Uncertainty and Bubbles in Cryptocurrencies: Evidence from Newly Developed Uncertainty Indices

ABSTRACT

In this paper, we examine whether newly developed crypto price and policy uncertainty indices based on news coverage (Lucey et al., 2022) are associated with the emergence of bubbles in cryptocurrencies. Using probit regressions, we show that these indices have a higher explanatory power than factors previously considered in the literature. Furthermore, using a random forest model, we show that these classifiers are associated with the largest information gain (reduction in the Gini impurity measure) of the model. While the COVID-19 pandemic has exacerbated the occurrence of bubbles, these crypto uncertainty indices remain the best predictors of bubbles both before and during the pandemic. These results are robust to alternative definitions of a bubble, variations in the time horizon, and the inclusion of various regressors known to be related to the price movements in crypto assets.

JEL classification: C32; F3; G15

Keywords: Cryptocurrencies; Bubbles; UCRY Price; UCRY Policy; Uncertainty; COVID-19

1. Introduction

New developments in cryptography and advances in distributed ledger technology have led to transformational changes in the financial services industry. Cryptocurrencies emerged as one of the main financial innovations of the 21st century as they afford the processing and recording of transactions between counterparties without the need for financial intermediaries. Because of these developments and in anticipation of further applications of cryptography to payment systems and financial contracts, the cryptocurrency market burgeoned.[1] These developments, however, were accompanied by extreme levels of volatility, swings in investor sentiment and market uncertainty. By the end of 2022 the major cryptocurrencies lost more than 70 percent of their value since their peak in 2021, and some commentators have predicted that the market is entering another "crypto-winter" – a prolonged period of depressed values and subdued investor expectations of wider market adoption or further innovations in the sector.[2]

Pricing cryptographic assets is inherently difficult as market participants cannot rely on standard asset pricing techniques to value these speculative assets. Rather than forming expectations about future cashflows, to value crypto, investors gauge the level of technological innovation in the sector and attempt to predict issues related to security and the speed of adoption of different crypto technologies. Hence, the value of cryptocurrencies is tied to their potential to serve as a store of value and medium of exchange, security considerations, as well as to the investors' expectations about the future demand for these assets. For these reasons, crypto assets

[1] During the COVID-19 pandemic, a substantial amount of capital has flows into the sector attracting funds both from institutional and amateur investors. Consequently, the market cap of its main constituent, Bitcoin, exceeded $1 trillion during the height of the crypto market rally in 2021 while the market cap of the entire sector exceeded $2.9 trillion (https://coinmarketcap.com/).

[2] On 12 November 2021, Bitcoin reached its peak with a closing price of $64,155.94 for that day, while a year later, on 12 November 2022 the closing price was $16,799.18. Similarly, Ethereum declined from $4,667.12 to $1,255.27 over the same period (https://coinmarketcap.com/).

are extremely volatile and susceptible to bubbles. The formation of bubbles in crypto assets, however, are phenomena that are still not well understood, and the factors contributing to bubbles are widely under researched.

In this paper, we aim to address this gap by examining whether uncertainly in the crypto market contributes to the emergence of bubbles in the major cryptocurrencies. Bubbles are viewed either as significant deviations from fundamental values (Stiglitz, 1990), or as periods of rapid price appreciation followed by a crash (Phillips, Shi, and Yu 2015, PSY hereafter). As cryptocurrencies do not lend themselves to traditional fundamental analysis, we opt for the latter formulation of a bubble and apply time series tests to identify the bubbles in each individual currency. We use the time series tests developed by PSY to identify episodes of explosive behavior and date-stamp these episodes as bubble periods. Our main question of interest is whether price and policy uncertainty specifically related to the crypto market are driving the formation of crypto bubbles. To that end, we analyze the price (UCRY Price) and policy (UCRY Policy) uncertainty indices developed by Lucey et al. (2022) for the crypto market. These uncertainty indices are based on newspapers and news-wire feeds from the Lexis-Nexis Business Database following the methodology by which economic policy uncertainty indices are developed for the economy (Baker et al., 2016).

Analyzing data from six major cryptocurrencies and 6,924 cryptocurrency-day observations, for the period from June 1, 2016, to May 31, 2021, we find that the crypto price and policy uncertainty indices are strong predictors for the future emergence of crypto bubbles. These results are robust to changes to the forecasting horizon, to alternative formulations of bubbles (i.e., a quantile regression of returns) and to the inclusion of various regressors known to be related to price movements in crypto assets, in particular the volatility index (VIX), commodity price (Brent oil and gold), measures of credit risk (TED-spread), economic policy uncertainty as well as the trading and the search volume for cryptocurrency keywords from Google Trends. Furthermore, estimating a random forest model, we find that the crypto price and policy uncertainty indices have the highest feature importance score of all regressors, i.e., are the classifiers contributing to the larges reduction of the weighted Gini impurity of the data.

2. Related Literature

The determinants of cryptocurrency returns and volatility have received much attention in the recent literature. While the literature has initially focused on Bitcoin (see, e.g. Grinberg, 2012; Baek and Elbeck, 2015; Balcilar et al., 2017; Kristoufek, 2015; Conrad et al., 2018; Pagnotta and Buraschi, 2018; Fang et al., 2019; and Biais et al., 2022) recent studies have expanded this set to include Ethereum, Litecoin, Ripple, Stellar and other major cryptocurrencies (see, e.g., Bouri et al., 2018; Demir et al., 2018; Walther et al., 2019; Gkillas et al. 2022). This literature strand has identified multiple macroeconomic factors with explanatory power for the return and volatility dynamics of crypto, the most important of which include global economic policy uncertainty,[3] trade policy uncertainty,[4] geopolitical risk,[5], investor sentiment,[6] measures of financial stress and real economic activity,[7] the price dynamics of energy and commodities, the performance of the stock market, as well as measures of credit risk in corporate bond markets. Furthermore, Fang et

[3] Indices of economic policy uncertainty were initially developed by Baker et al. (2016).
[4] Gozgor et al. (2019) find a positive relationship between US trade policy uncertainty and Bitcoin returns but show that the impact is negative during regime changes (2010–11 and 2017–18).
[5] Long et al. (2022) show that cryptocurrencies with the lowest exposure to geopolitical risk (lowest geopolitical beta) generate higher returns compared to the ones with the highest sensitivity to geopolitical risk.
[6] Corbet et al. (2020b) construct an investor sentiment index based on news and find that positive news regarding unemployment rates and durable goods result in high stock returns and low Bitcoin returns, while negative news contribute to high Bitcoin returns. Such a behavior would occur if the crypto market were dominated by active amateur investors.
[7] Real economic activity is a key driver of cryptocurrency volatility. Walther et al. (2019) study the CRIX index developed by Trimborn and Härdle (2018) for five cryptocurrencies (Bitcoin, Ethereum, Litecoin, Ripple, and Stellar) and show that global real economic activity outperforms all other predictors considered in their study.

al. (2020) show that the news-based implied volatility index (NVIX),[8] which is constructed from publications related to cryptocurrencies, is negatively associated with long-term volatility of cryptocurrencies, a result which holds after controlling for global economic policy uncertainty.[9]

In parallel to this literature, a growing strand has begun to emerge that explores the formation of bubbles (see, e.g. Cheung et al., 2015; Corbet et al., 2018; Bouri et al., 2019; Cheah and Fry, 2015; Fry and Cheah, 2016; Enoksen et al., 2020; Liu et al., 2020; Aloosh et al. 2022, Maouchi et al. 2022; and Chowdhury et al., 2022). Most closely related to the current analysis in terms of empirical method is the study by Enoksen et al. (2020) which also relies on PSY tests to identify bubble periods. They find, in accordance with expectations based on studies of traditional assets, that high volatility, high number of transactions and high trading volume contribute to the occurrence of bubbles. Surprisingly, while economic policy uncertainty (EPU) exhibits a positive association with bubble periods, the VIX-index shows a negative relationship with the occurrence of bubbles. We include all these variables in our analysis and find similar relationships in terms of directional impact and statistical significance. We show, however, that the impact of these variables is dwarfed by the impact of the crypto price and policy uncertainty indices (UCRY Price and UCRY Policy) constructed by Lucey et al. (2022). Furthermore, using a random forest model, we demonstrate that indeed these two features have the highest importance scores in this machine learning model. That is, partitioning the data sample on these variables leads to the highest information gain (largest reduction in the Gini impurity measure) in the model.

[8] NVIX is a text-based uncertainty index starting in 1980 using the cover page of *The Wall Street Journal* and developed by Manela and Moreira (2017).

[9] In addition to factors related to the macroeconomy and financial markets, the recent literature has also explored the impact of factors related to the distributed ledger technology. These include demand-driven factors such as censorship resistance, network effects and the ability for consumers to engage in exchanges as well as supply-side factors such as the marginal cost of mining and the competition among miners (Pagnotta and Buraschi 2018; Biais et al. 2022; Cong et al. 2021; Liu and Tsyvinski 2021).

