


Article

An Investigation of the Predictability of Uncertainty Indices on Bitcoin Returns

Jinghua Wang ^{1,*} , Geoffrey M. Ngene ², Yan Shi ³ and Ann Nduati Mungai ⁴

¹ Martin Tuchman School of Management, New Jersey Institute of Technology, 184-198 Central Ave, Newark, NJ 07103, USA

² Stetson School of Business and Economics, Mercer University, Macon, GA 31201, USA; ngene_gm@mercer.edu

³ Computer Science and Software Engineering Department, College of Engineering, Mathematics and Science, University of Wisconsin-Platteville, Platteville, WI 53181, USA; shiy@uwplatt.edu

⁴ Cameron School of Business, University of North Carolina Wilmington, 601 South College Street, Wilmington, NC 28403, USA; mungaia@uncw.edu

* Correspondence: jinghua.wang@njit.edu

Abstract: Policymakers and portfolio managers pay keen attention to sources of uncertainties that drive asset returns and volatility. The influence of uncertainty on Bitcoin has the potential to drive fluctuations in the entire cryptocurrency market. We investigate the predictability of thirteen economic policy uncertainty indices on Bitcoin returns. Using the Random Forest machine learning algorithm, we find that Singapore's economic policy uncertainty (EPU) has the strongest predictive power on Bitcoin returns, followed by financial crisis (FC) uncertainty and world trade uncertainty (WTU). We further categorize these uncertainties into different groups. Interestingly, the predictability of uncertainty indices on Bitcoin returns within the international trade group is stronger compared to other uncertainty categories. Additionally, we observed that internet-based uncertainty measures have more predictive power of Bitcoin returns than newspaper- and report-based measures. These results are robust using various additional machine learning methods. We believe that these findings could be valuable for policymakers and portfolio managers when making decisions related to uncertainty drivers of cryptocurrency prices and returns.

Keywords: Bitcoin; Singapore economic policy uncertainty; economic policy uncertainty; machine learning methods

JEL Classification: C32; G12



Citation: Wang, Jinghua, Geoffrey M. Ngene, Yan Shi, and Ann Nduati Mungai. 2023. An Investigation of the Predictability of Uncertainty Indices on Bitcoin Returns. *Journal of Risk and Financial Management* 16: 461. <https://doi.org/10.3390/jrfm16100461>

Academic Editor: Thanasis Stengos

Received: 26 August 2023

Revised: 15 October 2023

Accepted: 17 October 2023

Published: 23 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Since its launch in 2009, Bitcoin (BTC) has impressed market players with its decentralized innovative system, unprecedented returns, and high risk. BTC is the dominant cryptocurrency not just by market capitalization and trading volume but also by its causal impact on other major cryptocurrencies (Wang and Ngene 2020). Therefore, our study on the impact of multiple uncertainty measures on BTC returns has implications for the entire cryptocurrency market. Studies show that the cryptocurrency market is complex (Scagliarini et al. 2022), with multiple players and high-frequency trading. We investigate the impact of thirteen economic policies- macroeconomic news, international trade- and financial news-related uncertainty measures on BTC returns. These are particularly meaningful to capture the primary sources of uncertainty that drive BTC price returns. The proxy, country-level economic policy uncertainty (EPU) indices capture the risk associated with unclear government policies and regulatory frameworks. This phenomenon may lead to businesses and individuals delaying spending and investments because of uncertainty in the market.

EPU and its impact on BTC returns have recently attracted elevated attention among scholars, policymakers, and portfolio managers. [Wu et al. \(2021\)](#) find that China's EPU has a better explanatory power on Bitcoin returns than global and other national EPUs. [Wang et al. \(2019\)](#) find negligible risk spillover effect from EPU to BTC, even after accounting for different market conditions. Other studies have documented the predictive power of EPU on Bitcoin returns ([Demir et al. 2018](#)). Volatility index (VIX) has been documented to predict BTC returns over different frequencies ([Al-Yahyaee et al. 2019](#)). [Nguyen \(2021\)](#) finds that the US stock market significantly impacted Bitcoin returns during the global pandemic. Other studies have also found a significant impact of the newspaper-based EPU and internet-based uncertainty on BTC returns ([Bouri and Gupta 2021](#); [French 2021](#)).

While existing literature has explored some of the EPU variables impacting BTC returns, important variables that can potentially predict BTC returns have not been fully explored. The current studies center around the impact of a single EPU variable ([Bouri and Gupta 2021](#); [Demir et al. 2018](#); [Al-Yahyaee et al. 2019](#)) or five EPU variables ([Li et al. 2022](#)) on BTC returns. The limited variables can reveal only one side of the coin, resulting in the omission of the other important variables. [Cochrane \(2011\)](#) proposes an interesting asset pricing topic of Factor Zoo. According to the factor zoo theory, it is challenging to determine which factors should be adopted in the asset pricing model. We seek factors that can informatively explain the forecasting results without data overfitting problems. The current literature discusses various statistical and time series models used to identify the "King" factor within the factor zoo. [Smith \(2022\)](#) and [Bryzgalova et al. \(2023\)](#) develop Bayesian frameworks for the time-varying characteristics selection. Their work is consistent with [Harvey's \(2017\)](#) statement that a Bayesian method is a natural solution to the factor zoo. The factor zoo theory is also applied in the current study of the impact of uncertainties on BTC returns, similar to the study of factors used to predict the Chinese stock market ([Li et al. 2023](#)).

In the current study, we seek to answer the following questions. First, among multiple country-level and global-level uncertainty indices, are there key ("king") indices that can reasonably predict BTC returns? Are there critical predicting variables that might be missing in the existing literature? Second, a further concern has recently arisen. After a phased-out ban on cryptocurrency that resulted in an eventual blanket ban on all crypto transactions and mining activities in China since September 2021, which country's EPU has potentially taken over China's role in predicting Bitcoin returns?

