Research in International Business and Finance 64 (2023) 101905

Contents lists available at ScienceDirect

Research in International Business and Finance

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

Forecasting cryptocurrency returns with machine learning☆

Yujun Liu [a], Zhongfei Li [b,*], Ramzi Nekhili [c], Jahangir Sultan [d]

[a] School of Business, Sun Yat-Sen University, Guangzhou, China
[b] School of Business, Southern University of Science and Technology, Shenzhen, China
[c] Department of Accounting and Finance, Applied Science University, Bahrain
[d] McCallum Graduate School, Bentley University, MA, USA

ARTICLE INFO

JEL classification in this article:
C53
G11
G12
G15
G17

Keywords:
Cryptocurrency
Machine learning
eXtreme Gradient Boostine
SHapley Additive exPlanations

ABSTRACT

This article employs machine learning models to predict returns for 3703 cryptocurrencies for the 2013 – 2021 period. Based on daily data, we build an equal (capital)-weighted portfolio that generates 7.1 % (2.4 %) daily return with a 1.95 (0.27) Sharpe ratio. We obtain an out-of-sample R^2 of 4.855 %. Our results suggest that cryptocurrencies behave like conventional assets than fiat currencies since variables, including lagged returns, can predict future returns. As assets, cryptocurrencies are not weakly efficient, and production costs do not determine their prices. Returns for small cryptocurrencies are more predictable than larger ones. The predictive power of the 1-day lagged return is stronger than all other features (predictors) combined. The results offer new insights for crypto investors, traders, and financial analysts.

1. Introduction

Cryptocurrencies are one of the most volatile assets. The popular press and the academic community routinely highlight this market aspect. For instance, the market for all cryptocurrencies was about $500 billion in November 2020 (Fortune). It peaked at $2.84 trillion on November 9, 2021, suggesting a 500 % growth in 12 months. The current market cap is $856 billion (December 2, 2022), falling 70 % from the peak. Bitcoin, the flagship among cryptocurrencies, has an unusual volatility pattern. For example, between April 14, 2021 and May 29, 2021, the currency dropped 50 % from its peak price of $64,829. In fact, on Wednesday, May 19, 2021, Bitcoin dropped by more than 30% (CNBC, May 19, 2021). In early May 2022, Bitcoin fell 19 % in 10 days. This impressive volatility is a cause for concern for investors and hedgers. On the other hand, high volatility might present profitable trading opportunities for others.

The media has always been infatuated with cryptocurrencies' good, bad, and ugly. Recall legendary investor Warren Buffet's comment that Bitcoin is a 'bubble' and 'probably rat poison squared' (CNBC, May 4, 2019). We often read sensational headlines such as "Bitcoin loses $21 billion in market cap in 24 h" (CNBC, May 17, 2019) or "Majority of Bitcoin trading is a hoax, a new study finds"

Abbreviations: CNBC, Consumer News & Business Channel; EMH, Efficient Market Hypothesis; XGB, eXtreme Gradient Boosting; FFNN, Feed Forward Neural Network; GRNN, generalized regression neural architecture; GB, Gradient Boosting; LASSO, Least Absolute Shrinkage and Selection Operator; LightGBM, Light Gradient Boosting Machine; LSTM, long short-term memory; RF, Random Forest; SHAP, SHapley Additive exPlanations; SDAE, Stacked Denoising Autoencoders.

☆ This work was supported by Major Program of the National Natural Science Foundation of China [No. 71991474] and Innovative Research Group Project of National Natural Science Foundation of China [No. 71721001].

* Corresponding author.
E-mail address: lizf6@sustech.edu.cn (Z. Li).

https://doi.org/10.1016/j.ribaf.2023.101905
Received 15 November 2022; Received in revised form 5 February 2023; Accepted 9 February 2023
Available online 19 February 2023
0275-5319/

(CNBC, March 22, 2019) or "Why Bitcoin Could Be Headed To $10,000 And Then A Lot Higher After That" (Seeking Alpha, April 28, 2019). These headlines can be terrifying for some investors. While the current market price of Bitcoin at $23,368 (February 4, 2023) has proven many pessimists wrong, it cannot mask the reality that the currency has fallen about 65 % from its peak price of $67,566 on November 7, 2021. Despite such unprecedented volatility, interests in cryptocurrencies remain strong.