3. Research Design

3.1. Data Sample

We collect the daily closing prices of six cryptocurrencies: Bitcoin (BTC), Dogecoin (DOGE), Ethereum (ETH), Litecoin (LTC), Stellar (XLM), and Ripple (XRP) from the CoinMarketCap[10] database. Our sample spans the period from June 1, 2016 to May 31, 2021. Although we initially consider the top 20 cryptocurrencies based on their market cap, the requirement of an uninterrupted coverage during the entire sample period leaves us with the sample of the above six cryptocurrencies.[11] This sample includes also holiday and weekend observations. We used all these uninterrupted time series to derive the bubble periods. We were also able to obtain uninterrupted daily Google Trend Index data for these six major cryptocurrencies.

Our explanatory variables are collected from several other sources. The two main variables of interest are the news-based cryptocurrency uncertainty indices [UCRY Policy and UCRY Price] constructed by Lucey et al. (2022).[12] We collect trading volume data from the CoinMarketCap database. The 'volume' variable measures the daily volume of the six cryptocurrencies studied as reported in Coinmarketcap.com. Further, we include the TED-spread, the VIX-index, gold prices, Brent oil prices, and economic policy uncertainty index (EPU) from the FRED database.[13] The search volume for the names of the six crypto currencies is retrieved from the Google Trends.[14] Some of our control variables related to financial markets, however (e.g., the VIX index) are unavailable during weekends and holidays. We, therefore, removed these observations in our

[10] See https://coinmarketcap.com/
[11] Further details on our sample selection is provided in Table OA1 of the online appendix. This table shows that 14 of the coins have only partial data coverage in our sample period.
[12] The authors have made these indices publicly available at https://sites.google.com/view/cryptocurrency-indices/home?authuser=0.
[13] See https://fred.stlouisfed.org/
[14] See https://trends.google.com/trends/?geo=US

regression analysis in which we regressed the bubble dummy (or alternative proxy for a bubble) on the explanatory variables. This restriction leaves us with an overall sample of 6,924 cryptocurrency days.

3.2. Dependent Variable: Cryptocurrency Bubbles

Following prior studies (e.g., Cheung et al., 2015; Su et al., 2018; Corbet et al., 2018; and Bouri et al., 2018), we use the generalized supremum augmented Dickey–Fuller (GSADF) test by Phillips et al. (2015, PSY hereafter) to test for bubbles and date-stamp the periods of bubbles for each of the six cryptocurrencies. PSY is a widely used method to date stamp bubbles,[15] which has been applied to identify bubbles in markets such as energy, real estate, commodities, and financial assets.[16] The PSY test is based on the following augmented Dickey–Fuller (ADF) standard regression equation:

$$\Delta P_t = a_{r_1,r_2} + \beta_{r_1,r_2} P_{t-1} + \sum_{i=1}^{k} \varphi_{r_1,r_2}^{i} \Delta P_{t-i} + \varepsilon_t, \tag{1}$$

where P_t is the logarithmic of cryptocurrency price at time, t, a, β, φ are parameters calculated using OLS, k is the number of lags selected by BIC[17], ΔP_t is the first difference of P_t. The error term is assumed to follow a normal distribution [i.e., $\varepsilon \sim iidN(0, \sigma_{r_1,r_2}^2)$]. Further, $r_2 = r_1 + r_w$ ($r_w > 0$) is a rolling window, where the starting point r_1 and the ending point r_2 are denoted as

---

15 This method also accounts for nonlinearity and structural breaks in the time series.
16 See, e.g. Bouri et al. (2019).
17 Following Shi (2017), we select the number of lags k by BIC, whereby we allow for a maximum lag order of six. The lag terms of ΔP_t are included to control of potential serial correlation.

fractions of the total sample. Using equation (1), we estimate a sequence of test statistics $ADF_{r_1}^{r_2} = \hat{\beta}_{r_1,r_2}/\ s.e.\ (\hat{\beta}_{r_1,r_2})$ that depend on the fractions of the total sample r_1 and r_2.

PSY propose a generalized GSADF test and allow for flexible rolling windows in which both the starting and the ending date of the window can be varied as follows

$$GSADF\ (r_0) = \sup_{\substack{r_1 \in [0, r_2 - r_0] \\ r_2 \in [r_0, 1]}} ADF_{r_1}^{r_2}, \tag{2}$$

Using equation (2) we detect a bubble when the $GSADF_{r_0}$ statistic is greater than the corresponding right tail critical value obtained by Monte Carlo simulation. As a next step, we date-stamp the starting and ending points of the bubble period based on the following backward sup ADF statistics (BSADF),[18]

$$BSADF_{r_2}(r_0) = SADF_{r_1}^{r_2}\ , \tag{3}$$

where the origination of the subsample can vary from 0 to $r_2 - r_0$ and the end point of the sample is fixed at r_2. The beginning of an explosive behavior period, denoted by $\hat{r}_e$, is given by

$$\hat{r}_e = \left\{r_2 : BSADF_{r_2}(r_0) > scv_{r_2}^{\alpha}\right\}, \tag{4}$$

Likewise, the ending date of a bubble episode, denoted by $\hat{r}_f$, is the first observation after $\hat{r}_e + 3/T$[19]

[18] Note that the GSADF test is an ex-post method used to establish the existence of a bubble in time series rather than identifying the beginning and ending dates of exuberance. Because our goal for the paper is to date-stamp periods of exuberance, following Shi et al. (2016), we use the BSADF test to identify bubble periods.
[19] Following Etienne et al. (2014), we used 3/T in equation (6) to ensure that a period lasts for at least 3 days to be classified as a bubble period.

$$\hat{r}_f = \{r_2 : BSADF_{r_2}(r_0) < scv_{r_2}^{\alpha}\}, \quad (5)$$

where T is the sample size, $SCV_{r_2}^{\alpha}$ is the 100(1-α)% critical value of the sup ADF based on $\lfloor r_2 T \rfloor$ observations at a significance level α. Following PSY, we used the Monte Carlo simulation with 2,000 replications to estimate critical value sequences for the $BSADF_{r_2}$ test statistic.[20]

Further, following prior studies (e.g., Enoksen et al., 2020), we also used the BSADF statistics given in equation (3) as an additional dependent variable for our robustness tests.

3.3. Independent Variables: UCRY Policy and UCRY Price Indices and Control Variables

We use the measures of crypto price and policy uncertainty [UCRY Policy Index and the UCRY Price Index] constructed by Lucey et al. (2022). These indices track the frequency with which keywords such as "uncertainty", "regulator", "central bank" or "government" appear in conjunction with the main cryptocurrencies (Bitcoin, Ethereum, Ripple, Litecoin, etc.) in news articles from the LexisNexis Business database. The construction methodology of these indices is based on the methodology for the construction of economic policy uncertainty indices created by Baker et al. (2016). The UCRY Policy and UCRY Price indices are derived from 726.9 million news articles including many newspapers and newswire feeds. These indices respond to major global economic and political events as well as events related to the crypto market. Lucey et al. (2022) find movements in the UCRY indices associated with the Brexit vote, the 2016 US presidential election, but also with major hacking attacks on cryptocurrency exchanges and the onset of the COVID-19 pandemic. The UCRY indices measure uncertainty of the crypto market

[20] We selected the minimum window size based on the rule of thumb suggested by PSY, form r_0=0.01+1.8/√T.

beyond Bitcoin and capture the impact of price, policy, and regulatory debate on cryptocurrency returns and volatilities.

Crypto indices have also been developed by several other authors. Huang and Luk (2020) developed a new China economic policy uncertainty index, while Trimborn and Härdle (2018) developed the Royalton CRIX Index.[21]

We selected the Cryptocurrency Uncertainty Indices developed by Lucey et al. (2022) because they offer a comprehensive measure that encompasses both price and policy uncertainty in the cryptocurrency market. Previous research has primarily focused on uncertainty measures related to the stock market (such as volatility (VIX), economic policy uncertainty (EPU), and the TED-spread), and we included them in our study. As argued by Elsayed et al. (2022), these measures may not fully capture the heightened uncertainty unique to the cryptocurrency domain.

3.4. Empirical Model

We estimate the following probit regression model:

$$Prob\ (BUBBLE_{i,t} = 1) = \Phi(\beta_0 + \beta_1\, UCRY_{i,t-1} + \Sigma_k \beta_k Controls_{i,t-1}^k + \varepsilon_{it}) \tag{6}$$

Hereby $BUBBLE_{i,t}$ is an indicator variable taking on the value of 1 if cryptocurrency i is in a bubble on day t and 0 otherwise, while the function $\Phi(\cdot)$ is the standardized cumulative normal distribution function. Our main variable of interest is the cryptocurrency uncertainty, $UCRY_{i,t-1}$, which is either the price (UCRY Price) or the policy uncertainty (UCRY Policy) in the crypto

[21] See https://www.royalton-crix.com/

market (Lucey et al. 2022).[22] To facilitate comparison to existing studies (e.g., Enoksen et al., 2020; and Maouchi et al., 2022), we include the trading volume, TED-spread, VIX-index, gold, Brent oil, EPU, and Google Trends as control variables in our empirical specification. All variables are log-transformed for the regression analysis.[23] As a robustness check, we also estimated a linear regression by using the BSADF statistics as a dependent variable:

$$BSADF_{i,t} = \beta_0 + \beta_1\, UCRY_{i,t-1} + \Sigma_k \beta_k Controls_{i,t-1}^{k} + \varepsilon_{i,t}, \qquad (7)$$

Furthermore, we estimated a random forest (RF) model while using the same variables as bubble classifiers.