Machine learning algorithms can improve the forecasting results explaining future movement patterns ([Witten et al. 2005](#)), unlike traditional statistical methods, which only discover the inference of relationships among variables with no white noise ([Bzdok et al. 2018](#)). Random Forest (RF), proposed by [Breiman \(2001\)](#), is an ensemble algorithm that can improve the forecasting results while reducing data mining issues. Further, the current literature has documented the increased impact of Asian markets on Bitcoin ([Panagiotidis et al. 2019](#); [Wu et al. 2021](#)), especially after China banned cryptocurrency transactions and mining¹. We attempt to understand the impact of this policy that could potentially change the role of an Asian country's EPU in predicting Bitcoin returns.

We first employ an RF machine learning algorithm to evaluate the impact of uncertainty measures (features) on BTC returns. To measure the robustness of empirical results, we apply additional machine learning methods: rule fit (RuleFit), extra trees, and the gradient-boosted trees regressor (GBTR). Second, in line with EPU measurements created by [Baker et al. \(2016\)](#), we identify thirteen uncertainty features derived from newspapers, reports, and the internet². We decompose these features into the nine categories³ that could elaborate on individual categories' importance in predicting BTC returns.

Our study differs from prior studies in three main ways. First, we use the advanced algorithm machine learning methods to investigate the influential power of multiple domestic and global-level uncertainty indices. The supervised learning methods can capture the complex patterns of the dataset and choose the best features (factors) that explain the target (BTC returns) variable. Machine learning identifies the most suitable

features to make precise, out-of-sample forecasting results. These algorithms are crucial in making efficient long-run predictions and help our study uncover the most impactful uncertainty indices on BTC returns. To the best of our knowledge, there are no machine-learning methods used in the existing literature to examine the impact of uncertainty indices on BTC returns.

Second, we complement the existing literature by conducting a factor zoo analysis. The limitations of the traditional statistical models drive scholars and current studies to utilize limited datasets, focusing on the US EPU (Demir et al. 2018), equity market uncertainty (Wang et al. 2019), or trade policy uncertainty (Gozgor et al. 2019). Past studies have not incorporated a comprehensive list of uncertainty measures. We cover thirteen measures of uncertainty and compare their predictability on BTC returns to find the most and least impactful measures. The use of extensive uncertainty measures on BTC returns is missing in current studies (Bouri and Gupta 2021; Demir et al. 2018; Al-Yahyaee et al. 2019; Li et al. 2022).

Third, we decompose the uncertainty indices into nine categories based on their sources and the nature of the data (Table A2 in Appendix A). The uncertainty indices are grouped into different categories that further enrich the role of category variables in the predictability of BTC returns at the category level. Some of the uncertainty indices, such as the world trade uncertainty index and global pandemic, have not been explored in the existing studies. Identification of uncertainty measures with a significant impact on BTC returns would benefit investors, scholars, and policymakers in investment, regulatory, and policy decisions.

The remainder of the paper is organized as follows: Section 2 conducts a literature review, Section 3 explains our data and methodology, Section 4 discusses results, and Section 5 concludes.

2. Literature

With the broader application of big data, machine learning has become an efficient technology for complex data analysis in finance and economics. A typical approach in machine learning is training models to learn the dataset's characteristics, find the best performance of the learning algorithm, and then forecast complex relationships among variables. Unlike traditional statistical models, machine learning approaches efficiently deal with big datasets. Machine learning algorithms significantly enhance data-driven analysis in regression and classification (Witten et al. 2005).

In the past, machine learning algorithms were applied in engineering and other technological areas. Bertomeu et al. (2021) and Liu et al. (2021) recently applied machine learning in accounting and finance studies. Bertomeu et al. (2021) use the algorithmic gradient-boosted regression trees (GBRT) method to detect material misstatements in financial statements and analyze differences between misstatements and irregularities. Liu et al. (2021) apply a deep learning method to predict the price of Bitcoin.

Consistent with the current trend of using disruptive technology in finance, we apply RF algorithms. RF is the ensemble algorithm method consisting of multiple models to improve the accuracy of the forecasting results without progressively changing the training dataset. Developed by Breiman (2001), the RF method is a supervised learning algorithm widely used by engineers. It consists of many uncorrelated trees built on different data samples. RF involves the bagging of uncorrelated trees based on the optimization of randomized nodes. Specifically, the decision trees consist of nodes and branches. Each decision tree has the same nodes but different data and possibilities associated with the nodes, resulting in different outcomes. The predicting results reflect the average of all the decision trees. Zhu et al. (2019) use the RF algorithm to predict loan default, showing that the RF algorithm outperforms logistic regression and other machine learning methods. Their study helps identify potential borrowers who may default on future payments.

We use multiple dominant country-level EPU indices from the Asian markets based on the existing literature. Panagiotidis et al. (2019) find an increased impact of Asian markets'

EPU on BTC returns compared to other geographical markets. Wu et al. (2021) indicated that Singapore’s and Japan’s EPUs have a robust explanatory ability on BTC returns. We chose the EPUs of South Korea, Singapore, and Japan to investigate their predictability on BTC returns.

