluation criteria.

Under the MSE measure, RF models demonstrate superior performance, with RF-X_out, RF-X_st, and RF-X occupying the top three positions, respectively. In contrast, BMA models yield the weakest results. However, it is worth noting that the inclusion of explanatory variables with outliers removed significantly improved the accuracy of the BMA forecasts. Similarly, for both LASSO and RF, the addition of explanatory variables led to a substantial reduction in MSE. Unlike in the case of BMA, this improvement holds even when untransformed or merely standardised regressors are added.

Conversely, the MAE criterion favours LASSO models, with LASSO-X leading, followed by LASSO-X_out_st and LASSO-X_out. BMA consistently underperforms, ranking below even the HAR model. As with the MSE results, the inclusion of explanatory variables substantially enhances the performance of both LASSO and RF models under the MAE criterion.

The MCS test indicates a broad set of best-performing models for both MSE and MAE, likely due to the limited sample size of 169 weekly forecasts. For both error measures, BMA-X and BMA-X_st are significantly outperformed by other models. Additionally, under the MAE criterion, the HAR, BMA, and RF models are excluded from the MCS set.

To enhance robustness, we also apply two additional evaluation measures: QLIKE and R^2. The results of these measures are reported in Table A.3 in the Appendix. Overall, the outcomes based on QLIKE and R^2 are broadly consistent with those derived from MSE and MAE. Moreover, the values of the coefficient of determination are significantly lower for weekly forecasts than for daily forecasts, indicating that non-overlapping weekly data are substantially more challenging to predict.

Similarly to the daily forecasts, we also use the SPA test for pairwise model comparisons, detailed in Table 6.

The SPA test leads to very similar conclusions regardless of whether

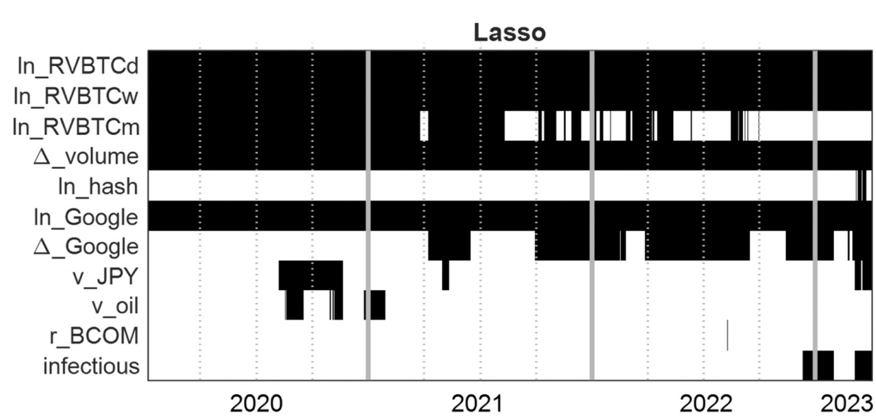

Fig. 5. Importance of BTC volatility drivers over time based on the significance of variables in LASSO.

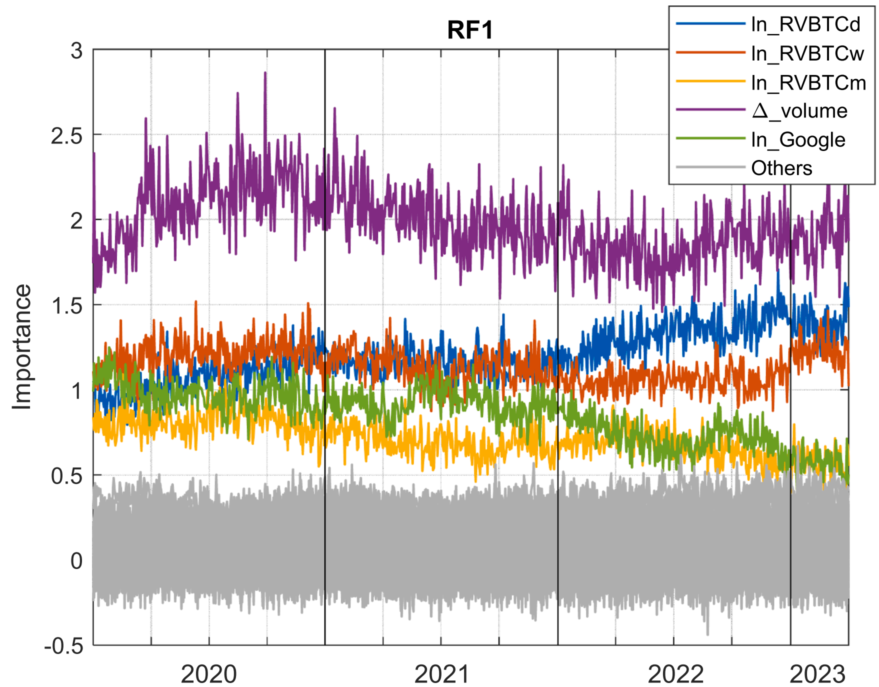

Fig. 6. Importance of BTC volatility drivers over time based on permutation across the out-of-bag observations for RF.

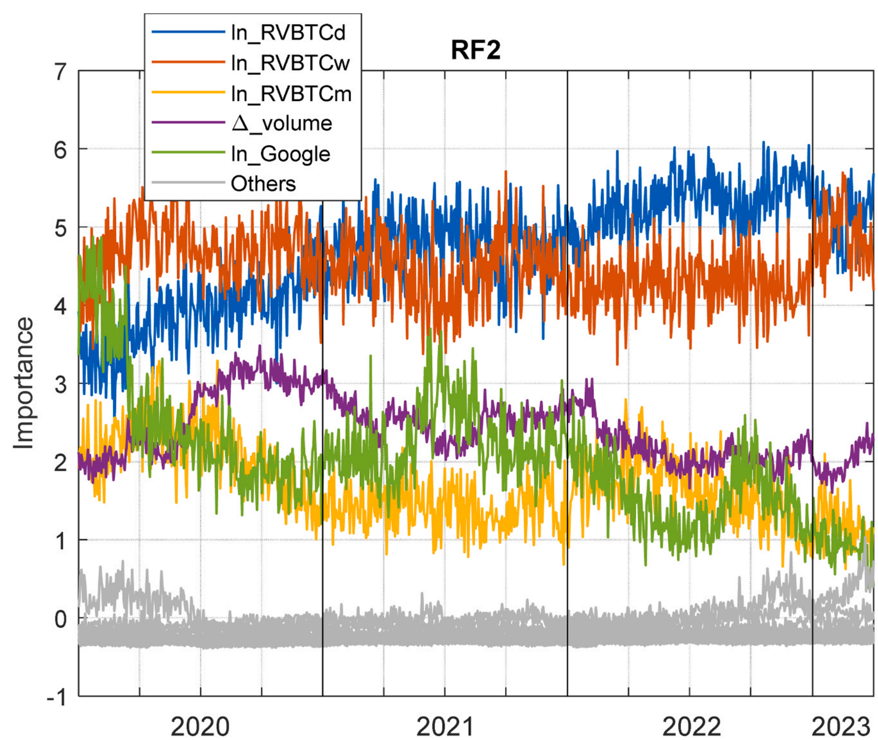

Fig. 7. Importance of BTC volatility drivers over time based on the improvement in the split criterion for RF (standardised data to enhance readability).

it is applied to the MSE or MAE criterion. Most notably, the results of this test for the BMA model differ markedly from those obtained for the LASSO and RF models. BMA shows limited improvement from exogenous variables, except when using outlier-replaced variables under the MSE criterion. In contrast, LASSO and RF models consistently benefit from both raw and transformed exogenous variables across both evaluation measures, highlighting their superior ability to leverage additional information.

Notably, for LASSO and RF models, untransformed exogenous variables significantly enhance weekly Bitcoin volatility forecasts. The SPA test suggests that further transformations, such as outlier replacement or standardisation, offer no additional forecasting accuracy improvements, rendering them unnecessary for these approaches.

Comparing the weekly and daily results, it can be observed that the RF model performs better in weekly forecasts than in daily ones. This improvement may be attributed to the model's ability to capture complex nonlinear patterns, which are more pronounced in aggregated weekly data. Weekly volatility tends to smooth out high-frequency noise

Table 5
Evaluation of weekly variance forecasts based on the MCS test.