4. Empirical Results

4.1. Descriptive Statistics and Correlations

The descriptive statistics for the six cryptocurrencies are presented in Table 1, and the descriptive statistics of the explanatory variables are reported in Table 2. The Pearson correlation matrix is presented in Table 3. We observe that the UCRY Price and the UCRY Policy uncertainty variables are positively correlated with the bubble variable with correlation coefficients of 0.250 and 0.258, respectively. As can be observed in column (1), UCRY Price and UCRY Policy are more strongly correlated with bubble periods relative to the above mentioned three uncertainty measures. The correlation between the bubble and other macro variables are relatively low, and

[22] The UCRY indices are weekly and all the remaining variables are daily. To link the weekly UCRY index with daily variables, we replicated the weekly values for each day within the corresponding week. For example, if the original dataset covered the period from January 4, 2021, to January 10, 2021, and the UCRY Policy had a value of 103.41, we assigned a value of 103.41 to the UCRY Policy for each day from January 4 to January 10 in 2021. We applied the same approach to the UCRY Price variable.

[23] Google Trends returns a zero search volume on some days and therefore we transform this variable using the function ln(1+GoogleTrend).

the correlations between the explanatory variables are not particularly high indicating that multicollinearity is not an issue for the empirical model.

4.2. Bubble Detection

The results from the PSY bubble detection procedure are presented in Figure 1. The blue lines represent the time series of cryptocurrency prices. The green line shows the sequence of BSADF statistics calculated using equations (4) and (5) with the corresponding 95% critical values given by the dash orange line. The bubble periods are represented by the green shaded areas. They are obtained when the BSADF statistics (the green line) exceed the 95% critical value given by the orange line. From the graph one can see that Bitcoin and Ethereum are more susceptible to bubbles and the bubble periods are longer compared to the other crypto currencies.

4.3. Baseline Regression Results

The baseline regression results of the model given in equation (6) for the price and the policy uncertainty indices are presented in Tables 4 and 5, respectively. In Column 1, we report the average marginal effects for the univariate model, and in Columns 2–8 we add one control variable at a time. Column 8 of Table 4 includes all regressors and shows that the marginal effect of *UCRY Price* is positive and highly statistically significant (coefficient = 5.509; $p < 0.01$). A one standard deviation (=1.367) increase in *UCRY Price* leads to 7.53% (=5.509×1.367%) increase in the probability of being in a bubble on the following day. Relative to the sample average probability of a bubble (=0.17), this amounts to an increase of 44.29% (=7.53%/0.17) for the probability of a bubble.[24] As a robustness check, we also estimate the model with a lag of seven days and obtained

[24] To ensure the robustness of our results, we re-estimated our baseline regression with all 16 cryptocurrencies that are not stable coins for the longest period for which data on all these cryptocurrencies are available. These results are reported in Online Appendix Table OA2 and Table OA3. They are very similar to our initial findings based on the

similar results (see Table A.1 in the Appendix). We also estimate a linear probability model and obtain qualitatively similar results (see Table A.3 in the Appendix).

The impact of UCRY Policy on bubbles is reported in Table 5, and this relationship is also positive and statistically significant (coefficient in Column 8 = 5.853; $p < 0.01$). The effects are of a similar magnitude. One standard deviation (=1.321) increase in UCRY Policy increases the probability of being in a crypto bubble on the following day by 7.73% (=5.853×1.321%). Relative to the sample average probability of a bubble (=0.17), this increases the probability of a bubble by 45.47% (=7.73%/0.17) relative to the mean. As a robustness check, we also estimate the model with a lag of seven days and obtained similar results (see Table A.2 in the Appendix).[25] We also fit a linear probability model and obtain largely similar results (see Table A.4 in the Appendix).

Further, our control variables (see Tables 4 and 5) show signs and significance in line with prior studies (e.g., Enoksen et al., 2020; and Maouchi et al., 2022). For instance, volume, EPU, and G-Trend elevate the probability of a bubble, whereas VIX, gold, Brent oil, and TED dampen the probability of a bubble. One way to explain these results is that EPU is a backward-looking measure obtained by measuring the frequency of keywords such as "Federal Reserve," "legislation," "regulation," and "congress" among others in leading US newspapers. If EPU proxies confidence in fiat currencies, cryptocurrencies might appear as a viable substitute which might lead to higher demand and the formation of bubbles. The VIX, on the other hand, incorporates the expected future volatility of the stock market, and by extension, of alternative

sample of six cryptocurrencies. In particular the UCRY Price and UCRY Policy remain highly statistically significant in all regression specifications. We are grateful to an anonymous referee for suggesting these robustness checks.

[25] While predictability exists also for somewhat longer lags, predictability naturally declines when one tries to predict further into the future (i.e. when the lags become too long). UCRY Price Index becomes insignificant after a lag of 35 days or 5 weeks while UCRY Policy Index becomes insignificant after a lag of 28 days or 4 weeks.

investments including crypto. Thus, the VIX index could imply a heightened risk for all investments, including crypto.

4.4. Random Forest Model

To account for nonlinear relationships and to assess the relative importance of the considered features, we also estimate a random forest model (Breiman, 2001). Due to its ability to account for complex relationships, this machine learning algorithm has the potential to provide more accurate forecasts than standard regressions (Strobl et al., 2008; Mullainathan and Spiess, 2017; Jones, 2017; Altman and Krzywinski, 2017; and Amini et al., 2021). The advantage over the current probit specification is that it offers more flexibility in allowing for non-linear relationship between the features and the assumed latent variable of the probit model. Furthermore, it avoids the well-known overfitting problem of the decision tree model (Liu and Wu, 2017).

Relative to the probit model, the random forest model allows more flexibility in the transformation of the latent variable into a probability which is not bound to be a standardized cumulative distribution function. We provide here a non-technical description of the random forest model along with the routines we followed. A more detailed description of the algorithm can be found in Hastie et al. (2008), pp. 587-604.

We apply the random forest model on the estimated indicator for bubbles based on the S&P Cryptocurrency Broad Digital Market (BDM) Index[26] as this index reflects the broader crypto market. For the current application, we use 70 percent of our data to train the model and 30 percent to test it. We perform bootstrapping (sampling with replacement) to randomly select three features (the square root from the set of the nine features in the dataset) to train 500 decision trees. The

[26] https://www.spglobal.com/spdji/en/indices/digital-assets/sp-cryptocurrency-broad-digital-market-index/#overview

prediction of the random forest model is generated by aggregation using the majority decision of the trees in the forest. This bootstrapping and aggregation (bagging) procedure ensures that we do not overfit the data and thus offers an advantage over the classification and regression tree model.

The ranking of importance of the variables depends on the information gain attained by the classifier (the mean reduction in the Gini impurity score) is presented in Table 6. The UCRY Price and UCRY Policy are the features associated the largest reduction of the Gini impurity score. We also performed this ranking for the subsamples prior to and during the COVID-19 pandemic. These results are presented Table 7. We observe that the importance score of UCRY Policy exceeds that of UCRY Price before the pandemic, while after the pandemic the ranking is reversed. The two features, however, dominate all other features in both subsamples. To shed light on the predictive ability of the random forest model, we report the associated confusion matrix in Table 8, which shows that the classification accuracy of the UCRY Price Model is 95.09% and of the UCRY Policy Model is 94.95%.

4.5. Role of Protocols

The reported results so far are based on the pooled data for all cryptocurrencies. Previous studies, however, have demonstrated that different digital assets respond differently to uncertainty (Lucey at al., 2020; and Corbet et al., 2020a). In particular, Irresberger et al. (2021) show that the price dynamics of cryptocurrencies depends on the consensus protocol that they are based on (see also Chowdhury et al., 2022). For the studied period, Bitcoin, Ethereum, Litecoin, and Dogecoin used a proof-of-work protocol while Ripple and Stellar use protocols that are not easily categorized into proof-of-work or proof-of-stake, and hence classed as nonstandard protocols.[27] Therefore, to

---

[27] Ethereum completed the switch to a proof-of-stake consensus protocol on September 15, 2022 which reduced its energy consumption by more than 99 percent.

examine the role of protocols, we divided the sample into nonstandard protocols (Ripple and Stellar) and other protocols (Bitcoin, Ethereum, Litecoin, and Dogecoin).

In Table 9 we report the coefficients on UCRY Price and UCRY Policy for nonstandard protocols than for other protocols. While the impact of UCRY Price is significant for both type of protocols, the effect is smaller for non-standard protocols (the coefficient reported in Column 1 is 3.553 while the coefficient in Column 2 is 0.845). Performing a Z-test, we reject the null that the coefficients are equal to each other at the 1% level.[28] The results for the UCRY Policy variable are also similar (see Column 3 and Column 4) with the effect being smaller for non-standard protocols and the difference in the coefficient being highly statistically significant. Our finding show that uncertainty has influence on digital assets using non-standard protocols.