3. Methodology

3.1. Random Forest Algorithm

Breiman (2001) stated that RF can be used for classification and regression analysis. In the classification case, RF is a classifier consisting of a series of structured classifiers, $\{\xi_m(x, \Omega_m), m = 1, \dots\}$ where Ω_m is the independent random vector for the m^{th} tree, and each tree chooses a unit vote for input x . The model votes for the most popular class after generating many trees.

$$\kappa(X, \hat{Y}) = av_m I(\xi_m(X) = \hat{Y}) - \max_{n \neq \hat{Y}} av_m I(\xi_m(X) = n) \tag{1}$$

Equation (1) is the margin function used to describe the accuracy of RF, where $\xi_1(x), \xi_2(x), \dots, \xi_m(x)$ are an ensemble of classifiers, $I(\cdot)$ is the indicator function, and m is the number of classifiers at each node. The training set is randomly drawn from the distribution of the random vector \hat{Y}, X . The margin function represents the difference between the proportion of the forest’s correct vote and the maximum proportion of all incorrect votes. A more positive margin function means a bigger cushion room in the forest’s correctness. For example, a margin of 0.2 at (X, \hat{Y}) means the forest has correctly picked a class \hat{Y} with 20% more total votes at \hat{Y} , than at any other incorrect classes.

The generalization error is formulated as follows:

$$GE = P_{X, \hat{Y}}(\kappa(X, \hat{Y}) < 0) \tag{2}$$

In the RF algorithm, $\xi_m(X) = \xi(X, \Omega_m)$. Breiman (2001) concludes that the generalization error converges to the following formula as the number of trees increases.

$$P_{X, \hat{Y}}(P_{\Omega}(\xi_m(X, \Omega) = \hat{Y}) - \max_{n \neq \hat{Y}} P_{\Omega}(\xi_m(X, \Omega) = n) < 0) \tag{3}$$

With the regression case, Breiman (2001) proposes that RF is formed by growing trees depending on a random vector Ω .

$$E_{X, \hat{Y}}(\hat{Y} - \xi(X))^2 \tag{4}$$

Equation (4) generates the mean-squared generalization error for any numerical predictor $\xi(x)$. The RF predictor is calculated by the average over m of the trees $\xi(x, \Omega_m)$. As the number of trees in the forest increases to infinity, the mean-squared generalization error is expressed as follows.

$$E_{X, \hat{Y}}(\hat{Y} - av_m \xi(X, \Omega_m))^2 \rightarrow E_{X, \hat{Y}}(\hat{Y} - E_{\Omega} \xi(X, \Omega))^2 \tag{5}$$

Based on Equation (5), a tree’s average generalized error (GE) is defined below.

$$GE(tree) = E_{\Omega} E_{X, \hat{Y}}(\hat{Y} - \xi(X, \Omega))^2 \tag{6}$$

Then, the forecast average generalized error is as follows.

$$GE(forest) \leq \rho GE(tree) \tag{7}$$

where ρ is the weighted correlation between the residuals $\hat{Y} - \xi(X, \Omega)$ and $\hat{Y} - \xi(X, \Omega')$ and Ω, Ω' are independent.

3.2. Cross-Validation

To enhance the accuracy of the forecasting results, we apply the five-fold method for the cross-validation analysis. In supervised learning algorithms, the K-fold cross-validation (CV) (Burman 1989; Stone 1974, 1977; Geisser 1975; Zhang 1993) extends the training, validation, and testing process. The method trains models to capture the underlying patterns without too much noise. The advantage of obtaining a K-fold CV is to solve the issue of overfitting the data.

The study designs the model with five iterations to deal with the problem of model overfitting. The value of the five-fold CV is computed on the average performance of these five iterations. The cross-validation conducts the unbiased dataset assessment by dividing it into training, validation, and testing parts. The observations in the training and validation sets are further divided into five folds: one-fifth for validation and the remainder for training. Eighty percent of the observations are used for model training and validation. The remaining twenty percent of the observations are used to test the model's performance.

4. Empirical Results

4.1. Data and Summary Statistics

We collect monthly data on BTC prices from investing.com and calculate the natural log returns on month t using BTC prices, P . BTC returns are defined as $\ln \ln P_t - \ln \ln P_{t-1}$. We collect the EPU indices data from policyuncertainty.com (accessed on October 2021). The EPU indices dataset measures the role of economic policy by counting the frequency of related articles in newspapers (Baker et al. 2016). Table A1 describes the methodology of collecting economic policy-related uncertainties. There are thirteen uncertainty indices in our study, including business investment and sentiment (BIS), exchange rates (ER), financial crises (FC), financial regulation (FR), interest rates (IR), Japan EPU (JEP), macroeconomic news and outlook (MNEMV), overall equity market volatility (OEMV), Singapore EPU (SEPU), new South Korean EPU (SKE), twitter-based EPU (TEU), world pandemic (WPUI) and world trade uncertainty index (WTUI). The research period is from July 2011 to January 2021, with 1596 monthly observations across the thirteen indices. We convert the quarterly and daily EPU data to monthly data, consistent with the monthly frequency data used in the study. We use the Akaike information criterion (AIC) approach to choose the prior lag parameters for the uncertainty indices, as the algorithm models rely on the lagged modeling dataset.

We further investigate the predictability of the uncertainty at the group (category) level. Table A2 in Appendix A presents the categorized features at various levels on their source and nature. These categories consist of the Asian region index, economic policy index, FC index, newspaper-based index, report-based index, internet-based index, overall US equity index, regulation index, and international trade index. We group Singapore, South Korea, and Japan in the Asian region category. We identify ten indices (BIS, BMPU, ER, FC, FR, IR, JEP, MNEMV, OEMV, and SKE) that rely on the newspapers to collect the uncertainty data as the newspaper-based category. WTUI index relies on the reports to collect the data. Thus, we group it in the report-based category. TEU uses Twitter (X), and we group it in an internet-based category. SEPU is a trade-weighted average of national EPU indices for 21 countries, and WTUI are the trade-related indices grouped in the international trade category.