Method	MSE	Rank	P-value	MAE	Rank	P-value
HAR	0.240	12	0.150*	0.503	13	0.025
BMA	0.240	13	0.106*	0.504	14	0.052
BMA-X	0.311	16	0.068	0.576	16	0.076
BMA-X_out	0.233	10	0.325*	0.483	10	0.425*
BMA-X_st	0.295	15	0.068	0.568	15	0.032
BMA-X_out_st	0.233	9	0.411*	0.482	9	0.553*
LASSO	0.245	14	0.068	0.498	12	0.315*
LASSO-X	0.225	4	0.883*	**0.459**	1	1.000*
LASSO-X_out	0.229	7	0.570*	0.462	3	0.834*
LASSO-X_st	0.228	6	0.591*	0.466	4	0.747*
LASSO-X_out_st	0.229	8	0.520*	0.461	2	0.834*
RF	0.236	11	0.244*	0.495	11	0.071
RF-X	0.225	3	0.883*	0.468	5	0.826*
RF-X_out	**0.224**	1	1.000*	0.471	7	0.661*
RF-X_st	0.224	2	0.971*	0.470	6	0.687*
RF-X_out_st	0.226	5	0.660*	0.472	8	0.661*

Note: The values of MSE are multiplied by 10^3, the values of MAE are multiplied by 10^2, the lowest values of MSE and MAE are in bold, p-value is for the MCS test, * indicates that models belong to MCS with a confidence level of 0.90. The evaluation period is January 1, 2020 - March 31, 2023.

Table 6
Evaluation of weekly variance forecasts based on the SPA test.

Compared methods	MSE	MAE
	P-value	P-value
BMA vs. BMA-X	0.837	0.845
BMA vs. BMA-X_out	**0.049**	0.109
BMA vs. BMA-X_st	0.824	0.849
BMA vs. BMA-X_out_st	**0.045**	**0.098**
BMA-X vs. BMA-X_out	0.154	0.143
BMA-X vs. BMA-X_st	0.152	0.186
LASSO vs. LASSO-X	**0.058**	**0.027**
LASSO vs. LASSO-X_out	**0.047**	**0.003**
LASSO vs. LASSO-X_st	**0.051**	**0.024**
LASSO vs. LASSO-X_out_st	**0.051**	**0.004**
LASSO-X vs. LASSO-X_out	0.798	0.645
LASSO-X vs. LASSO-X_st	0.901	0.547
RF vs. RF-X	**0.045**	**0.006**
RF vs. RF-X_out	**0.030**	**0.015**
RF vs. RF-X_st	**0.026**	**0.006**
RF vs. RF-X_out_st	**0.057**	**0.013**
RF-X vs RF-X_out	0.288	0.783
RF-X vs. RF-X_st	0.325	0.739

Note: The table presents p-values of the SPA test for pairs of models (displayed at the left). A p-value lower than a significance level means that the forecasts from the second model are more accurate than the forecasts from the first model, which is used as a benchmark model (the p-values lower than 0.1 are in bold). The evaluation period is January 1, 2020 - March 31, 2023.

present in daily returns, thereby emphasizing underlying structural relationships that are better suited for nonparametric models such as RF. In contrast, daily volatility is more influenced by short-term fluctuations and random shocks, which may obscure the nonlinear dependencies that RF is designed to exploit. While this issue merits further examination, such an extension is constrained by the limited availability of non-overlapping weekly observations.

The BMA model performs worse than LASSO in both daily and weekly forecasts, likely because it incorporates too many explanatory variables, including those with limited or no predictive power. This can introduce estimation noise and reduce forecast precision, especially when multicollinearity is present. LASSO, by contrast, performs variable selection and regularization simultaneously, improving generalization by retaining only the most informative predictors.

5.3. Contextualizing our results in the literature

The results we have obtained are unique, as other studies in the literature are different in methods, explanatory variables, research periods and frequency of data. In particular, it is not possible to directly confront our results with other studies, as, to the best of our knowledge, no previous work has compared BMA, LASSO, and RF in out-of-sample forecasting of BTC volatility using explanatory variables. Lehrer, Xie, and Yi [61] compared LASSO and RF models using a sentiment index derived from Twitter. For all forecast horizons considered, i.e., 1, 2, 4, and 7 days, the RF model significantly outperformed the LASSO model. Their study used data from May 20, 2015, to August 20, 2017, which precedes the period analysed in our study.

Bakas, Magkonis, and Oh [28] examined the importance of 22 potential determinants of Bitcoin volatility over the period from August 2010 to May 2020 using dynamic BMA. They identified Google Trends, total Bitcoin circulation, U.S. consumer confidence, and the S&P 500 index as the most important factors influencing Bitcoin volatility. However, they used monthly data and did not conduct an out-of-sample forecasting study.

Research on the accuracy of out-of-sample volatility forecasts that jointly considers a wide range of potential drivers within a single model remains scarce (see Section 2 for the detailed review). Zhou, Xie, Wang, Gong, and Zhu [99] demonstrated a significant impact of the 10 analysed variables on daily BTC volatility; however, their results varied across different time periods – namely, before and during the COVID-19 pandemic – as well as across the adopted time scales. Notably, the authors did not account for BTC-specific factors such as trading volume and Google search activity, which emerged as the most important predictors in our study.

Wang, Ma, Bouri, and Guo [19] utilized 17 economic variables and 3 technical strategies to forecast BTC volatility. According to the LASSO model, the most frequently selected variables in their study were the trade-weighted USD index return, S&P 500 RV, and the global real economic activity index. However they relied on monthly data and in contrast to our study, they did not consider BTC-specific factors such as trading volume, and did not apply nonlinear ML methods.

Feng, Qi, and Lucey [21] used 19 economic variables and indicators to forecast daily BTC volatility. The most important drivers of volatility were the hash rate and the effective federal funds rate. While these results differ significantly from ours, their out-of-sample forecasting period was much narrower, namely, from October 19, 2021, to December 31, 2022. Moreover, the authors did not use nonlinear ML methods.

The most comprehensive study was conducted by Wang, Andreeva, and Martin-Barragán [88], who used 27 external and 6 internal determinants to forecast daily, weekly, and monthly BTC volatility. Unfortunately, their analysis is divided into two parts: one focusing on internal variables (this set includes, for example, BTC trading volume), and another on external drivers. The adjusted closing prices of the NASDAQ and S&P 500 indices, as well as Google search activity, were identified as the most influential external determinants. However, the analysis was conducted separately for the two sets of variables rather than within a unified model.

5.4. Discussion

1. The main goal of this work is to identify the primary drivers of Bitcoin volatility using various predictive models. It must be emphasised that this task is very challenging due to the nature of the time series data for both Bitcoin volatility and the exogenous variables, which lack distinctive patterns, strong seasonality, and exhibit chaotic trends and large random fluctuations. Key factors contributing to the high volatility of cryptocurrencies include the lack of fundamental value, making their prices highly dependent on volatile market demand, and the strong influence of market sentiment and speculation driven by news events and social media. Additionally, the absence of regulatory oversight, market immaturity, low capitalisation, and limited historical data further complicate accurate

price forecasting and increase susceptibility to significant price fluctuations.

2. It should be noted that the results, specifically the importance of considered exogenous variables as BTC volatility drivers, depend on the predictive model used. Different models combine variables in various ways, leading to discrepancies in the perceived importance of each variable. For instance, RF constructs more complex models than linear models such as LASSO and BMA. RF can capture intricate relationships between variables, which linear models might overlook due to their simpler structure. Consequently, linear models can only identify drivers that have a linear impact on BTC volatility, whereas more complex models can reveal nonlinear influences.
3. Note that the predictive models we used have fundamentally different mechanisms for estimating variable importance, and thus, the importance determined by each model carries a different meaning. Here is a clarification of these differences:
   - BMA accounts for model uncertainty in the variable selection process by averaging over the best models within the model class according to their approximate posterior model probability. Variables with higher model-averaged estimates are considered more important, as they consistently contribute to the best-performing models.
   - LASSO minimises a regularised loss function, which is particularly useful for feature importance estimation because it can shrink some coefficients in a linear model to zero, effectively performing variable selection. The non-zero coefficients provide an indication of the importance of the corresponding variables. This method highlights variables that have a significant linear relationship with the target variable.
   - RF employs two methods for calculating variable importance:
     a. Permutation Importance: This method utilises out-of-bag observations. It involves shuffling the values of a variable in the out-of-bag set and measuring how much the permutation decreases the model's accuracy. The variable importance is calculated as the difference between the baseline accuracy and the accuracy after shuffling, averaged over all trees. This method reflects how much the model relies on each variable for making accurate predictions.
     b. Improvement in the split criterion: Each time a variable is used to split a node, the reduction in the split criterion (MSE) is calculated for the child nodes. The variable importance is computed as the total reduction of the criterion brought by that variable, averaged over all trees in the forest. This method indicates how well each variable contributes to making splits that improve the purity of the nodes.