4.6. Role of COVID-19

We also investigate the role of the COVID-19 pandemic and present the results in Table 10. Following Maouchi et al. (2022), we define the period before January 22, 2022 as the pre-pandemic period (taking on the value of 0) and the period starting on that day as the COVID-19 period (taking on the value of 1). The results, which are presented in Table 10, show that COVID-19 increased the probability of cryptocurrency bubbles.

5. Additional Robustness Checks

In this section we show that our results are robust to alternative measures of a bubble and alternative data samples.

[28] The Z-statistic given by $Z = {(b_1 - b_1)}\Big/{\sqrt{SEb_1^2 + SEb_1^2}}$ (see Paternoster et al., 1998) is 4.15 for the probit model and 3.67 for the linear model with the critical value being 2.58 at the 1% significance level.

5.1. Alternative Measures of a Bubble

We begin by considering alternative definitions of a bubble. Instead of using the PSY test we directly use the BSADF statistics derived from equation (4) as a proxy for a bubble (see also Enoksen et al., 2020 for such a robustness test). Our findings show that a bubble in both UCRY Price and UCRY Policy are positively and significantly related to bubbles (see Tables A.5 and A.6 in the Appendix).

As a second alternative to the PSY test, we identify bubble periods as periods in which return fall within the 95% quintile.[29] These regression results are reported in Table 11. The UCRY Price index has a positive and significant impact on the probability of a bubble (Column 2; coefficient = 3.371; $p < 0.01$). The UCRY Policy index is also positive and statistically significant (Column 4; coefficient = 3.118; $p < 0.01$). Thus, our results are not specifically linked to the PSY methodology for identifying bubbles.

5.2. Alternative Data Sources

Because of data availability and convergence issues with the probit model, we examined bubbles for six cryptocurrencies. To mitigate potential concerns with our sample selection, we identified bubbles in the S&P Cryptocurrency Broad Digital Market (BDM) Index.[30] The impact of Cryptocurrency Uncertainty Indices on these bubbles is reported in Table 12 (see columns 2 and 4). The coefficient for UCRY Price is 10.20 and for UCRY Policy is 12.07 with both indices being highly statistically significant ($p < 0.01$). The effect of the uncertainty indices on the

[29] Applying Granger-causality in quantiles tests, Corbet et al. (2020) find a bidirectional causal relationship between Bitcoin and seven altcoins, with the predictability results being stronger during bull markets.
[30] Walther et al. (2019) also focuses on Bitcoin, Ethereum, Litecoin, Ripple, and Stellar as well as the Cryptocurrency index CRIX.

formation of bubbles in the BDM index is even greater. One standard deviation (=1.367) increase in UCRY Price results in a 13.94% (=10.20×1.367%) increase in the probability of bubbles in cryptocurrencies. Because the sample average for S&P cryptocurrencies bubble is 0.234, this economic impact is translated into an increase in the probability of bubbles by 59.57% (=13.94%/0.234) relative to the mean. Further, the coefficient (=12.07) of UCRY Policy in Table 11, column 4, implies that a standard deviation (=1.367) increase in UCRY Policy is associated with a 16.50% (=12.07×1.367%) increase in the probability of bubbles in cryptocurrencies. This corresponds to an increase in the probability of bubbles by 70.51% (=16.50%/0.234) relative to the mean.

6. Conclusion

Cryptocurrencies are speculative assets exhibiting high levels of volatility and market bubbles. While bubbles have been a widely observed phenomenon in digital markets, their determinants and origin are not well understood. In this paper, we examine whether newly developed measures of price and policy uncertainty based on the frequency with which keywords appear in major news articles can explain the emergence of bubbles.

Using daily observations for a recent five-year period and six major cryptocurrencies, we show that recently developed cryptocurrency uncertainty price and policy indices are a better predictor for bubbles than factors previously considered in the literature, including VIX, economic policy uncertainty, the prices of commodities such as Brent oil and gold, or the intensity of search for keywords related to cryptocurrencies retrieved from Google Trends. Furthermore, using a random forest model, we show that the cryptocurrency uncertainty indices are the features with the highest Gini importance scores relative to the factors previously established in the literature. These results are robust to various alternative specifications of bubbles, data sample and time periods, including the COVID-19 pandemic. While we show that the frequency of bubbles intensifies with the onset of the pandemic, we also show that the two measures of uncertainty are robust predictors of bubbles prior to and during the pandemic.

Acknowledgement

The paper benefitted from extensive discussions and from the comments of participants in the session on uncertainty and volatility of to the Crypto Research Conference 2022 in Durham University. We thank the Cryptocurrency Research Conference 2022 participants at Durham University Business School, Durham, UK, for their valuable comments. Also, special thanks to Matt Brown for his detailed comments and suggestions.

References

Aloosh, A., Ouzan, S., & Shahzad, S. J. H. (2022). Bubbles across meme stocks and cryptocurrencies, *Finance Research Letters*, 49, 103155.

Altman, N., & Krzywinski, M. (2017). Ensemble methods: bagging and random forests. *Nature Methods*, *14*(10), 933-935.

Amini, S., Elmore, R., Öztekin, Ö., & Strauss, J. (2021). Can machines learn capital structure dynamics?. *Journal of Corporate Finance*, *70*, 102073.

Baek, C., & Elbeck, M. (2015). Bitcoins as an investment or speculative vehicle? A first look. *Applied Economics Letters*, *22*(1), 30-34.

Baker, S.R., Bloom, N., & Davis, S.J. (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics*, 131(4), 1593-1636

Balcilar, M., Bouri, E., Gupta, R., & Roubaud, D. (2017). Can volume predict Bitcoin returns and volatility? A quantiles-based approach. *Economic Modelling*, *64*, 74-81.

Biais, B., Bisière, C., Bouvard, M., Casamatta, C., & Menkveld, A. J. (2021). Equilibrium Bitcoin pricing. forthcoming *Journal of Finance*. Available at SSRN: https://ssrn.com/abstract=3261063

Bouri, E., Das, M., Gupta, R., & Roubaud, D. (2018). Spillovers between Bitcoin and other assets during bear and bull markets. *Applied Economics*, *50*(55), 5935-5949.

Bouri, E., Shahzad, S. J. H., & Roubaud, D. (2019). Co-explosivity in the cryptocurrency market. *Finance Research Letters*, *29*, 178-183.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Cheah, E. T., & Fry, J. (2015). Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. *Economics letters*, *130*, 32-36.

Cheung, A., Roca, E., & Su, J. J. (2015). Crypto-currency bubbles: an application of the Phillips–Shi–Yu (2013) methodology on Mt. Gox bitcoin prices. *Applied Economics*, *47*(23), 2348-2358.

Chowdhury, M. S. R., Damianov, D. S., & Elsayed, A. H. (2022). Bubbles and crashes in cryptocurrencies: Interdependence, contagion, or asset rotation? *Finance Research Letters*, *46*, 102494.

Cong, L. W., Li, Y., & Wang, N. (2021). Tokenomics: Dynamic adoption and valuation. *Review of Financial Studies*, *34*, 1105-1155.

Conrad, C., Custovic, A., & Ghysels, E. (2018). Long-and short-term cryptocurrency volatility components: A GARCH-MIDAS analysis. *Journal of Risk and Financial Management*, *11*(2), 23.

Corbet, S., Katsiampa, P., & Lau, C. K. M. (2020). Measuring quantile dependence and testing directional predictability between Bitcoin, altcoins, and traditional financial assets. *International Review of Financial Analysis*, 71, 101571.

Corbet, S., Lucey, B., & Yarovaya, L. (2018). Datestamping the Bitcoin and Ethereum bubbles. *Finance Research Letters*, *26*, 81-88.

Corbet, S., Larkin, C., Lucey, B., Meegan, A., & Yarovaya, L. (2020a). Cryptocurrency reaction to fomc announcements: Evidence of heterogeneity based on blockchain stack position. *Journal of Financial Stability*, *46*, 100706.

Corbet, S., Larkin, C., Lucey, B. M., Meegan, A., & Yarovaya, L. (2020b). The impact of macroeconomic news on Bitcoin returns. *European Journal of Finance*, *26*(14), 1396-1416.

Demir, E., Gozgor, G., Lau, C. K. M., & Vigne, S. A. (2018). Does economic policy uncertainty predict the Bitcoin returns? An empirical investigation. *Finance Research Letters*, *26*, 145-149.

Elsayed, A. H., Gozgor, G., & Yarovaya, L. (2022). Volatility and return connectedness of cryptocurrency, gold, and uncertainty: Evidence from the cryptocurrency uncertainty indices. Finance Research Letters, 47, 102732.

Enoksen, F. A., Landsnes, C. J., Lučivjanská, K., & Molnár, P. (2020). Understanding risk of bubbles in cryptocurrencies. *Journal of Economic Behavior & Organization*, *176*, 129-144.