Table A3 presents the summary statistics of BTC returns and the thirteen uncertainty indices. BTC returns and uncertainty indices have non-normal distribution because the Jarque-Bera statistic is significantly different from zero. SKE has the highest standard deviation (79.99), followed by SEPU (77.14). In addition, all the variables except SEPU demonstrate leptokurtic distribution due to the excess kurtosis. The excess kurtosis describes the higher probability for more outliers than the normal distribution in the sample period, perhaps reflecting a sudden rise or the speculations in the uncertainty indices or BTC prices. The skewness of all variables is greater than zero except for the skewness of

FR and SEPU, which exhibit a near-symmetric distribution. Therefore, the non-normal distribution of these two variables may mainly be driven by their kurtosis.

4.2. The Feature Importance

Figure 1 depicts the predictive power of various uncertainty indices on BTC returns using an RF regressor. The feature importance is essential in generating and predicting BTC returns using machine learning algorithms. Figure 1 provides the statistical magnitude between the feature and the target variable. The most important feature is given a one, and the least important feature is given the most negligible value close to zero. The higher value of the feature indicates the critical role of the uncertainty index in predicting BTC returns and vice versa. SEPU (1.0000) is the most important feature in predicting BTC returns, followed by FC (0.8167), TEU (0.7781), and WTUI (0.6177). It is noteworthy that Singapore’s EPU has a stronger explanatory power than Japan and South Korea in the Asian region. Due to political intervention, China has faded away from the cryptocurrency market, and Singapore has replaced its dominant role of impacting the BTC return. The financial crisis feature is the second significant dominant power driving the BTC returns. Investors’ fear of the unstable financial situation causes significant changes in BTC prices. Investors who believe in BTC would purchase it more when the financial crisis news is mentioned frequently in the market. Conversely, investors who fear holding BTC would liquidate their BTC positions.

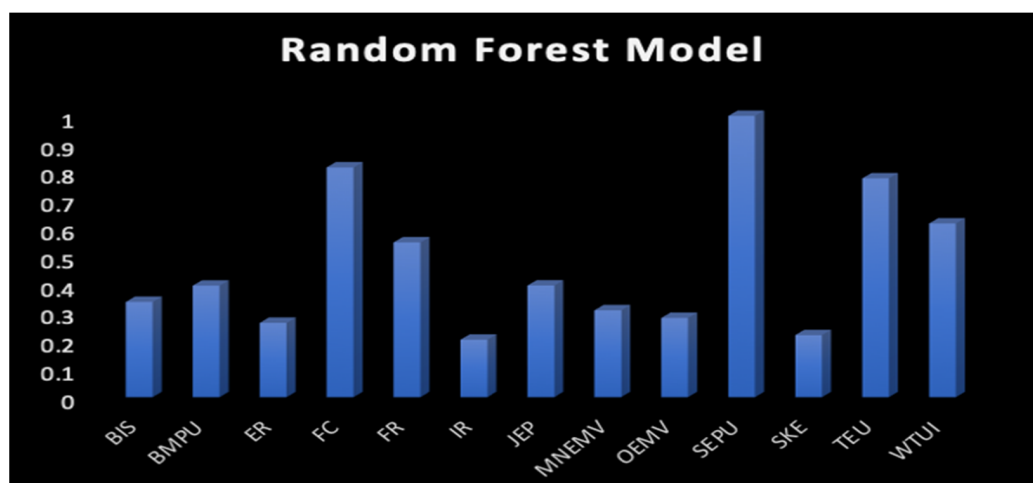


Figure 1. The feature importance. Note: The figure reflects the importance of features in the range of 0 and 1 using the RF algorithm. The Y-axis indicates the magnitude of the importance of features in predicting returns. The X-axis represents the features, including 13 uncertainty-related indices. The higher value indicates the features’ more robust ability to predict BTC returns.

Further, it is not surprising to realize that Twitter-based measure of uncertainty is a significant driver of the BTC returns. Twitter generates revenue mainly from advertisement and data licensing. In 2022, Twitter generated \$4.4 billion in advertising revenue, an 11% decline from the 2021 revenue⁴. It has over 300 million monthly active users⁵. With the growing number of Twitter users, it is generally inferred that investors use internet-based social media more to exchange their attitudes towards investments and help them to make investment decisions in the crypto market dominated by BTC.

The existing literature does not explore the role of the world trade-related uncertainty index. World trade is one of the major economic indicators reflecting the global economic performance. We collect the World trade data from the Economic Intelligence Unit country reports globally (see Table A1). BTC is considered a medium of exchange in executing international trade and payments. BTC is also an investment vehicle, and its market capitalization has significantly increased from \$1.02 billion in the second quarter of 2013 to around \$526.6 billion by mid-September 2023⁶. It is traded in 44 countries worldwide.⁷ The

adoption of BTC in international trade makes it necessary to analyze the impact of world trade-related uncertainty on BTC returns. We find that the interest rate-related uncertainty index plays the most negligible role in predicting BTC returns, suggesting that monetary policy does not impact changes in BTC prices.

Figure 2 demonstrates the importance of category features. We find that financial crises (0.8167) contribute the most in explaining BTC returns, followed by international trade (0.8089) and internet-based index (0.7781). The value of the category reflects the average of the importance of features. The international trade and financial crisis (FC) categories have almost equal strength and magnitude in predicting BTC returns. The uncertainty related to international trade indicates the average importance between SEPU and WTUI. (see Figure 2 and Table A2). We find that the internet-based economic uncertainty index (0.7781) is superior to the newspaper-based (0.3780) and report-based (0.6177) measures in predicting the BTC returns. Our results confirm the findings by Bouri and Gupta (2021). It is interesting to notice that the U.S. equity index-related uncertainty (0.2819) is of low importance in predicting BTC returns. BTC is traded worldwide, 7 days a week and 24 h a day, so the uncertainty index limited to the U.S. equity market is unsuitable for predicting BTC returns. BTC investors should not rely on the performance of the U.S. equity market to make investment decisions in the crypto market.