     In summary, while BMA focuses on averaging model contributions to account for uncertainty, LASSO highlights variables with significant linear relationships, and RF emphasises the reduction in prediction error and split criterion. Understanding these different mechanisms helps in interpreting the importance of variables in a more nuanced way, depending on the model used.
4. The influence of the factors affecting Bitcoin volatility was time-varying. The impact of BTC trading volume and Google searches for Bitcoin was much higher during the COVID-19 pandemic. The obtained results extend the previous findings in the literature. Guzmán, Pinto-Gutiérrez, and Trujillo [120] found that during the coronavirus pandemic investors traded more BTC on days with low mobility connected with lockdown mandates. Moreover, Corbet, Hou, Hu, Larkin, Lucey, and Oxley [121] analysed the connections between cryptocurrency volatility and liquidity (measured by trading volume) during the outbreak of the pandemic. Results suggest that liquidity increased sharply, in accordance with the WHO announcements of a worldwide pandemic.

   According to the results of our study, the significance of Google searches was stronger during the COVID-19 pandemic and diminished during 2022–23. This conclusion expands research results from Salisu, and Ogbonna [77], who analysed the influence of information from Google Trends on the volatility of Bitcoin, Ethereum, Litecoin and Ripple. They showed a positive impact of news on the volatility of cryptocurrencies, which was higher during the coronavirus pandemic than the period before it.

   According to PIP in the BMA procedure the WTI crude oil variance influenced significantly BTC volatility in 2020. Our results yet again extend the findings of previous results in the literature. Maghyereh, and Abdoh [67] investigated volatility interlinkages between Bitcoin and various financial assets. They found that the volatility coherence between Bitcoin and crude oil was limited and occurred only during the COVID-19 pandemic - primarily in the early period from January to May 2020, and later in November 2020. A significant increase in posterior inclusion probabilities for the crude oil index was also observed in 2020 by Bakas, Magkonis, and Oh [28]. Furthermore, according to the study by Zhou, Xie, Wang, Gong, and Zhu [99], the short-term influence of oil volatility on Bitcoin volatility was positive before the COVID-19 pandemic but turned negative during the pandemic. Moreover, Elsayed, Gozgor, and Lau [48] observed changes in the direction of return and volatility spillovers between oil and BTC during the pandemic; however, these changes were very short-term.
5. As traditional financial markets operate from Monday to Friday, we excluded all data from Saturdays and Sundays from our analysis. To account for weekend trading effects in future studies, several approaches could be considered:
   - separate modelling of weekend volatility, e.g., estimating volatility on weekends separately or including weekend-specific variance components;
   - using dummy variables to capture potential weekend spillover effects;
   - employing mixed-frequency models (such as GARCH-MIDAS) that can accommodate differing trading calendars or allow for continuous crypto data to inform models of lower-frequency (weekday-only) data.

   These adjustments could improve the understanding of inter-market volatility transmission across different trading calendars.

6. Conclusions

The cryptocurrency market is constantly developing, and its importance is growing every year. This is confirmed by the increasing number of crypto assets, growing capitalisation and the increasing number of new derivatives and other instruments.

Despite the growing development of the cryptocurrency market, the volatility of these instruments is still enormous compared to other financial assets. This high variability is a serious problem because it plays an important role in many financial strategies. It is crucial, for example, for the construction of portfolios, hedging, valuation of derivatives, and risk management.

Forecasting the huge volatility of cryptocurrencies is very difficult, and the application of standard volatility models such as the GARCH model does not give satisfactory results (see [15]). That is why we apply in this study more advanced forecasting methods: HAR, BMA, LASSO and RF, and use the widest in the literature set of explanatory variables which potentially can influence Bitcoin volatility. Three of the methods used, namely, BMA, LASSO and RF have built-in mechanisms for variable selection, allowing us to evaluate variable importance. We show that LASSO and RF models with exogenous variables significantly increase the accuracy of both daily and weekly BTC variance forecasts in comparison to the models which include only lagged volatilities of BTC. Of the wide set of variables taken for the study, the ones with the greatest impact on BTC volatility are lagged daily, weekly and monthly realised variances of BTC, trading volume of BTC and Google searches for Bitcoin. Moreover, this influence of the variables on BTC volatility is time-varying and underscores the evolving nature of Bitcoin's relationship with broader economic indicators and market sentiment.

For daily forecasts, linear models BMA and LASSO are unable to use the information contained in the untransformed exogenous variables effectively. In contrast, the RF model handles this quite well probably because it is well-suited to nonlinear relationships in the data and makes robust predictions. The most accurate daily forecasts of Bitcoin variance are based on the LASSO model, provided that outliers have been replaced in exogenous variables.

In the case of weekly forecasts, the results depend on the adopted error measure. According to the MSE criterion, the best forecasts are from the RF model, while the MAE measure indicates the LASSO model. For RF and LASSO, incorporating untransformed exogenous variables significantly enhances the accuracy of weekly forecasts compared to models without exogenous inputs. However, further processing through outlier replacement and standardisation does not lead to additional improvements in accuracy.

The study's results offer valuable insights for investors and other participants in the crypto-asset markets. The identified relationships can aid investment decision-making, particularly in evaluating volatility and managing investment risk. The findings suggest that several persistent factors with predictive capabilities can be identified, along with other factors that may show such characteristics during specific periods. The findings further suggest that, beyond price movements, investors and other market participants should continuously monitor additional cryptocurrency market characteristics, such as volatility and trading volume, as these exhibit strong predictive properties concerning future market volatility. In addition, market sentiment, as proxied by the Google Trends index, remains an important subject of ongoing analysis by market participants. The proposed methods allow for including these factors and their adjustment over time. These results may also benefit supervisory authorities, market makers, and crypto-asset market organisers in monitoring market conditions and preparing for increased or decreased volatility.

Our research can be further developed and expanded in several directions in the future. First, alternative forecasting methods and different variable selection techniques could be explored. Our study focused solely on forecasting models that incorporate an internal mechanism for predictor evaluation. However, other variable selection methods exist and could be applied to a broader class of forecasting models (e.g., [122]). Second, it would be valuable to consider other cryptocurrencies to determine whether our findings for BTC volatility can be generalised to other digital assets. Third, our study highlighted the significant impact of data preprocessing. Therefore, it seems crucial to also consider alternative preprocessing methods beyond outlier replacement and standardisation, which could further enhance forecast accuracy.

CRediT authorship contribution statement

Witold Orzeszko: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Piotr Fiszeder:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Grzegorz Dudek:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Radosław Pietrzyk:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used AI-based language tools (ChatGPT, Claude, and Gemini) in order to improve language and readability. After using this tools/services, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Piotr Fiszeder supported by the National Science Centre project number 2021/43/B/HS4/00353 entitled "Robust methods for range-based models - Risk and comovement analysis on the cryptocurrency market" and by Institutional support of the Prague University of Economics and Business no. IP100040. Witold Orzeszko supported by the National Science Centre project number 2019/35/B/HS4/00642 entitled "Forecasting financial markets using a machine learning approach".

Acknowledgements

Piotr Fiszeder and Radosław Pietrzyk supported by the National Science Centre project number 2021/43/B/HS4/00353 entitled "Robust methods for range-based models - Risk and comovement analysis on the cryptocurrency market"; Piotr Fiszeder funded by Institutional support of the Prague University of Economics and Business no. IP100040; Witold Orzeszko supported by the National Science Centre project number 2019/35/B/HS4/00642 entitled "Forecasting financial markets using a machine learning approach".