Etienne, X. L., Irwin, S. H., & Garcia, P. (2014). Bubbles in food commodity markets: Four decades of evidence. *Journal of International Money and Finance*, *42*, 129-155.

Fang, L., Bouri, E., Gupta, R., & Roubaud, D. (2019). Does global economic uncertainty matter for the volatility and hedging effectiveness of Bitcoin?. *International Review of Financial Analysis*, *61*, 29-36.

Fang, T., Su, Z., & Yin, L. (2020). Economic fundamentals or investor perceptions? The role of uncertainty in predicting long-term cryptocurrency volatility. *International Review of Financial Analysis*, *71*, 101566.

Fry, J., & Cheah, E. T. (2016). Negative bubbles and shocks in cryptocurrency markets. *International Review of Financial Analysis*, *47*, 343-352.

Gozgor, G., Tiwari, A.K., Demir, E. & Akron, S. (2019). The Relationship between Bitcoin Returns and Trade Policy Uncertainty. *Finance Research Letters*, 29, 75–82. (SSCI).

Gkillas, K., Katsiampa, P., Konstantatos, C., & Tsagkanos, A. (2022). Discontinuous movements and asymmetries in cryptocurrency markets. *European Journal of Finance*, DOI: 10.1080/1351847X.2021.2015416

Grinberg, R. (2012). Bitcoin: An innovative alternative digital currency. Hastings Sci. & Tech. LJ, 4, 159.

Hastie, T., Tibshirani, R., & Friedman, J. (2008). The Elements of Statistical Learning: Data Mining, Inference and Prediction. Second Edition. *Springer Series in Statistics*.

Huang, Y., & Luk, P. (2020). Measuring economic policy uncertainty in China. *China Economic Review*, *59*, 101367.

Irresberger, F., John, K., Mueller, P., & Saleh, F. (2021). The public blockchain ecosystem: An empirical analysis. *NYU Stern School of Business*.

Jones, S. (2017). Corporate bankruptcy prediction: a high dimensional analysis. *Review of Accounting Studies*, *22*(3), 1366-1422.

Kristoufek, L. (2015). What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. *PloS one*, *10*(4), e0123923.

Liu, Y., & Tsyvinski, A. (2021). Risk and return of cryptocurrency. *Review of Financial Studies*, 34, 2689-2727.

Liu, Y., & Wu, H. (2017). Water bloom warning model based on random forest. *Journal of Information and Communication Engineering (JICE), 3(5)*, 178-182.

Liu, R., Wan, S., Zhang, Z., & Zhao, X. (2020). Is the introduction of futures responsible for the crash of Bitcoin?. *Finance Research Letters*, 34, 101259.

Long, H., Demir, E., Będowska-Sójka, B., Zaremba, A., & Shahzad, S. J. H. (2022). Is geopolitical risk priced in the cross-section of cryptocurrency returns? *Finance Research Letters*, 49, 103131.

Lucey, B. M., Vigne, S. A., Yarovaya, L., & Wang, Y. (2022). The cryptocurrency uncertainty index. *Finance Research Letters*, 45, 102147.

Manela, A., & Moreira, A. (2017). News implied volatility and disaster concerns. *Journal of Financial Economics*, *123*(1), 137-162.

Maouchi, Y., Charfeddine, L., & El Montasser, G. (2022). Understanding digital bubbles amidst the COVID-19 pandemic: Evidence from DeFi and NFTs. *Finance Research Letters*, 47, 102584.

Mullainathan, D.J., & Spiess, J. (2017). Machine learning: An applied econometric approach, *Journal of Economic Perspectives*, 31, 87-106.

Pagnotta, E. S., & Buraschi, A. (2018). An equilibrium valuation of Bitcoin and decentralized network assets, Available at SSRN: https://ssrn.com/abstract=3142022

Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients, *Criminology*, 36(4), 859-866,

Phillips, P. C., Shi, S., & Yu, J. (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review*, 56(4), 1043-1078.

Shi, S. (2017). Speculative bubbles or market fundamentals? An investigation of US regional housing markets. *Economic Modelling*, 66, 101-111.

Stiglitz, J. (1990). Symposium on Bubbles. *Journal of Economic Perspectives*, 4(2), 13-18.

Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC bioinformatics*, *9*(1), 1-11.

Trimborn, S., & Härdle, W. K. (2018). CRIX an Index for cryptocurrencies. *Journal of Empirical Finance*, *49*, 107-122.

Walther, T., Klein, T., & Bouri, E. (2019). Exogenous drivers of Bitcoin and Cryptocurrency volatility–A mixed data sampling approach to forecasting. *Journal of International Financial Markets, Institutions and Money*, 63, 101133.

Table 1

Descriptive statistics for currency-specific variables.

	BTC		DOGE		ETH		LTC		XLM		XRP	
	Mean	St. Dev	Mean	St. Dev	Mean	St. Dev	Mean	St. Dev	Mean	St. Dev	Mean	St. Dev
Bubble	0.304	0.460	0.157	0.364	0.264	0.441	0.138	0.345	0.064	0.245	0.085	0.279
BSADF	0.035	1.040	-0.447	1.344	-0.016	1.104	-0.339	1.082	-0.587	0.826	-0.639	1.046
Return	0.232	4.129	4.129	8.076	0.290	5.745	0.202	5.961	0.306	7.8189	0.285	7.439
Volume	18.400	22.500	0.640	3.580	8.330	11.000	2.190	2.570	0.360	0.740	2.050	3.720

Notes: 'Bubble' takes on the value of 1 in bubble periods and 0 otherwise. 'BSADF' is the backward sup ADF statistics defined in equation (3). 'Return' is the annualized return over the sample period from June 1, 2016 to May 31, 2021. Volume is measured in $10^{\wedge 9}$ USD.

Table 2

Descriptive statistics for key variable of interests and macroeconomic factors.

	Mean	St. Dev.	Min	Max
UCRY Price	100.415	1.367	99.244	109.178
UCRY Policy	100.379	1.321	99.221	108.258
Volume	5.31	22.50	0.04	351.00
VIX	18.123	9.104	9.140	76.450
Gold	1444.744	241.873	1129.8	2058.4
Brent oil	53.323	11.081	8.910	77.41
TED	0.314	0.185	0.110	1.420
EPU	137.228	111.277	4.7	807.66
G-Trend	7.757	10.727	0	100

Notes: 'UCRY Price' and 'UCRY Policy' are the price and policy uncertainty indices created by Lucey et al. (2022). 'Volume' is the trading volume of the studied six cryptocurrencies measured in $10^{\wedge 9}$ USD. 'VIX' is the 30-day expected volatility of the U.S. stock market. 'Gold' is the price of a troy ounce of gold. 'Brent oil' is the price of a barrel of Brent crude oil. 'TED' is the TED spread. 'EPU' is the economic policy uncertainty index. 'G-Trend' is the search volume for the studied six cryptocurrencies derived from Google Trends.

Table 3

Pearson correlations among variables.

	Bubble	UCRY Price	UCRY Policy	Volume	VIX	Gold	Brent oil	TED	EPU	G-Trend
Bubble	1.000									
UCRY Price	**0.250*****	1.000								
UCRY Policy	**0.258*****	0.986***	1.000							
Volume	**0.178*****	0.389*	0.377*	1.000						
VIX	-0.209***	0.239***	0.232***	0.357***	1.000					
Gold	-0.038***	0.474***	0.465***	0.505***	0.667***	1.000				
Brent oil	0.048***	0.161***	0.143***	-0.035***	-0.571***	-0.399***	1.000			
TED	-0.109***	-0.368***	-0.350***	-0.413***	-0.133***	-0.567***	-0.151***	1.000		
EPU	-0.072***	0.128***	0.144***	0.255***	0.635***	0.610***	-0.614***	-0.182***	1.000	
G-Trend	0.241***	0.356***	0.353***	0.285***	-0.002	0.113***	0.126***	-0.148***	-0.014	1.000

Notes. ‘UCRY Price’ and ‘UCRY Policy’ are the price and policy uncertainty indices created by Lucey et al. (2022). ‘Volume’ is the trading volume of the studied six cryptocurrencies measured in $10^{\wedge 9}$ USD. ‘VIX’ is the 30-day expected volatility of the U.S. stock market. ‘Gold’ is the price of a troy ounce of gold. ‘Brent oil’ is the price of a barrel of Brent crude oil. ‘TED’ is the TED spread. ‘EPU’ is the economic policy uncertainty index. ‘G-Trend’ is the search volume for the studied six cryptocurrencies derived from Google Trends. *** significance is at the 1% level; ** significance is at the 5% level; and * significance is at the 10% level.

Table 4

Baseline probit regression results for UCRY Price.