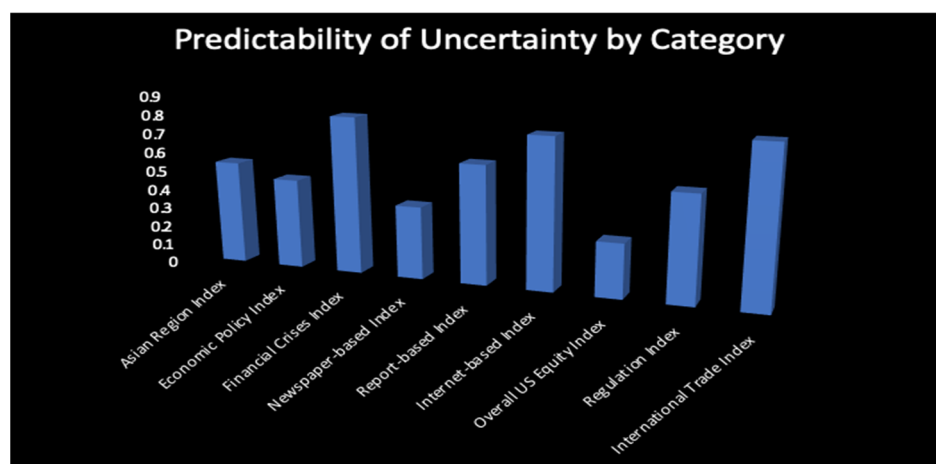


Figure 2. The predictability of uncertainty by category. Note: The categorized uncertainty indices are presented in Figure 2. The Y-axis presents the value of the predictability by category. The X-axis represents features in groups. The predictability is reflected by the magnitude of the feature’s importance. The importance by category is based on the average importance of features in the same category.

4.3. Robustness Checks

To further investigate the predictability of uncertainty indices on BTC returns, we use alternative machine learning algorithms to check the robustness of the empirical results. Table 1 presents the importance of RF, GBTR, extra trees, and RuleFit algorithms. Figure 3 demonstrates the 3-D feature importance to compare the feature importance.

Three models (RF, GBTR, and extra tree) choose SEPU as the most crucial feature in predicting BTC returns. SEPU is also superior to any of the other features in the study in explaining BTC returns. FC feature plays the second most important role (0.8167) based on RF ranking but the third most important role (0.8188) based on GBTR ranking. The TEU feature plays the most, second most, and third most important role based on Rulefit (1.000) Extra tree (0.6635) RF (0.7781), respectively. WTUI plays the fourth important role (0.6177) using RF and GBTR (0.8007) but also the second most crucial predictive role (0.7677) based on RuleFit. The robustness checks further confirm that IR plays the least significant role in predicting BTC returns (0.2046, 0.4005, 0.0472) using RF, GBTR, and extra tree methods,

respectively, and the second most minor role (0.1173) using RuleFit. Our results are robust in explaining the strong predictability of SEPU, FC, TEU, and WTUI on BTC returns.

Table 1. The feature importance using four machine learning algorithms.

Feature	Random Forest Model	GBTR Model	Extra Tree Model	Rule Fit Model
BIS	0.3392	0.5840	0.3261	0.4712
BMPU	0.3967	0.8196	0.2291	0.4131
ER	0.2643	0.5815	0.0541	0.1173
FC	0.8167	0.8188	0.1558	0.6016
FR	0.5504	0.6834	0.2910	0.7521
IR	0.2046	0.4005	0.0472	0.3028
JEP	0.3972	0.6025	0.4807	0.7437
MNEMV	0.3084	0.4979	0.1981	0.6465
OEMV	0.2819	0.6073	0.0908	0.3359
SEPU	1.0000	1.0000	1.0000	0.5150
SKE	0.2208	0.5958	0.1234	0.4311
TEU	0.7781	0.6054	0.6635	1.0000
WTUI	0.6177	0.8007	0.1625	0.7677

Note: The values refer to the relative importance score ranging from 0 to 1. The feature names are sorted alphabetically to facilitate comparison. The value of the relative importance is normalized. The most important feature is normalized to 1. The higher value indicates that the feature is more important in explaining BTC returns, hence a more substantial predictive power. The lower value indicates that the feature has a lower predictive power in the model.

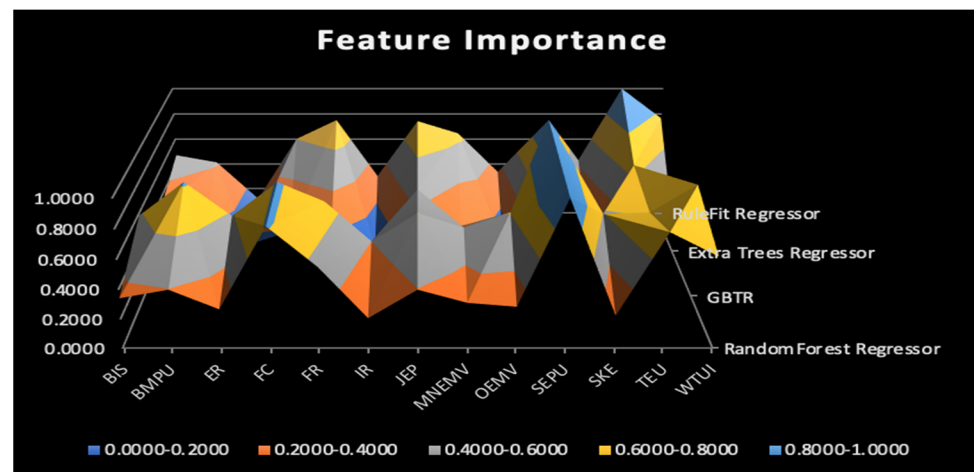


Figure 3. The comparison of feature importance using four machine learning algorithms. Note: The 3-D figure compares the value of the feature importance using the RuleFit Regressor, Extra Trees, RF Regressor, and GBTR models separately. The x-axis indicates the feature’s name, the y-axis indicates the value ranging from 0 to 1, and the z-axis shows the name of an algorithm model.