Appendix

Table A.1
Summary statistics of analysed data

Variable	Mean	Minimum	Maximum	Standard deviation	Skewness	Excess kurtosis
ln_RVBTCd	-6.9006	-10.7399	-2.0680	1.1302	0.11	0.61
ln_RVBTCw	-6.7185	-9.6766	-3.2473	0.9867	0.11	0.78
ln_RVBTCm	-6.5401	-8.5368	-4.4751	0.8165	0.29	0.11
BTC-specific factors						
ln_volume	23.6402	20.5061	26.5840	0.9535	-0.78	-0.07
Δ_volume	0.1802	-1.8737	1.8624	0.2283	0.25	7.38
ln_transactions	12.5611	11.7896	13.1035	0.1751	-0.42	0.08
Δ_transactions	0.0234	-0.5606	0.4753	0.0923	-0.06	4.15
ln_block	0.1343	-0.5681	0.9135	0.1885	-0.10	2.14
Δ_block	0.0538	-0.5584	0.4529	0.1034	-0.02	2.40

(continued on next page)

Table A.1 (*continued*)

Variable	Mean	Minimum	Maximum	Standard deviation	Skewness	Excess kurtosis
ln_capital	26.2492	24.5166	27.8760	0.8613	0.21	-1.13
Δ_capital	0.1651	-0.2929	0.1824	0.0436	-0.29	3.84
ln_hash	18.2476	15.2227	19.8023	0.9378	-0.94	0.31
Δ_hash	0.2750	-0.5486	0.5803	0.1228	-0.02	0.73
ln_addresses	13.3246	12.6153	13.8858	0.1839	-0.28	-0.20
Δ_addresses	0.0273	-0.4786	0.4051	0.0824	-0.20	3.02
ln_Google	2.6803	1.5539	4.6052	0.5249	0.52	-0.21
Δ_Google	0.0097	-1.0537	1.4697	0.1941	1.16	9.34
Financial markets						
r_S&P	0.0335	-0.1277	0.0897	0.0130	-0.83	14.53
v_S&P	0.0084	0.0001	0.2577	0.0180	7.84	83.88
r_Nasdaq	0.0432	-0.1315	0.0893	0.0153	-0.63	7.61
v_Nasdaq	0.0119	0.0001	0.3341	0.0212	6.78	68.87
r_STOXX	0.0147	-0.1324	0.0883	0.0125	-0.99	14.21
v_STOXX	0.0096	0.0000	0.4799	0.0242	11.96	195.47
r_FTSE	0.0022	-0.1151	0.0867	0.0107	-1.15	15.93
v_FTSE	0.0076	0.0001	0.3505	0.0185	10.85	160.42
r_Nikkei	0.0225	-0.0627	0.0773	0.0117	-0.13	4.15
v_Nikkei	0.0057	0.0000	0.2984	0.0133	12.74	228.19
r_SSE	-0.0002	-0.0804	0.0555	0.0106	-0.66	5.69
v_SSE	0.0070	0.0003	0.1364	0.0087	5.34	52.94
r_EUR	-0.0056	-0.0206	0.0212	0.0046	-0.03	1.42
v_EUR	0.0023	0.0000	0.0292	0.0026	4.21	27.73
r_JPY	0.0125	-0.0386	0.0316	0.0051	-0.42	8.10
v_JPY	0.0027	0.0000	0.0750	0.0052	7.72	79.87
r_CNY	0.0014	-0.0162	0.0158	0.0029	-0.13	3.64
v_CNY	0.0005	0.0000	0.0088	0.0007	4.48	31.31
ln_USDX	4.7497	4.6681	4.8545	0.0376	0.41	0.01
r_USDX	0.0057	-0.0191	0.0187	0.0031	0.23	3.00
r_US2Y	0.0753	-0.3452	0.3535	0.0466	-0.19	12.26
v_US2Y	0.2628	0.0000	12.0719	0.7982	8.00	84.31
r_US10Y	0.0294	-0.3241	0.3678	0.0362	0.06	24.91
v_US10Y	0.1295	0.0000	20.0079	0.6547	21.36	592.96
r_oil	0.0266	-0.3454	0.3196	0.0337	-0.96	29.97
v_oil	0.1024	0.0025	10.2820	0.4147	15.07	298.70
r_gold	0.0301	-0.0590	0.0430	0.0087	-0.40	3.78
v_gold	0.0082	0.0000	0.3033	0.0139	10.07	161.42
r_gas	-0.0192	-0.3005	0.3817	0.0409	0.12	9.08
v_gas	0.1296	0.0026	9.6801	0.2884	25.03	815.68
r_BCOM	0.0199	-0.0526	0.0406	0.0097	-0.62	3.52
v_BCOM	0.0065	0.0003	0.2500	0.0100	11.66	248.33
Market and policy uncertainty						
ln_VIX	2.9501	2.2127	4.4151	0.3625	0.36	0.38
Δ_VIX	0.0418	-0.2662	0.7682	0.0817	1.55	8.59
ln_VSTOXX	2.9729	2.3684	4.4499	0.3425	0.71	0.69
Δ_VSTOXX	0.0239	-0.3452	0.4857	0.0769	1.02	4.12
ln_GVZ	2.7203	2.1838	3.8914	0.2801	0.45	0.18
Δ_GVZ	-0.0001	-0.2657	0.2977	0.0518	0.71	3.51
ln_EPU	4.7738	2.3906	6.6941	0.6055	0.28	0.34
Δ_EPU	0.0015	-1.7103	1.9415	0.4669	0.12	0.72
ln_TMU-SCA	4.0229	1.2415	6.3766	0.5599	0.38	1.70
Δ_TMU-SCA	-0.1053	-2.3297	2.2601	0.4266	0.42	2.95
ln_TEU-SCA	4.0605	1.2687	5.9715	0.5757	0.15	0.58
Δ_TEU-SCA	0.0156	-1.8670	2.3983	0.4033	0.15	2.99
ln_GPR	4.6338	2.2504	6.2965	0.4687	-0.12	1.03
Δ_GPR	-0.0228	-2.9959	2.3449	0.4767	-0.16	1.74
ln_RA_BEX	1.0962	0.8860	3.2718	0.2047	4.28	31.47
Δ_RA_BEX	0.0110	-1.1791	1.2980	0.0774	3.20	127.58
ln_UNC_BEX	-10.2622	-11.3851	-8.6406	0.4254	0.44	1.57
Δ_UNC_BEX	0.0316	-0.2775	0.4355	0.0479	1.63	15.09
infectious	8.1746	0.0000	68.3700	10.2598	1.85	4.57
Δ_infectious	0.6691	-42.9800	40.9300	6.7918	0.05	7.92

Note: The mean values for variables Δ_, r_, v_ and minimum, maximum, standard deviation for variables v_ are multiplied by 10^2. The analysed period is August 1, 2017 - March 31, 2023.

Table A.2
Evaluation of daily variance forecasts using QLIKE and R^2

Method	QLIKE	Rank	P-value	R^2	Rank
HAR	-5.562	14	0.000	0.153	11
BMA	-5.562	13	0.000	0.153	10
BMA-X	-5.599	9	0.000	0.057	14
BMA-X_out	-5.591	12	0.001	0.169	8
BMA-X_st	-5.597	10	0.000	0.094	13
BMA-X_out_st	-5.591	11	0.000	0.169	7
LASSO	-5.532	15	0.001	0.161	9
LASSO-X	-5.731	3	0.022	0.030	16
LASSO-X_out	-5.754	2	0.599*	0.435	2
LASSO-X_st	-5.726	4	0.010	0.046	15
LASSO-X_out_st	**-5.757**	1	1.000*	**0.510**	1
RF	-5.530	16	0.001	0.131	12
RF-X	-5.695	7	0.017	0.219	5
RF-X_out	-5.702	5	0.027	0.233	3
RF-X_st	-5.696	6	0.020	0.228	4
RF-X_out_st	-5.686	8	0.022	0.195	6

Note: The lowest values of QLIKE and the highest value of R^2 are in bold, p-value is for the MCS test, * indicates that models belong to MCS with a confidence level of 0.90. The evaluation period is January 1, 2020 - March 31, 2023.