Dependent Variable	***BUBBLE***$_{i,t}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
UCRY_Price$_{i,t-1}$	**5.323*****	**4.114*****	**5.072*****	**5.695*****	**6.411*****	**6.054*****	**6.183*****	**5.509*****
	(0.327)	(0.329)	(0.311)	(0.361)	(0.409)	(0.404)	(0.402)	(0.419)
Volume$_{i,t-1}$		0.014***	0.025***	0.026***	0.028***	0.026***	0.026***	0.024***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
VIX$_{i,t-1}$			-0.379***	-0.335***	-0.355***	-0.336***	-0.353***	-0.341***
			(0.012)	(0.016)	(0.016)	(0.016)	(0.016)	(0.017)
Gold$_{i,t-1}$				-0.189***	-0.255***	-0.373***	-0.436***	-0.422***
				(0.046)	(0.050)	(0.050)	(0.050)	(0.050)
Brent oil $_{i,t-1}$					-0.106***	-0.116***	-0.109***	-0.108***
					(0.021)	(0.021)	(0.020)	(0.020)
TED$_{i,t-1}$						-0.076***	-0.075***	-0.075***
						(0.009)	(0.009)	(0.009)
EPU$_{i,t-1}$							0.039***	0.039***
							(0.007)	(0.007)
G-Trends$_{i,t-1}$								0.027***
								(0.003)
Observations	6,924	6,924	6,924	6,924	6,924	6,924	6,924	6,924
McFadden's R-Squared	0.055	0.066	0.212	0.215	0.234	0.240	0.244	0.251

Notes. The table reports the average marginal effect with corresponding robust standard errors in brackets. All variables are log transformed. The sample spans the period from June 1, 2016 to May 31, 2021. The text in bold indicates the main variable of interest. *** significance is at the 1% level; ** significance is at the 5% level; and * significance is at the 10% level.

Table 5

Baseline probit regression results for UCRY Policy.

Dependent Variable				***BUBBLE***$_{i,t}$				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
UCRY_Policy$_{i,t-1}$	**5.701*****	**4.501*****	**5.400*****	**6.089*****	**6.782*****	**6.423*****	**6.519*****	**5.853*****
	(0.333)	(0.336)	(0.312)	(0.361)	(0.405)	(0.402)	(0.399)	(0.416)
Volume$_{i,t-1}$		0.013***	0.024***	0.026***	0.028***	0.026***	0.026***	0.023***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
VIX$_{i,t-1}$			-0.378***	-0.331***	-0.351***	-0.331***	-0.346***	-0.335***
			(0.012)	(0.016)	(0.016)	(0.016)	(0.016)	(0.017)
Gold$_{i,t-1}$				-0.202***	-0.263***	-0.386***	-0.443***	-0.430***
				(0.046)	(0.049)	(0.049)	(0.050)	(0.050)
Brent oil $_{i,t-1}$					-0.104***	-0.115***	-0.108***	-0.108***
					(0.021)	(0.021)	(0.020)	(0.020)
TED$_{i,t-1}$						-0.078***	-0.077***	-0.077***
						(0.009)	(0.009)	(0.009)
EPU$_{i,t-1}$							0.035***	0.036***
							(0.007)	(0.007)
G-Trends$_{i,t-1}$								0.026***
								(0.003)
Observations	6,924	6,924	6,924	6,924	6,924	6,924	6,924	6,924
McFadden's R-Squared	0.059	0.069	0.216	0.219	0.238	0.244	0.247	0.254

Notes. The table reports the average marginal effect with corresponding robust standard errors in brackets. All variables are log transformed. The sample spans the period from June 1, 2016 to May 31, 2021. The text in bold indicates the main variable of interest. *** significance is at the 1% level; ** significance is at the 5% level; and * significance is at the 10% level.

Table 6

Random forest model: relative variable importance for the entire sample.

Variables	Score	Rank
UCRY Policy	**100**	**1**
UCRY Price	**71**	**2**
Brent oil	32	3
Gold	32	4
VIX	30	5
TED	29	6
EPU	12	7
G-Trends	3	8
Volume	3	9

Notes: The dependent variable is the S&P Cryptocurrency Broad Digital Market (BDM) index. Results are based on training 500 decision trees using a bagging procedure with a random selection of 3 out of the 9 features.

Table 7

Random forest model: relative variable importance before and after the COVID-19 pandemic.

Variables	Score	Rank
Panel A: Before COVID-19		
UCRY Policy	**100**	**1**
UCRY Price	**72**	**2**
VIX	65	3
Brent oil	63	4
TED	55	5
Gold	50	6
EPU	23	7
Volume	7	8
G-Trend	6	9
Panel B: After COVID-19		
UCRY Price	**100**	**1**
UCRY Policy	**98**	**2**
Brent oil	47	3
VIX	33	4
Gold	18	5
TED	16	6
EPU	13	7
G-Trend	2	8
Volume	1	9

Notes: The dependent variable is the S&P Cryptocurrency Broad Digital Market (BDM) index. Results are based on training 500 decision trees using a bagging procedure with a random selection of 3 out of the 9 features.

Table 8

Random forest model: out-of-sample forecasting performance.

	UCRY_Price Model		***UCRY_Policy Model***	
	PSY Framework: normal versus bubble days			
	Normal	Bubble	Normal	Bubble
	(1)	(2)	(3)	(4)
Model prediction: normal versus bubble days				
Normal	1686	27	1684	29
Bubble	74	273	75	272
Classification Accuracy	**95.09%**		**94.95%**	

Notes: The dependent variable is the S&P Cryptocurrency Broad Digital Market (BDM) index. Results are based on training 500 decision trees using a bagging procedure with a random selection of 3 out of the 9 features.

Table 9

Probit regression results: the role of protocols.

Dependent Variable	$BUBBLE_{i,t}$			
	Other protocols	Nonstandard protocols	Other protocols	Nonstandard protocols
	(1)	(2)	(3)	(4)
UCRY_Price$_{i,t-1}$	**3.553*** **	**0.845*** **		
	(0.663)	(0.256)		
UCRY_Policy$_{i,t-1}$			**3.905*** **	**0.997*** **
			(0.649)	(0.259)
Other controls	Yes	Yes	Yes	Yes
Observations	4,616	2,308	4,616	2,308
McFadden's R-Squared	0.360	0.262	0.360	0.262
Coefficient equality test	$\beta_1 = \beta_2$ *p-value*=0.000		$\beta_3 = \beta_4$ *p-value*=0.000	

Notes. Ripple and Stellar use nonstandard protocols. Bitcoin, Litecoin, and Dogecoin use proof of work, and Ethereum adopted a proof of stake protocol at the end of 2020. These cryptocurrencies form the group of currencies that use other protocols. The table reports the average marginal effect with corresponding robust standard errors in brackets. The other controls are $Volume_{i,t-1}$; $VIX_{i,t-1}$; $Gold_{i,t-1}$; Brent $oil_{i,t-1}$; $TED_{i,t-1}$; $EPU_{i,t-1}$; and G-$Trends_{i,t-1}$. All variables are log transformed. Our sample spans the period from June 1, 2016 to May 31, 2021. The text in bold indicates the main variable of interest. *** significance is at the 1% level; ** significance is at the 5% level; and * significance is at the 10% level.

Table 10

Probit regression results: the role of COVID-19.

Dependent Variable	***BUBBLE***$_{i,t}$	
	(1)	(2)
COVID-19	**0.230*****	**0.219*****
	(0.0187)	(0.0185)
UCRY_Price$_{i,t-1}$	**4.570*****	
	(0.402)	
UCRY_Policy$_{i,t-1}$		**4.862*****
		(0.404)
Other controls	Yes	Yes
Observations	6,924	6,924
McFadden's R-Squared	0.258	0.261

Notes. The table reports the average marginal effect with corresponding robust standard errors in brackets. The other controls are $Volume_{i,t-1}$; $VIX_{i,t-1}$; $Gold_{i,t-1}$; Brent oil $_{i,t-1}$; $TED_{i,t-1}$; $EPU_{i,t-1}$; and G-Trends$_{i,t-1}$. All variables are log transformed. Our sample spans the period from June 1, 2016 to May 31, 2021. The text in bold indicates the main variable of interest. *** significance is at the 1% level; ** significance is at the 5% level; and * significance is at the 10% level.

Table 11

Robustness test: alternative measure of a bubble (return at upper quantile).

Dependent Variable	***$Ret_{i,t}$ at Upper Quantile τ = 0.95***			
	(1)	(2)	(3)	(4)
UCRY_Price $_{i,t-1}$	**1.683*** **	**3.371*** **		
	(0.390)	(0.457)		
UCRY_Policy $_{i,t-1}$			**1.854*** **	**3.118*** **
			(0.394)	(0.454)
Intercept	-7.657***	-13.81***	-10.06***	-12.91***
	(1.795)	(2.035)	(1.373)	(1.330)
Other controls	No	Yes	No	Yes
Observations	6,924	6,924	6,924	6,924
Pseudo R-Squared	0.018	0.059	0.020	0.060

Notes. The table reports the coefficients with corresponding bootstrapped standard errors with 1,000 replications in brackets. The other controls are $Volume_{i,t-1}$; $VIX_{i,t-1}$; $Gold_{i,t-1}$; Brent oil $_{i,t-1}$; $TED_{i,t-1}$; $EPU_{i,t-1}$; and G-$Trends_{i,t-1}$. All variables are log transformed. The sample period is from June 1, 2016 to May 31, 2021. The text in bold indicates the main variable of interest. *** significance is at the 1% level; ** significance is at the 5% level; and * significance is at the 10% level.