4.4. Model Accuracy

For computing efficiency, we check the model accuracy. The fitted forecasting model can explain the returns statistically. Table 2 presents the empirical results of model accuracy in forecasting the target BTC returns using four machine-learning algorithms. The root mean squared error (RMSE) measures the model’s inaccuracy. A low RMSE indicates strong predictive accuracy of the model on the target and vice versa. The four models report RMSE values of less than 0.4, suggesting that the models can be implemented to predict BTC returns. RF has the lowest RMSE value (0.2734), meaning that it is the best-fitting model among the four models.

On the other hand, three models (RF, GBTR, and extra tree) report R-squared values above zero. RF has the highest R-squared value (0.1391), further supporting that it is the best model. The RuleFit model reports a negative R-squared value, suggesting the worst performance on its forecasting ability. Based on the model accuracy of our analysis, we can

rely on the results of RF, GBTR, and extra tree models while ignoring the RuleFit model. We use machine learning algorithms to analyze the forecasting accuracy of the individual feature on the target variable, not forecasting the target variable.

Table 2. The comparison of the model accuracy.

Model	Random Forest Model	GBTR Model	Extra Tree Model	Rule Fit Model
Sample Size	64.15%	64.15%	64.15%	64.15%
RMSE (Cross Validation)	0.2734	0.2915	0.2859	0.3553
Residual Mean	0.0095	0.0053	−0.1501	−0.1878
R^2	0.1391	0.0212	0.0462	−0.0979

Note: Sample size is the percentage of the observations used to train and validate the model. We carry out validation after model training. Validation is the process of evaluating a trained model with a testing dataset. RMSE is the root mean squared error used to assess the inaccuracy of predicted mean values. Lower values of RMSE suggest higher predictive accuracy on the target variable. The residual mean of parameters is the average of all residuals, using the total of residuals divided by the number of residuals. R^2 is the coefficient of determination, explaining the variation of the actual series and predicted series. The higher the number, the more accurate the model.

5. Conclusions

Our study expands the existing analysis of the uncertainty factors impacting BTC returns. Specifically, we investigate, using innovative machine learning methods, the predictive power of thirteen economic policy-, financial crisis- international trade-, interest rate- and market volatility-related uncertainties in explaining changes in BTC prices. The study fills the several blanks in the existing literature. First, we find that Singapore EPU is the “King” factor in predicting the BTC returns, possibly due to the Chinese government’s ban on crypto trading and mining. Second, our study is the first attempt at using machine learning algorithms, Random Forest, to analyze the predictability of multiple uncertainty measures on the BTC returns. Third, considering the category of EPU, we find that financial crisis-related uncertainty contributes the most to predicting BTC returns. Fourth, we find that international trade-related uncertainty is missing from the current literature on the study of the predictive power of uncertainty measures on BTC returns. This gap has not been explored until now. Last, we use the RF algorithm to prove the importance of the internet-based uncertainty measure, which is superior to the reports- and newspaper-based uncertainty measures in predicting BTC returns.

The study results provide empirical evidence to researchers, practitioners, portfolio managers, and policymakers on domestic- and global-level uncertainties driving the cryptocurrency returns for regulatory and portfolio management decisions. Future studies can extend the current research study to explore the impact of uncertainties on cryptocurrency volatility using machine learning methods such as explainable Artificial Intelligence.

Author Contributions: Conceptualization, J.W. and G.M.N.; Methodology, J.W.; Software, J.W. and Y.S.; Validation, Y.S.; Formal analysis, J.W.; Resources, G.M.N.; Writing—original draft, J.W.; Writing—review & editing, G.M.N., Y.S. and A.N.M. All authors have read and agreed to the published version of the manuscript.

Funding: The APC was funded by New Jersey Institute of Technology.

Data Availability Statement: Data is unavailable due to privacy.

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

Appendix A

Table A1. Features Definitions.

	Abbreviation	Description	Data Frequency
Business Investment and Sentiment	BIS	Based on the counts of newspaper articles containing the keywords in the category—business investment and sentiment from eleven major U.S. newspapers, multiplying the contemporaneous equity market volatility tracker value that moves with the CBOE Volatility Index and the realized volatility of returns on the S&P 500.	Monthly
BBD Measuring Economic Policy Uncertainty Index	BMPU	The Baker-Bloom-Davis MPU, based on access to the world news indices for the United States, implements the approach developed for measuring economic policy uncertainty.	Monthly
Exchange Rates	ER	Based on the counts of newspaper articles containing the keywords in the category—exchange rates from eleven major U.S. newspapers, multiplying the contemporaneous equity market volatility tracker value that moves with the CBOE Volatility Index and the realized volatility of returns on the S&P 500.	Monthly
Financial Crises	FC	Based on the counts of newspaper articles containing the keywords in the category—financial crises from eleven major U.S. newspapers.	Monthly
Financial Regulation	FR	Based on the counts of newspaper articles containing the keywords in the category—financial regulation from eleven major U.S. newspapers, multiplying the contemporaneous equity market volatility tracker value that moves with the CBOE Volatility Index and the realized volatility of returns on the S&P 500.	Monthly
Interest Rates	IR	Based on the counts of newspaper articles containing the keywords in the category—interest rates.	Monthly
Japan Economic Policy Uncertainty Index	JEP	The index consists of the articles in four major Japanese newspapers.	Monthly
Macroeconomic News and Outlook	MNEMV	The index is based on the counts of newspaper articles containing the keywords in the category—macroeconomic news and outlook from eleven major U.S. newspapers, multiplying the contemporaneous equity market volatility tracker value that moves with the CBOE Volatility Index and the realized volatility of returns on the S&P 500.	Monthly
Overall Equity Market Volatility	OEMV	Based on the average of the standardized scaled counts of newspaper articles containing the keywords to match the mean value of the CBOE Volatility Index. The index tracks the overall equity market volatility for eleven major U.S. newspapers.	Monthly
Singapore Economic Policy Uncertainty Index	SEPU	A trade-weighted average of national EPU indices for 21 countries.	Monthly
New South Korean Economic Policy Uncertainty Index	SKE	The New South Korean Economic Policy Uncertainty (EPU) Index uses six major newspapers in South Korea.	Monthly
Twitter-based Economic Uncertainty Index	TEU	The index extracts all messages (tweets) in English sent on Twitter since June 2011 that contain keywords related to Uncertainty and the Economy.	Daily
World Trade Uncertainty Index	WTUI	It measures trade uncertainty globally using the Economic Intelligence Unit country reports.	Quarterly