Table A.3
Evaluation of weekly variance forecasts using QLIKE and R^2

Method	QLIKE	Rank	P-value	R^2	Rank
HAR	-3.444	13	0.001	0.082	14
BMA	-3.441	14	0.001	0.081	15
BMA-X	-3.534	6	0.631*	0.090	10
BMA-X_out	-3.340	16	0.408*	0.088	12
BMA-X_st	-3.488	12	0.225*	0.092	9
BMA-X_out_st	-3.340	15	0.394*	0.089	11
LASSO	-3.495	11	0.686*	0.067	16
LASSO-X	-3.499	10	0.729*	0.129	3
LASSO-X_out	-3.507	8	0.753*	**0.141**	1
LASSO-X_st	-3.538	5	0.904*	0.120	7
LASSO-X_out_st	-3.501	9	0.731*	0.140	2
RF	-3.524	7	0.868*	0.083	13
RF-X	-3.705	4	1.000*	0.123	6
RF-X_out	-3.727	3	1.000*	0.123	5
RF-X_st	**-3.743**	1	1.000*	0.125	4
RF-X_out_st	-3.733	2	1.000*	0.119	8

Note: The lowest values of QLIKE and the highest value of R^2 are in bold, p-value is for the MCS test, * indicates that models belong to MCS with a confidence level of 0.90. The evaluation period is January 1, 2020 - March 31, 2023.

Data availability

The data and code generated in this study have been made publicly available and can be accessed via the links provided in the article.

References

[1] L. Ryll, S. Seidens, 'Evaluating the performance of machine learning algorithms in financial market forecasting: a comprehensive survey', arXiv (2019) https://doi.org/10.48550/ARXIV.1906.07786.
[2] P. Fiszeder, W. Orzeszko, Covariance matrix forecasting using support vector regression, Appl. Intell. 51 (10) (Oct. 2021) 7029–7042, https://doi.org/10.1007/s10489-021-02217-5.
[3] A.M. Ozbayoglu, M.U. Gudelek, O.B. Sezer, Deep learning for financial applications: a survey, Appl. Soft Comput. 93 (Aug. 2020) 106384, https://doi.org/10.1016/j.asoc.2020.106384.
[4] O.B. Sezer, M.U. Gudelek, A.M. Ozbayoglu, Financial time series forecasting with deep learning: a systematic literature review: 2005–2019, Appl. Soft Comput. 90 (May 2020) 106181, https://doi.org/10.1016/j.asoc.2020.106181.
[5] I. Gurrib, Machine learning and portfolio management: a review', Ann. Math. Comput. Sci. 5 (Mar. 2022) 31–43.
[6] T.O. Kehinde, F.T.S. Chan, S.H. Chung, Scientometric review and analysis of recent approaches to stock market forecasting: two decades survey, Expert Syst. Appl. 213 (Mar. 2023) 119299, https://doi.org/10.1016/j.eswa.2022.119299.
[7] N. Nazareth, Y.V. Ramana Reddy, Financial applications of machine learning: a literature review, Expert Syst. Appl. 219 (Jun. 2023) 119640, https://doi.org/10.1016/j.eswa.2023.119640.
[8] J.V.R. Ferro, R.J.R.D. Santos, E. De Barros Costa, J.R. Da Silva Brito, Machine learning techniques via ensemble approaches in stock exchange index prediction: systematic review and bibliometric analysis, Appl. Soft Comput. 167 (Dec. 2024) 112359, https://doi.org/10.1016/j.asoc.2024.112359.
[9] Y. Chen, L. Zhang, Z. Xie, W. Zhang, Q. Li, Unraveling asset pricing with AI: A systematic literature review, Appl. Soft Comput. 175 (May 2025) 112978, https://doi.org/10.1016/j.asoc.2025.112978.
[10] A. Bâra, S.-V. Oprea, M. Panait, Insights into Bitcoin and energy nexus. A Bitcoin price prediction in bull and bear markets using a complex meta model and SQL analytical functions, Appl. Intell. 54 (8) (Apr. 2024) 5996–6024, https://doi.org/10.1007/s10489-024-05474-2.
[11] A. Bâra, S.-V. Oprea, An ensemble learning method for Bitcoin price prediction based on volatility indicators and trend, Eng. Appl. Artif. Intell. 133 (Jul. 2024) 107991, https://doi.org/10.1016/j.engappai.2024.107991.
[12] S.A. Basher, P. Sadorsky, Forecasting Bitcoin price direction with random forests: how important are interest rates, inflation, and market volatility? Mach. Learn. Appl. 9 (Sep. 2022) 100355 https://doi.org/10.1016/j.mlwa.2022.100355.
[13] S.B. Akbar, K. Thanupillai, V. Govindarajan, Forecasting Bitcoin price using time opinion mining and bi-directional GRU, J. Intell. Fuzzy Syst. 42 (3) (Feb. 2022) 1825–1833, https://doi.org/10.3233/JIFS-211217.
[14] F. Corsi, A simple approximate long-memory model of realized volatility, J. Financ. Econ. 7 (2) (Nov. 2008) 174–196, https://doi.org/10.1093/jjfinec/nbp001.
[15] G. Dudek, P. Fiszeder, P. Kobus, W. Orzeszko, Forecasting cryptocurrencies volatility using statistical and machine learning methods: a comparative study, Appl. Soft Comput. 151 (Jan. 2024) 111132, https://doi.org/10.1016/j.asoc.2023.111132.
[16] S. Peng, C. Prentice, S. Shams, T. Sarker, A systematic literature review on the determinants of cryptocurrency pricing, China Account. Financ. Rev. 26 (1) (Mar. 2024) 1–30, https://doi.org/10.1108/CAFR-05-2023-0053.