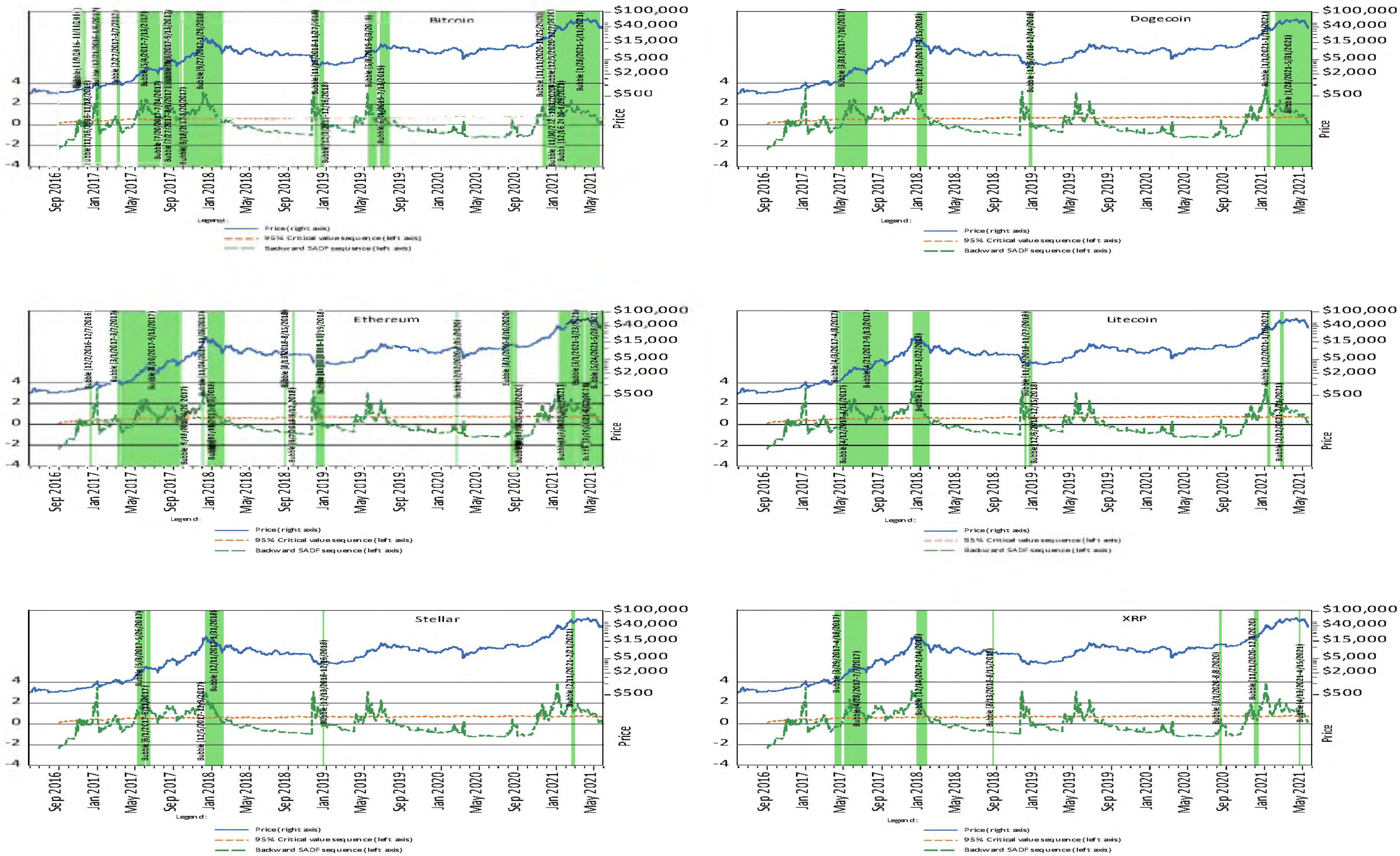
Table 12

Robustness test: probit regressions results using bubbles derived from the S&P broad cryptocurrency index.

Dependent Variable	***S&P_BUBBLE***$_{i,t}$			
	(1)	(2)	(3)	(4)
UCRY_Price$_{i,t-1}$	**13.74*** **	**10.20*** **		
	(0.858)	(0.872)		
UCRY_Policy$_{i,t-1}$			**15.32*** **	**12.07*** **
			(0.891)	(0.920)
Other Controls	No	Yes	No	Yes
Observations	5,832	5,832	5,832	5,832
Pseudo R-Squared	0.240	0.339	0.260	0.414

Notes. The table reports the average marginal effect with corresponding robust standard errors in brackets. The other controls are $\text{Volume}_{i,t-1}$; $\text{VIX}_{i,t-1}$; $\text{Gold}_{i,t-1}$; Brent oil $_{i,t-1}$; $\text{TED}_{i,t-1}$; $\text{EPU}_{i,t-1}$; and $\text{G-Trends}_{i,t-1}$. All variables are log transformed. The sample period is from June 6, 2017 to May 31, 2021. Data is based on data availability. The text in bold indicates the main variable of interest. *** significance is at the 1% level; ** significance is at the 5% level; and * significance is at the 10% level.

Fig. 1. Date stamping of bubble periods for cryptocurrencies, 2016m6–2021m5.



Notes. Bubble periods (shaded in green) are periods for which the BSADF statistics exceed the 95% critical value for at least 3 days.

Appendix A. Additional Tables

Table A.1

Probit regression results for UCRY Price (using 7 days lag).

Dependent Variable	*BUBBLE*$_{i,t}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
UCRY_Price$_{i,t-7}$	**5.230*****	**4.209*****	**5.315*****	**5.740*****	**6.476*****	**6.083*****	**6.229*****	**5.698*****
	(0.336)	(0.342)	(0.326)	(0.371)	(0.420)	(0.412)	(0.409)	(0.427)
Volume$_{i,t-7}$		0.011***	0.022***	0.024***	0.026***	0.023***	0.024***	0.021***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
VIX$_{i,t-7}$			-0.380***	-0.350***	-0.371***	-0.350***	-0.368***	-0.359***
			(0.012)	(0.016)	(0.016)	(0.016)	(0.017)	(0.017)
Gold$_{i,t-7}$				-0.132***	-0.197***	-0.338***	-0.416***	-0.406***
				(0.047)	(0.050)	(0.050)	(0.0503)	(0.0505)
Brent oil $_{i,t-7}$					-0.110***	-0.121***	-0.113***	-0.113***
					(0.022)	(0.022)	(0.021)	(0.0213)
TED$_{i,t-7}$						-0.089***	-0.088***	-0.088***
						(0.009)	(0.009)	(0.009)
EPU$_{i,t-7}$							0.045***	0.046***
							(0.008)	(0.008)
G-Trends$_{i,t-7}$								0.021***
								(0.003)
Observations	6,900	6,900	6,900	6,900	6,900	6,900	6,900	6,900
McFadden's R-Squared	0.050	0.058	0.201	0.202	0.222	0.230	0.234	0.239

Notes. The table reports the average marginal effect with corresponding robust standard errors in brackets. All variables are log transformed. The sample period is from June 1, 2016 to May 31, 2021. The text in bold indicates the main variable of interest. *** significance is at the 1% level; ** significance is at the 5% level; and * significance is at the 10% level.

Table A.2

Probit regression results for UCRY Policy (using 7 days lag).

Dependent Variable	***BUBBLE***$_{i,t}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
UCRY_Policy$_{i,t-7}$	**5.627******	**4.647******	**5.636******	**6.156******	**6.868******	**6.478******	**6.591******	**6.085******
	(0.339)	(0.346)	(0.323)	(0.366)	(0.411)	(0.407)	(0.402)	(0.420)
Volume$_{i,t-7}$		0.011***	0.022***	0.024***	0.026***	0.023***	0.023***	0.021***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
VIX$_{i,t-7}$			-0.379***	-0.343***	-0.364***	-0.342***	-0.360***	-0.351***
			(0.012)	(0.016)	(0.016)	(0.016)	(0.017)	(0.017)
Gold$_{i,t-7}$				-0.156***	-0.219***	-0.362***	-0.435***	-0.426***
				(0.046)	(0.050)	(0.050)	(0.050)	(0.050)
Brent oil $_{i,t-7}$					-0.109***	-0.121***	-0.113***	-0.112***
					(0.022)	(0.022)	(0.021)	(0.021)
TED$_{i,t-7}$						-0.089***	-0.089***	-0.088***
						(0.009)	(0.009)	(0.009)
EPU$_{i,t-7}$							0.043***	0.0441***
							(0.008)	(0.008)
G-Trends$_{i,t-7}$								0.020***
								(0.003)
Observations	6,900	6,900	6,900	6,900	6,900	6,900	6,900	6,900
McFadden's R-Squared	0.055	0.062	0.206	0.208	0.228	0.236	0.240	0.244

Notes. The table reports the average marginal effect with corresponding robust standard errors in brackets. All variables are log transformed. The sample period is from June 1, 2016 to May 31, 2021. The text in bold indicates the main variable of interest. *** significance is at the 1% level; ** significance is at the 5% level; and * significance is at the 10% level.