Note: The data source is Economic Policy Uncertainty on the website policyuncertainty.com. The quarterly and daily data frequency is converted to the monthly data to be consistent with most data. The linear interpolation method converts the data to the monthly data.

Table A2. Uncertainty by Category.

Category	Feature
Asian Region Index	SEPU, SKE, JEP
Economic Policy Index	BIS, BMPU, ER, FC, IR, JEP, MNEMV, OEMV, SEPU, SKE), TEU, WTUI
Financial Crises Index	FC
Newspaper-based Index	BIS, BMPU, ER, FC, FR, IR, JEP, MNEMV, OEMV, SKE
Report-based Index	WTUI
Internet-based Index	TEU
Overall US Equity Index	OEMV
Regulation Index	FR
International Trade Index	SEPU, WTUI

Note: The total thirteen features are categorized in the various levels. The features belong to one or more categories based on the data sources. For example, SEPU belongs to both levels of the Asian region and is Trade-based.

Table A3. Summary Statistics.

	Mean	Median	Maximum	Minimum	Std. Dev.	Skewness	Kurtosis	Jarque-Bera
BIS	0.496348	0.348	3.9517	0	0.543014	3.151029	17.71081	1131.213 ***
BMPU	74.84245	58.08498	304.0693	18.68333	51.09273	1.840589	7.544355	151.0599 ***
BTC	0.08203	0.064265	1.562375	−0.470078	0.285	1.641923	9.688176	245.1928 ***
ER	0.245037	0.14665	3.8576	0	0.440176	5.786614	44.95514	8365.932 ***
FC	3.950982	3.451	20.4663	1.6502	2.207639	4.272334	31.03031	3792.633 ***
FR	2.408818	2.4325	5.6821	0.7559	0.932568	0.704123	3.784667	11.4783 ***
IR	5.284479	4.6696	19.0173	1.7408	2.744067	2.319844	9.788747	298.6276 ***
JEP	114.818	108.8493	212.6997	62.28234	31.8555	1.044502	4.070694	24.33725 ***
OEMV	19.14626	17.01115	63.3638	9.5696	7.656087	2.560262	12.82233	541.9162 ***
SEPU	179.1436	153.8134	407.7419	82.86535	77.13623	0.880181	2.834384	13.80784 ***
SKE	161.363	137.604	538.1768	55.90073	79.98659	1.74158	7.292713	134.9724 ***
TEU	87.959	71.87471	445.7241	24.56089	67.18413	3.105647	16.07644	925.6157 ***
WTUI	19.95698	1.43	174.34	0.04	39.57344	2.240248	6.887645	155.4164 ***
MNEMV	13.62065	12.23125	46.6632	6.9832	5.732917	2.656577	13.19266	583.5295 ***

Note: The table describes the key summary statistics for thirteen features and BTC. The paper uses the continuously compounded return of BTC derived as the natural log difference. The research period runs from July 2011 to January 2021, with 1596 monthly observations. Jarque-Bera tests show the statistics of the normality. The symbol *** indicates statistical significance at the 1% level.

Notes

- China banned cryptocurrencies on 6/2009, 11/2013, 4/2014, 2/2017, 9/2017 and 5/2021. In September of 2021, China's central bank and its National Development and Reform Commission harshly banned crypto mining and crypto transactions. (source: http://www.gov.cn/zhengce/zhengceku/2021-10/08/content_5641404.htm, accessed in October 2021).
- The features were selected based on the findings of the existing literature and the availability of the data. A full list of features can be seen in Table A1 of Appendix A.
- Table A2 in Appendix A lists nine categories.
- The quarterly earnings report was released by Twitter Inc. on 2 October 2022. <https://investor.twitterinc.com/financial-information/quarterly-results/default.aspx>, accessed in October 2022.
- <https://www.businessofapps.com/data/twitter-statistics/> (accessed on 2 October 2022).
- Data source is from "Bitcoin market capitalization quarterly 2013–2022", <https://www.statista.com>.
- <https://www.statista.com/statistics/1195753/bitcoin-trading-selected-countries/> (accessed on 2 October 2022).