[17] E.S. Gunnarsson, H.R. Isern, A. Kaloudis, M. Risstad, B. Vigdel, S. Westgaard, Prediction of realized volatility and implied volatility indices using AI and machine learning: a review, Int. Rev. Financ. Anal. 93 (May 2024) 103221, https://doi.org/10.1016/j.irfa.2024.103221.
[18] N.A. Kyriazis, A Survey on volatility fluctuations in the decentralized cryptocurrency financial assets, J. Risk Financ. Manag 14 (7) (Jun. 2021) 293, https://doi.org/10.3390/jrfm14070293.
[19] J. Wang, F. Ma, E. Bouri, Y. Guo, Which factors drive Bitcoin volatility: macroeconomic, technical, or both? J. Forecast 42 (4) (2023) 970–988, https://doi.org/10.1002/for.2930.
[20] A. Benhamed, A.S. Messai, G. El Montasser, On the Determinants of Bitcoin returns and volatility: what we get from gets? Sustainability 15 (3) (Jan. 2023) 1761, https://doi.org/10.3390/su15031761.
[21] L. Feng, J. Qi, B. Lucey, Enhancing cryptocurrency market volatility forecasting with daily dynamic tuning strategy, Int. Rev. Financ. Anal. 94 (Jul. 2024) 103239, https://doi.org/10.1016/j.irfa.2024.103239.
[22] H.A. Aalborg, P. Molnár, J.E. De Vries, What can explain the price, volatility and trading volume of Bitcoin? Financ. Res. Lett. 29 (Jun. 2019) 255–265, https://doi.org/10.1016/j.frl.2018.08.010.
[23] M. Al Guindy, Cryptocurrency price volatility and investor attention, Int. Rev. Econ. Financ. 76 (Nov. 2021) 556–570, https://doi.org/10.1016/j.iref.2021.06.007.
[24] M.S. Alam, A. Amendola, V. Candila, S.D. Jabarabadi, Is monetary policy a driver of cryptocurrencies? Evidence from a structural break GARCH-MIDAS approach, Econometrics 12 (1) (Jan. 2024) 2, https://doi.org/10.3390/econometrics12010002.
[25] N. Aslanidis, A.F. Bariviera, Ó.G. López, The link between cryptocurrencies and Google Trends attention, Financ. Res. Lett. 47 (Jun. 2022) 102654, https://doi.org/10.1016/j.frl.2021.102654.
[26] A.F. Aysan, E. Demir, G. Gozgor, C.K.M. Lau, Effects of the geopolitical risks on Bitcoin returns and volatility, Res. Int. Bus. Financ. 47 (Jan. 2019) 511–518, https://doi.org/10.1016/j.ribaf.2018.09.011.
[27] V. Babalos, E. Bouri, R. Gupta, Does the introduction of US spot Bitcoin ETFs affect spot returns and volatility of major cryptocurrencies? Q. Rev. Econ. Financ. 102 (Jun. 2025) 102006 https://doi.org/10.1016/j.qref.2025.102006.
[28] D. Bakas, G. Magkonis, E.Y. Oh, What drives volatility in Bitcoin market? Financ. Res. Lett. 50 (Dec. 2022) 103237 https://doi.org/10.1016/j.frl.2022.103237.
[29] M. Balcilar, E. Bouri, R. Gupta, D. Roubaud, Can volume predict Bitcoin returns and volatility? A quantiles-based approach, Econ. Model 64 (Aug. 2017) 74–81, https://doi.org/10.1016/j.econmod.2017.03.019.
[30] B. Będowska-Sójka, J. Górka, D. Hemmings, A. Zaremba, Uncertainty and cryptocurrency returns: A lesson from turbulent times, Int. Rev. Financ. Anal. 94 (Jul. 2024) 103330, https://doi.org/10.1016/j.irfa.2024.103330.
[31] B.M. Blau, Price dynamics and speculative trading in bitcoin, Res. Int. Bus. Financ. 41 (Oct. 2017) 493–499, https://doi.org/10.1016/j.ribaf.2017.05.010.
[32] J. Bleher, T. Dimpfl, Today I got a million, tomorrow, I don't know: On the predictability of cryptocurrencies by means of Google search volume, Int. Rev. Financ. Anal. 63 (May 2019) 147–159, https://doi.org/10.1016/j.irfa.2019.03.003.
[33] D. Bourghelle, F. Jawadi, P. Rozin, Do collective emotions drive bitcoin volatility? A triple regime-switching vector approach, J. Econ. Behav. Organ 196 (Apr. 2022) 294–306, https://doi.org/10.1016/j.jebo.2022.01.026.
[34] D. Bourghelle, F. Jawadi, P. Rozin, Can Collective Emotions Improve Bitcoin Volatility Forecasts? Bank. Mark. Invest N° 171 (4) (Dec. 2022) 10–19, https://doi.org/10.54695/bmi.171.8459.
[35] E. Bouri, K. Gkillas, R. Gupta, C. Pierdzioch, Forecasting Realized Volatility of Bitcoin: The Role of the Trade War, Comput. Econ. 57 (1) (Jan. 2021) 29–53, https://doi.org/10.1007/s10614-020-10022-4.
[36] E. Bouri, L. Kristoufek, N. Azoury, Bitcoin and S&P500: Co-movements of high-order moments in the time-frequency domain', PLOS ONE 17 (11) (Nov. 2022) e0277924 https://doi.org/10.1371/journal.pone.0277924.
[37] E. Bouri, C.K.M. Lau, B. Lucey, D. Roubaud, Trading volume and the predictability of return and volatility in the cryptocurrency market, Financ. Res. Lett. 29 (Jun. 2019) 340–346, https://doi.org/10.1016/j.frl.2018.08.015.
[38] A. Brauneis, M. Sahiner, Crypto volatility forecasting: mounting a HAR, sentiment, and machine learning horserace, ' AsiaPac. Financ. Mark. (Dec. 2024), https://doi.org/10.1007/s10690-024-09510-6.
[39] V. Candila, Multivariate Analysis of Cryptocurrencies, Econometrics 9 (3) (Jul. 2021) 28, https://doi.org/10.3390/econometrics9030028.
[40] T. Conlon, S. Corbet, R.J. McGee, The Bitcoin volume-volatility relationship: A high frequency analysis of futures and spot exchanges, Int. Rev. Financ. Anal. 91 (Jan. 2024) 103013, https://doi.org/10.1016/j.irfa.2023.103013.
[41] C. Conrad, A. Custovic, E. Ghysels, Long- and Short-Term Cryptocurrency Volatility Components: A GARCH-MIDAS Analysis, J. Risk Financ. Manag 11 (2) (May 2018) 23, https://doi.org/10.3390/jrfm11020023.
[42] S. Corbet, G. McHugh, A. Meegan, The influence of central bank monetary policy announcements on cryptocurrency return volatility, Invest. Manag. Financ. Innov. 14 (4) (Dec. 2017) 60–72, https://doi.org/10.21511/imfi.14(4).2017.07.
[43] M. Di, K. Xu, COVID-19 vaccine and post-pandemic recovery: Evidence from Bitcoin cross-asset implied volatility spillover, Financ. Res. Lett. 50 (Dec. 2022) 103289, https://doi.org/10.1016/j.frl.2022.103289.
[44] I.K. Dias, J.M.R. Fernando, P.N.D. Fernando, Does investor sentiment predict bitcoin return and volatility? A quantile regression approach, Int. Rev. Financ. Anal. 84 (Nov. 2022) 102383, https://doi.org/10.1016/j.irfa.2022.102383.