Table A.3

Linear regression results for UCRY Price.

Dependent Variable	*BUBBLE*$_{i,t}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
UCRY_Price$_{i,t-1}$	**7.016*** **	**5.988*** **	**7.097*** **	**7.684*** **	**9.107*** **	**9.136*** **	**9.265*** **	**8.295*** **
	(0.462)	(0.474)	(0.475)	(0.480)	(0.588)	(0.602)	(0.579)	(0.585)
Volume$_{i,t-1}$		0.011***	0.024***	0.026***	0.028***	0.027***	0.027***	0.024***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
VIX$_{i,t-1}$			-0.322***	-0.283***	-0.356***	-0.332***	-0.353***	-0.343***
			(0.010)	(0.012)	(0.019)	(0.017)	(0.016)	(0.0164)
Gold$_{i,t-1}$				-0.185***	-0.282***	-0.428***	-0.496***	-0.472***
				(0.032)	(0.041)	(0.063)	(0.054)	(0.054)
Brent oil $_{i,t-1}$					-0.195***	-0.215***	-0.200***	-0.198***
					(0.048)	(0.054)	(0.054)	(0.053)
TED$_{i,t-1}$						-0.058***	-0.058***	-0.059***
						(0.010)	(0.010)	(0.010)
EPU$_{i,t-1}$							0.041***	0.041***
							(0.009)	(0.009)
G-Trends$_{i,t-1}$								0.035***
								(0.003)
Intercept	-32.17***	-27.66***	-32.12***	-33.63***	-38.43***	-37.52***	-37.84***	-33.57***
	(2.130)	(2.176)	(2.182)	(2.184)	(2.457)	(2.436)	(2.391)	(2.415)
Observations	6,924	6,924	6,924	6,924	6,924	6,924	6,924	6,924
R-Squared	0.063	0.070	0.169	0.172	0.198	0.201	0.204	0.213

Notes. The table reports the coefficients with corresponding robust standard errors in brackets. All variables are log transformed. The sample period is from June 1, 2016 to May 31, 2021. The text in bold indicates the main variable of interest. *** significance is at the 1% level; ** significance is at the 5% level; and * significance is at the 10% level.

Table A.4

Linear regression results for UCRY Policy.

Dependent Variable	$BUBBLE_{i,t}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$UCRY\_Policy_{i,t-1}$	**7.463*** **	**6.447*** **	**7.540*** **	**8.160*** **	**9.435*** **	**9.495*** **	**9.572*** **	**8.604*** **
	(0.469)	(0.477)	(0.472)	(0.474)	(0.558)	(0.574)	(0.555)	(0.560)
$Volume_{i,t-1}$		0.011***	0.024***	0.026***	0.029***	0.027***	0.027***	0.024***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$VIX_{i,t-1}$			-0.322***	-0.282***	-0.352***	-0.327***	-0.345***	-0.336***
			(0.010)	(0.012)	(0.018)	(0.017)	(0.016)	(0.016)
$Gold_{i,t-1}$				-0.193***	-0.281***	-0.438***	-0.496***	-0.473***
				(0.031)	(0.040)	(0.061)	(0.053)	(0.053)
$Brent\ oil_{i,t-1}$					-0.188***	-0.210***	-0.19***	-0.195***
					(0.046)	(0.053)	(0.052)	(0.052)
$TED_{i,t-1}$						-0.062***	-0.062***	-0.062***
						(0.010)	(0.010)	(0.010)
$EPU_{i,t-1}$							0.036***	0.036***
							(0.009)	(0.009)
$G\text{-}Trends_{i,t-1}$								0.035***
								(0.003)
Intercepts	-34.23***	-29.77***	-34.16***	-35.77***	-39.99***	-39.15***	-39.27***	-35.00***
	(2.160)	(2.193)	(2.171)	(2.163)	(2.355)	(2.341)	(2.309)	(2.328)
Observations	6,924	6,924	6,924	6,924	6,924	6,924	6,924	6,924
R-Squared	0.067	0.074	0.173	0.176	0.201	0.204	0.206	0.215

Notes. The table reports the coefficients with corresponding robust standard errors in brackets. All variables are log transformed. from the sample period is June 1, 2016 to May 31, 2021. The text in bold indicates the main variable of interest. *** significance is at the 1% level; ** significance is at the 5% level; and * significance is at the 10% level.

Table A.5.

Linear regression results for UCRY Price: an alternative measure of a bubble (BSADF).

Dependent Variable	$BSADF_{i,t}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
UCRY_Price$_{i,t-1}$	**25.95*****	**17.99*****	**21.97*****	**26.62*****	**28.75*****	**28.91*****	**29.11*****	**26.22*****
	(1.181)	(1.129)	(1.141)	(1.183)	(1.317)	(1.366)	(1.334)	(1.367)
Volume$_{i,t-1}$		0.087***	0.134***	0.150***	0.154***	0.147***	0.147***	0.138***
		(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
VIX$_{i,t-1}$			-1.155***	-0.848***	-0.957***	-0.833***	-0.866***	-0.835***
			(0.029)	(0.032)	(0.042)	(0.042)	(0.042)	(0.043)
Gold$_{i,t-1}$				-1.467***	-1.612***	-2.377***	-2.484***	-2.414***
				(0.084)	(0.095)	(0.142)	(0.134)	(0.133)
Brent oil $_{i,t-1}$					-0.292***	-0.399***	-0.374***	-0.370***
					(0.080)	(0.103)	(0.102)	(0.101)
TED$_{i,t-1}$						-0.304***	-0.305***	-0.306***
						(0.027)	(0.026)	(0.026)
EPU$_{i,t-1}$							0.066**	0.066**
							(0.026)	(0.026)
G-Trends$_{i,t-1}$								0.106***
								(0.011)
Intercept	-120.0***	-85.01***	-101.0***	-113.0***	-120.2***	-115.4***	-115.9***	-103.2***
	(5.443)	(5.182)	(5.241)	(5.345)	(5.690)	(5.735)	(5.674)	(5.816)
Observations	6,924	6,924	6,924	6,924	6,924	6,924	6,924	6,924
R-Squared	0.096	0.146	0.289	0.307	0.314	0.323	0.324	0.333

Notes. The dependent variable is the *BSADF/PSY* statistic. The table reports the coefficients with corresponding robust standard errors in brackets. All variables are log transformed. The sample period is from June 1, 2016 to May 31, 2021. The text in bold indicates the main variable of interest. *** significance is at the 1% level; ** significance is at the 5% level; and * significance is at the 10% level.

Table A.6.

Linear regression results for UCRY Policy: an alternative measure of a bubble (BSADF).

Dependent Variable	$\boldsymbol{BSADF_{i,t}}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\boldsymbol{UCRY\_Policy_{i,t-1}}$	**26.69*****	**18.66*****	**22.57*****	**27.26*****	**29.05*****	**29.36*****	**29.46*****	**26.52*****
	(1.154)	(1.106)	(1.091)	(1.119)	(1.207)	(1.257)	(1.234)	(1.265)
$Volume_{i,t-1}$		0.088***	0.135***	0.151***	0.155***	0.147***	0.148***	0.139***
		(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
$VIX_{i,t-1}$			-1.152***	-0.847***	-0.946***	-0.816***	-0.841***	-0.812***
			(0.028)	(0.032)	(0.0418)	(0.042)	(0.042)	(0.042)
$Gold_{i,t-1}$				-1.454***	-1.578***	-2.379***	-2.457***	-2.389***
				(0.084)	(0.092)	(0.137)	(0.131)	(0.130)
$Brent\ oil_{i,t-1}$					-0.264***	-0.377***	-0.357***	-0.355***
					(0.074)	(0.096)	(0.097)	(0.096)
$TED_{i,t-1}$						-0.317***	-0.317***	-0.316***
						(0.027)	(0.026)	(0.026)
$EPU_{i,t-1}$							0.0491*	0.051**
							(0.026)	(0.026)
$G\text{-}Trends_{i,t-1}$								0.108***
								(0.011)
Intercept	-123.4***	-88.12***	-103.8***	-116.0***	-122.0***	-117.7***	-117.8***	-104.9***
	(5.321)	(5.079)	(5.015)	(5.067)	(5.264)	(5.314)	(5.275)	(5.411)
Observations	6,924	6,924	6,924	6,924	6,924	6,924	6,924	6,924
R-Squared	0.095	0.147	0.289	0.307	0.313	0.323	0.323	0.333

Notes. The dependent variable is the *BSADF/PSY* statistic. The table reports the coefficients with corresponding robust standard errors in brackets. All variables are log transformed. The sample period is from June 1, 2016 to May 31, 2021. The text in bold indicates the main variable of interest. *** significance is at the 1% level; ** significance is at the 5% level; and * significance is at the 10% level.