References

- Al-Yahyaee, Khamis, Mobeen Ur Rehman, Walid Mensi, and Idries Mohammad Wanas Al-Jarrah. 2019. Can uncertainty indices predict Bitcoin prices? A revisited analysis using partial and multivariate wavelet approaches. *The North American Journal of Economic Finance* 49: 47–56. [CrossRef]
- Baker, Scott R., Nicholas Bloom, and Steven J. Davis. 2016. Measuring economic policy uncertainty. *The Quarterly Journal of Economics* 131: 1593–636. [CrossRef]

- Bertomeu, Jeremy, Edwige Cheynel, Eric Floyd, and Wenqiang Pan. 2021. Using machine learning to detect misstatements. *Review of Accounting Studies* 26: 468–519. [\[CrossRef\]](#)
- Bouri, Elie, and Rangan Gupta. 2021. Predicting Bitcoin returns: Comparing the roles of newspaper- and internet-based measures of uncertainty. *Finance Research Letters* 38: 101398. [\[CrossRef\]](#)
- Breiman, Leo. 2001. Random Forests. *Machine Learning* 45: 5–32. [\[CrossRef\]](#)
- Bryzgalova, Svetlana, Jiantao Huang, and Christian Julliard. 2023. Bayesian solutions for the factor zoo: We just ran two quadrillion models. *The Journal of Finance* 78: 487–557. [\[CrossRef\]](#)
- Burman, Prabir. 1989. A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika* 76: 503–14. [\[CrossRef\]](#)
- Bzdok, Danilo, Naomi Altman, and Martin Krzywinski. 2018. Statistics versus machine learning. *Nature Methods* 15: 233–34. [\[CrossRef\]](#)
- Cochrane, John H. 2011. Presidential address: Discount rates. *Journal of Finance* 66: 1047–108. [\[CrossRef\]](#)
- Demir, Ender, Giray Gozgor, Chi Keung Marco Lau, and Samuel A. Vigne. 2018. Does economic policy uncertainty predict the Bitcoin returns? An empirical investigation. *Finance Research Letters* 26: 145–49. [\[CrossRef\]](#)
- French, Joseph. 2021. #Bitcoin, #COVID-10: Twitter-based uncertainty and Bitcoin before and during the pandemic. *International Journal of Financial Studies* 9: 28.
- Gozgor, Giray, Aviral Kumar Tiwari, Ender Demir, and Sagi Akron. 2019. The relationship between Bitcoin returns and trade policy uncertainty. *Finance Research Letters* 29: 75–82. [\[CrossRef\]](#)
- Geisser, Seymour. 1975. The predictive sample reuse method with applications. *Journal of the American Statistical Association* 70: 320–28. [\[CrossRef\]](#)
- Harvey, Campbell R. 2017. Presidential address: The scientific outlook in financial economics. *The Journal of Finance* 72: 1399–440. [\[CrossRef\]](#)
- Li, Zheng-Zheng, Chi-Wei Su, and Meng Nan Zhu. 2022. How Does Uncertainty Affect Volatility Correlation between Financial Assets? Evidence from Bitcoin, Stock and Gold. *Emerging Markets Finance & Trade* 58: 2682–94.
- Li, Zhiyong, Yifan Wan, Tianyi Wang, and Mei Yu. 2023. Factor-timing in the Chinese factor zoo: The role of economic policy uncertainty. *Journal of International Financial Markets, Institutions & Money* 85: 101782.
- Liu, Mingxi, Guowen Li, Jianping Li, Xiaoqian Zhu, and Yinhong Yao. 2021. Forecasting the price of Bitcoin using deep learning. *Finance Research Letters* 40: 101755. [\[CrossRef\]](#)
- Nguyen, Khanh Quoc. 2021. The correlation between the stock market and Bitcoin during COVID-19 and other uncertainty periods. *Finance Research Letters* 46: 102284. [\[CrossRef\]](#)
- Panagiotidis, Theodore, Thanasis Stengos, and Orestis Vravosinos. 2019. The effects of markets, uncertainty and search intensity on bitcoin returns. *International Review of Financial Analysis* 63: 220–42. [\[CrossRef\]](#)
- Scagliarini, Tomas, Giuseppe Pappalardo, Alessio Emanuele Biondo, Alessandro Pluchino, Andrea Rapisarda, and Sebastiano Stramaglia. 2022. Pairwise and high-order dependencies in the cryptocurrency trading network. *Scientific Reports* 12: 18483. [\[CrossRef\]](#)
- Smith, Simon. 2022. Time-variation, multiple testing, and the factor zoo. *International Review of Financial Analysis* 84: 102394. [\[CrossRef\]](#)
- Stone, M. 1974. Cross-validated choice and assessment of statistical predictions. *Journal of Royal Statistical Society, Series B* 36: 111–47. [\[CrossRef\]](#)
- Stone, M. 1977. An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of Royal Statistical Society, Series B* 39: 44–47. [\[CrossRef\]](#)
- Wang, Gang-Jin, Chi Xie, Danyan Wen, and Longfeng Zhao. 2019. When Bitcoin meets economic policy uncertainty (EPU): Measuring risk spillover effect from EPU to Bitcoin. *Finance Research Letters* 31: 489–97. [\[CrossRef\]](#)
- Wang, Jinghua, and Geoffrey M. Ngene. 2020. Does Bitcoin still own the dominant power? An intraday analysis. *International Review of Financial Analysis* 71: 101551. [\[CrossRef\]](#)
- Witten, Ian H., Eibe Frank, Mark A. Hall, and Christopher Pal. 2005. *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed. Burlington: Morgan Kaufmann.
- Wu, Chang-Che, Shu-Ling Ho, and Chih-Chiang Wu. 2021. The determinants of Bitcoin returns and volatility: Perspectives on global and national economic policy uncertainty. *Financial Research Letters* 45: 202175. [\[CrossRef\]](#)
- Zhang, Ping. 1993. Model selection via multifold cross validation. *The Annals of Statistics* 21: 299–313. [\[CrossRef\]](#)
- Zhu, Lin, Dafeng Qiu, Daji Ergu, Cai Ying, and Kuiyi Liu. 2019. A study on predicting loan default based on the random forest algorithm. *Procedia Computer Science* 162: 503–13. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.