[45] S. Ding, X. Wu, T. Cui, J.W. Goodell, A.M. Du, Modeling climate policy uncertainty into cryptocurrency volatilities, Int. Rev. Financ. Anal. 102 (Jun. 2025) 104030, https://doi.org/10.1016/j.irfa.2025.104030.
[46] A.H. Dyhrberg, Bitcoin, gold and the dollar – A GARCH volatility analysis, Financ. Res. Lett. 16 (Feb. 2016) 85–92, https://doi.org/10.1016/j.frl.2015.10.008.
[47] C. Eom, T. Kaizoji, S.H. Kang, L. Pichl, Bitcoin and investor sentiment: Statistical characteristics and predictability, ' Phys. Stat. Mech. Appl. 514 (Jan. 2019) 511–521, https://doi.org/10.1016/j.physa.2018.09.063.
[48] A.H. Elsayed, G. Gozgor, C.K.M. Lau, Risk transmissions between bitcoin and traditional financial assets during the COVID-19 era: The role of global uncertainties, Int. Rev. Financ. Anal. 81 (May 2022) 102069, https://doi.org/10.1016/j.irfa.2022.102069.
[49] A.H. Elsayed, G. Gozgor, L. Yarovaya, Volatility and return connectedness of cryptocurrency, gold, and uncertainty: Evidence from the cryptocurrency uncertainty indices, Financ. Res. Lett. 47 (Jun. 2022) 102732, https://doi.org/10.1016/j.frl.2022.102732.
[50] L. Fang, E. Bouri, R. Gupta, D. Roubaud, Does global economic uncertainty matter for the volatility and hedging effectiveness of Bitcoin? Int. Rev. Financ. Anal. 61 (Jan. 2019) 29–36, https://doi.org/10.1016/j.irfa.2018.12.010.
[51] T. Fang, Z. Su, L. Yin, Economic fundamentals or investor perceptions? The role of uncertainty in predicting long-term cryptocurrency volatility, Int. Rev. Financ. Anal. 71 (Oct. 2020) 101566, https://doi.org/10.1016/j.irfa.2020.101566.
[52] G. Figá-Talamanca, M. Patacca, Does market attention affect Bitcoin returns and volatility? Decis. Econ. Financ. 42 (1) (Jun. 2019) 135–155, https://doi.org/10.1007/s10203-019-00258-7.
[53] A.D. Gbadebo, A.O. Adekunle, W. Adedokun, A.-O.A. Lukman, J. Akande, BTC price volatility: Fundamentals versus information, Cogent Bus. Manag 8 (1) (Jan. 2021) 1984624, https://doi.org/10.1080/23311975.2021.1984624.
[54] M. Ghani, U. Ghani, S. Ali, M. Mustafa, R. Kosar, Economic Uncertainty and Bitcoin Volatility: Evidence During COVID-19, J. Predict. Mark. 17 (2) (Dec. 2023) 107–122, https://doi.org/10.5750/jpm.v17i2.2087.
[55] K. Gkillas, M. Tantoula, M. Tzagarakis, Transaction activity and bitcoin realized volatility, Oper. Res. Lett. 49 (5) (Sep. 2021) 715–719, https://doi.org/10.1016/j.orl.2021.06.016.
[56] D. Güler, The Impact of Investor Sentiment on Bitcoin Returns and Conditional Volatilities during the Era of Covid-19, J. Behav. Financ. 24 (3) (Jul. 2023) 276–289, https://doi.org/10.1080/15427560.2021.1975285.
[57] W. Kristjanpoller, M.C. Minutolo, A hybrid volatility forecasting framework integrating GARCH, artificial neural network, technical analysis and principal components analysis, Expert Syst. Appl. 109 (Nov. 2018) 1–11, https://doi.org/10.1016/j.eswa.2018.05.011.
[58] L. Kristoufek, Will Bitcoin ever become less volatile? Financ. Res. Lett. 51 (Jan. 2023) 103353 https://doi.org/10.1016/j.frl.2022.103353.
[59] A. Kufo, A. Gjeci, A. Pilkati, Unveiling the Influencing Factors of Cryptocurrency Return Volatility, J. Risk Financ. Manag 17 (1) (Dec. 2023) 12, https://doi.org/10.3390/jrfm17010012.
[60] N. Kyriazis, S. Papadamou, P. Tzeremes, S. Corbet, The differential influence of social media sentiment on cryptocurrency returns and volatility during COVID-19, Q. Rev. Econ. Financ. 89 (Jun. 2023) 307–317, https://doi.org/10.1016/j.qref.2022.09.004.
[61] S.F. Lehrer, T. Xie, G. Yi, Do the Hype of the Benefits from Using New Data Science Tools Extend to Forecasting Extremely Volatile Assets? in: S. Consoli, D. Reforgiato Recupero, M. Saisana (Eds.), in Data Science for Economics and Finance: Methodologies and Applications Springer International Publishing, Cham, 2021, pp. 287–330, https://doi.org/10.1007/978-3-030-66891-4_13.
[62] C. Liang, Y. Zhang, X. Li, F. Ma, Which predictor is more predictive for Bitcoin volatility? And why? Int. J. Financ. Econ. 27 (2) (Apr. 2022) 1947–1961, https://doi.org/10.1002/ijfe.2252.
[63] S. (Cheng) Long, I. Chatziantoniou, D. Gabauer, B. Lucey, Do social media sentiments drive cryptocurrency intraday price volatility? New evidence from asymmetric TVP-VAR frequency connectedness measures, Eur. J. Financ. 30 (13) (Sep. 2024) 1470–1489, https://doi.org/10.1080/1351847X.2024.2314085.
[64] S. Long, Y. Xie, Z. Zhou, B. Lucey, A. Urquhart, From whales to waves: Social media sentiment, volatility, and whales in cryptocurrency markets, Br. Account. Rev. (May 2025) 101682, https://doi.org/10.1016/j.bar.2025.101682.
[65] M.Á. López-Cabarcos, A.M. Pérez-Pico, J. Piñeiro-Chousa, A. Šević, Bitcoin volatility, stock market and investor sentiment. Are they connected? Financ. Res. Lett. 38 (Jan. 2021) 101399 https://doi.org/10.1016/j.frl.2019.101399.
[66] Š. Lyócsa, P. Molnár, T. Plíhal, M. Širaňová, Impact of macroeconomic news, regulation and hacking exchange markets on the volatility of bitcoin, J. Econ. Dyn. Control 119 (Oct. 2020) 103980, https://doi.org/10.1016/j.jedc.2020.103980.
[67] A. Maghyereh, H. Abdoh, COVID-19 and the volatility interlinkage between bitcoin and financial assets, Empir. Econ. 63 (6) (Dec. 2022) 2875–2901, https://doi.org/10.1007/s00181-022-02223-7.
[68] P.E. Mandaci, E.C. Cagli, Herding intensity and volatility in cryptocurrency markets during the COVID-19, Financ. Res. Lett. 46 (May 2022) 102382, https://doi.org/10.1016/j.frl.2021.102382.
[69] K. Mokni, When, where, and how economic policy uncertainty predicts Bitcoin returns and volatility? A quantiles-based analysis, Q. Rev. Econ. Financ. 80 (May 2021) 65–73, https://doi.org/10.1016/j.qref.2021.01.017.
[70] A. Omura, A. Cheung, J.J. Su, Does natural gas volatility affect Bitcoin volatility? Evidence from the HAR-RV model, Appl. Econ. 56 (4) (Jan. 2024) 414–425, https://doi.org/10.1080/00036846.2023.2168608.
[71] T.V.H. Nguyen, T.V.H. Nguyen, T.C. Nguyen, T.T.A. Pham, Q.M.P. Nguyen, Stablecoins versus traditional cryptocurrencies in response to interbank rates,

Financ. Res. Lett. 47 (Jun. 2022) 102744, https://doi.org/10.1016/j.frl.2022.102744.
[72] J.B. Nouir, H.B.H. Hamida, How do economic policy uncertainty and geopolitical risk drive Bitcoin volatility? Res. Int. Bus. Financ. 64 (Jan. 2023) 101809 https://doi.org/10.1016/j.ribaf.2022.101809.
[73] S. Papadamou, N.A. Kyriazis, P.G. Tzeremes, Non-linear causal linkages of EPU and gold with major cryptocurrencies during bull and bear markets, North Am. J. Econ. Financ. 56 (Apr. 2021) 101343, https://doi.org/10.1016/j.najef.2020.101343.
[74] N. Sabah, Cryptocurrency accepting venues, investor attention, and volatility, Financ. Res. Lett. 36 (Oct. 2020) 101339, https://doi.org/10.1016/j.frl.2019.101339.
[75] F.F. Said, R.S. Somasuntharam, M.R. Yaakub, T. Sarmidi, Impact of Google searches and social media on digital assets' volatility, Humanit. Soc. Sci. Commun. 10 (1) (Nov. 2023) 885, https://doi.org/10.1057/s41599-023-02400-8.
[76] A.A. Salisu, U.B. Ndako, X.V. Vo, Oil price and the Bitcoin market, Resour. Policy 82 (May 2023) 103437, https://doi.org/10.1016/j.resourpol.2023.103437.
[77] A.A. Salisu, A.E. Ogbonna, The return volatility of cryptocurrencies during the COVID-19 pandemic: Assessing the news effect, Glob. Financ. J. 54 (Nov. 2022) 100641, https://doi.org/10.1016/j.gfj.2021.100641.
[78] N. Sapkota, News-based sentiment and bitcoin volatility, Int. Rev. Financ. Anal. 82 (Jul. 2022) 102183, https://doi.org/10.1016/j.irfa.2022.102183.
[79] M. Seo, G. Kim, Hybrid Forecasting Models Based on the Neural Networks for the Volatility of Bitcoin, Appl. Sci. 10 (14) (Jul. 2020) 4768, https://doi.org/10.3390/app10144768.
[80] D. Shen, A. Urquhart, P. Wang, Does twitter predict Bitcoin? Econ. Lett. 174 (Jan. 2019) 118–122, https://doi.org/10.1016/j.econlet.2018.11.007.
[81] L.A. Smales, Investor attention in cryptocurrency markets, Int. Rev. Financ. Anal. 79 (Jan. 2022) 101972, https://doi.org/10.1016/j.irfa.2021.101972.
[82] M. Teterin, A. Peresetsky, Google Trends and Bitcoin volatility forecast, J. N. Econ. Assoc. 65 (4) (2024) 118–135.
[83] K.-Y. Tzeng, Y.-K. Su, Can U.S. macroeconomic indicators forecast cryptocurrency volatility? North Am. J. Econ. Financ. 74 (Sep. 2024) 102224 https://doi.org/10.1016/j.najef.2024.102224.
[84] A. Urquhart, What causes the attention of Bitcoin? Econ. Lett. 166 (May 2018) 40–44, https://doi.org/10.1016/j.econlet.2018.02.017.
[85] G. Uzonwanne, Volatility and return spillovers between stock markets and cryptocurrencies, Q. Rev. Econ. Financ. 82 (Nov. 2021) 30–36, https://doi.org/10.1016/j.qref.2021.06.018.
[86] T. Walther, T. Klein, E. Bouri, Exogenous drivers of Bitcoin and Cryptocurrency volatility – A mixed data sampling approach to forecasting, J. Int. Financ. Mark. Inst. Money 63 (Nov. 2019) 101133, https://doi.org/10.1016/j.intfin.2019.101133.
[87] Y. Wan, Y. Song, X. Zhang, Z. Yin, Asymmetric volatility connectedness between cryptocurrencies and energy: Dynamics and determinants, Front. Environ. Sci. 11 (Jan. 2023) 1115200, https://doi.org/10.3389/fenvs.2023.1115200.
[88] Y. Wang, G. Andreeva, B. Martin-Barragan, Machine learning approaches to forecasting cryptocurrency volatility: Considering internal and external determinants, Int. Rev. Financ. Anal. 90 (Nov. 2023) 102914, https://doi.org/10.1016/j.irfa.2023.102914.
[89] C.-C. Wu, S.-L. Ho, C.-C. Wu, The determinants of Bitcoin returns and volatility: Perspectives on global and national economic policy uncertainty, Financ. Res. Lett. 45 (Mar. 2022) 102175, https://doi.org/10.1016/j.frl.2021.102175.
[90] R. Wu, M.A. Hossain, H. Zhang, Factors affecting the volatility of bitcoin prices, Econ. Financ. Lett. 11 (2) (Apr. 2024) 107–125, https://doi.org/10.18488/29.v11i2.3730.
[91] X. Wu, X. Yin, Z. Umar, N. Iqbal, Volatility forecasting in the Bitcoin market: A new proposed measure based on the VS-ACARR approach, North Am. J. Econ. Financ. 67 (Jul. 2023) 101948, https://doi.org/10.1016/j.najef.2023.101948.
[92] Y. Xia, C. Sang, L. He, Z. Wang, The role of uncertainty index in forecasting volatility of Bitcoin: Fresh evidence from GARCH-MIDAS approach, Financ. Res. Lett. 52 (Mar. 2023) 103391, https://doi.org/10.1016/j.frl.2022.103391.
[93] K.-C. Yen, H.-P. Cheng, Economic policy uncertainty and cryptocurrency volatility, Financ. Res. Lett. 38 (Jan. 2021) 101428, https://doi.org/10.1016/j.frl.2020.101428.
[94] L. Yin, J. Nie, L. Han, Understanding cryptocurrency volatility: The role of oil market shocks, Int. Rev. Econ. Financ. 72 (Mar. 2021) 233–253, https://doi.org/10.1016/j.iref.2020.11.013.
[95] I. Yousaf, S. Ali, M. Marei, M. Gubareva, Spillovers and hedging effectiveness between islamic cryptocurrency and metal markets: Evidence from the COVID-19 outbreak, Int. Rev. Econ. Financ. 92 (Apr. 2024) 1126–1151, https://doi.org/10.1016/j.iref.2024.02.075.
[96] M. Yu, Forecasting Bitcoin volatility: The role of leverage effect and uncertainty, ' Phys. Stat. Mech. Appl. 533 (Nov. 2019) 120707, https://doi.org/10.1016/j.physa.2019.03.072.
[97] W. Zhang, P. Wang, Investor attention and the pricing of cryptocurrency market, Evol. Inst. Econ. Rev. 17 (2) (Jul. 2020) 445–468, https://doi.org/10.1007/s40844-020-00182-1.
[98] S. Zhou, Exploring the driving forces of the Bitcoin currency exchange rate dynamics: an EGARCH approach, Empir. Econ. 60 (2) (Feb. 2021) 557–606, https://doi.org/10.1007/s00181-019-01776-4.
[99] Y. Zhou, C. Xie, G.-J. Wang, J. Gong, Y. Zhu, Forecasting cryptocurrency volatility: a novel framework based on the evolving multiscale graph neural network, Financ. Innov. 11 (1) (Feb. 2025) 87, https://doi.org/10.1186/s40854-025-00768-x.
[100] P. Zhu, X. Zhang, Y. Wu, H. Zheng, Y. Zhang, Investor attention and cryptocurrency: Evidence from the Bitcoin market, PLOS ONE 16 (2) (Feb. 2021) e0246331, https://doi.org/10.1371/journal.pone.0246331.
[101] E.E. Leamer, Regression Selection Strategies and Revealed Priors, J. Am. Stat. Assoc. 73 (363) (Sep. 1978) 580–587.
[102] T.J. Mitchell, J.J. Beauchamp, Bayesian variable selection in linear regression', J. Am. Stat. Assoc. 83 (404) (Dec. 1988) 1023–1032.
[103] A.E. Raftery, D. Madigan, J.A. Hoeting, Bayesian _odels, J. Am. Stat. Assoc. 92 (437) (Mar. 1997) 179–191.
[104] J.A. Hoeting, D. Madigan, A.E. Raftery, C.T. Volinsky, Bayesian model averaging: a tutorial', Stat. Sci. 14 (4) (1999) 382–417.
[105] A. Zellner, On Assessing Prior Distributions and Bayesian regression analysis with g-prior distributions', in: P.K. Goel, A. Zellner (Eds.), in Basic Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti, Amsterdam, 1986, pp. 233–243.
[106] R. Tibshirani, Regression shrinkage and selection via the Lasso, ' J. R. Stat. Soc. Ser. B Methodol. 58 (1) (1996) 267–288.
[107] A.E. Hoerl, R.W. Kennard, Ridge regression: biased estimation for nonorthogonal problems, Technometrics 12 (1) (1970) 55–67.
[108] I. Gurrib, F. Kamalov, O. Starkova, E.E. Elshareif, D. Contu, Drivers of the next-minute Bitcoin price using sparse regressions, Stud. Econ. Financ. 41 (2) (Jun. 2024) 410–431, https://doi.org/10.1108/SEF-04-2023-0182.
[109] M. Kuhn and K. Johnson, Applied predictive modeling. New York: Springer, 2013.
[110] T.J. Hastie, R. Tibshirani, and J.H. Friedman, The elements of statistical learning: data mining, inference, and prediction, 2nd ed. in Springer series in statistics. New York: Springer, 2009.
[111] B. Clarke, E. Fokoue, H.H. Zhang, Principles and Theory for Data Mining and Machine Learning. in Springer Series in Statistics, Springer New York, New York, NY, 2009, https://doi.org/10.1007/978-0-387-98135-2.
[112] L. Breiman, Random Forests, Mach. Learn 45 (1) (Oct. 2001) 5–32, https://doi.org/10.1023/A:1010933404324.
[113] S.M. Amini, C.F. Parmeter, Bayesian model averaging in R, J. Econ. Soc. Meas. 36 (4) (Nov. 2011) 253–287, https://doi.org/10.3233/JEM-2011-0350.
[114] S. Zeugner, M. Feldkircher, Bayesian model averaging employing fixed and flexible priors: the BMS Package for R, J. Stat. Softw. 68 (4) (2015), https://doi.org/10.18637/jss.v068.i04.
[115] P.R. Hansen, A. Lunde, J.M. Nason, The model confidence set', Econometrica 79 (2) (2011) 453–497.
[116] P.R. Hansen, A Test for Superior Predictive Ability, J. Bus. Econ. Stat. 23 (4) (Oct. 2005) 365–380, https://doi.org/10.1198/073500105000000063.
[117] C. Alexander, M. Dakos, A critical investigation of cryptocurrency data and analysis, Quant. Financ. 20 (2) (Feb. 2020) 173–188, https://doi.org/10.1080/14697688.2019.1641347.
[118] M. Parkinson, The extreme value method for estimating the variance of the rate of return, J. Bus. 53 (1) (1980) 61–65.
[119] N. Tripathy, S. Hota, D. Singh, B.M. Acharya, S.K. Nayak, A comprehensive analysis of Bitcoin volatility forecasting using time-series econometric models, Appl. Soft Comput. 178 (Jun. 2025) 113339, https://doi.org/10.1016/j.asoc.2025.113339.
[120] A. Guzmán, C. Pinto-Gutiérrez, M.-A. Trujillo, Trading Cryptocurrencies as a pandemic pastime: COVID-19 lockdowns and Bitcoin volume, Mathematics 9 (15) (Jul. 2021) 1771, https://doi.org/10.3390/math9151771.
[121] S. Corbet, Y. (Greg) Hou, Y. Hu, C. Larkin, B. Lucey, L. Oxley, Cryptocurrency liquidity and volatility interrelationships during the COVID-19 pandemic, Financ. Res. Lett. 45 (Mar. 2022) 102137, https://doi.org/10.1016/j.frl.2021.102137.
[122] S. Kamolov, D. Iskhakov, B. Ziyaev, Machine learning methods in time series forecasting: a review, Ann. Math. Comput. Sci. 2 (Aug. 2021) 10–14, https://doi.org/10.56947/amcs.v2.13.