The cryptocurrency market has been resilient to global economic, political, and pandemic-related crises. While the global economy was poised for a healthy recovery from the supply chain and the pandemic crises, the recent Russian invasion of Ukraine has led to a renewed interest in the use of cryptocurrencies for trade, as safe-haven currencies, and for avoiding sanctions. On top of this, inflation continues to hammer traditional assets as investors react to eroding values of their wealth. As a result, despite high volatility, cryptocurrencies may continue to attract attention from investors and speculators.

Forecasting cryptocurrency returns in a choppy market is a daunting task. But accurate predictions can provide a reference for traders, investors, and financial analysts. Many articles have predicted the prices or returns of bitcoin and other cryptocurrencies. For example, Chowdhury et al. (2020) use artificial neural networks, K-nearest neighbors, gradient boosting trees, and ensemble algorithms to predict the Cryptocurrency Index 30 and its nine component cryptocurrencies. They claim that their best model performs better than other machine learning models in previous literature. Sun et al. (2020) apply the Light Gradient Boosting Machine (LightGBM) method to predict the price for a market consisting of 41 cryptocurrencies. Akyildirim et al. (2021) apply logistic regression, support vector machines, artificial neural networks, and random forests to predict the price trends of the twelve most liquid cryptocurrencies and find that support vector machines perform best. Liu et al. (2021) use a deep learning method called Stacked Denoising Autoencoders (SDAE) to predict the future price of Bitcoin and find that it performs better than back propagation neural networks and support vector machines. Lahmiri and Bekiros (2019) use deep learning techniques to predict the price of Bitcoin, Digital Cash, and Ripple. They find that long short-term memory (LSTM) performs much better than generalized regression neural architecture (GRNN). Indera et al. (2018) predict Bitcoin price using a multi-layer perceptron-based nonlinear autoregressive with an exogenous input model. Zhang et al. (2021) proposed an improved machine learning model to predict cryptocurrencies' prices, which performs better than other prevailing machine learning models. Yae and Tian (2022) forecast the returns of Bitcoin, Ethereum, and Ripple using correlation with stock markets and achieve out-of-sample R^2 up to 2.75 % using Rank regressions. Bouri et al. (2022) use three-pass regression filter techniques to forecast the returns of the top five cryptocurrencies and find that the models perform better than a benchmark random-walk model. Jiang et al. (2022) propose a new model to forecast the returns and risks of Bitcoin, Ripple, and Litecoin. The model combines a normal-heavy-tailed mixture with a time-varying weight and an accelerated GAS (Generalized Autoregressive Score) approach. They find that the forecasting performance of their model, in terms of Value-at-Risk, performs better than forecasts based on other models, such as GARCH. Using machine learning techniques, Wang et al. (2022) study whether informed trading could forecast the returns of 12 cryptocurrencies. They find that informed trading generally does not have significant prediction ability for the whole market. Ren et al. (2022) report an exhaustive literature review covering studies using machine learning algorithms in predicting cryptocurrency prices. According to the authors, researchers generally use several machine learning techniques seeking model accuracy and better forecasts. Recently, Orte et al. (2023) applied the Random Forest technique to forecast Bitcoin price direction. They find that this technique better predicts long positions than short ones.

Section 2 will address selected articles that study what factors can predict or influence cryptocurrency returns or prices. However, the existing studies only focus on the cryptocurrency market or several cryptocurrencies. Few of them study more than 50 cryptocurrencies. On the other hand, the total size of the cryptocurrency market is $856 billion, and there are more than 9314 active out of 21,844 cryptocurrencies (explodingtopics.com, December 2, 2022). While Bitcoin's market cap has fallen from 85 % in 2010 to just 38 % as of this writing, small-cap cryptocurrencies have become increasingly important. In addition, few of these articles compare the predictive power of the factors, and few study nonlinear relationships between cryptocurrency returns and factors.

This study uses a rich set of variables to predict returns of 3703 cryptocurrencies over a long period. Ordinary Least Squares (OLS) and eXtreme Gradient Boosting (XGB) are our main methods. In the latter part of this paper, we demonstrate that XGB performs the best among several competing machine learning methods, including Least Absolute Shrinkage and Selection Operator (LASSO), Ridge, Random Forest (RF), Gradient Boosting (GB), and Feed Forward Neural Network (FFNN). We further discuss the advantage of XGB over these methods in Section 4.6.

This study uncovers some interesting results using daily data. First, the relationships between our input features[1] and forecast cryptocurrency return are nonlinear. Second, eXtreme Gradient Boosting, a tree-based method, is always better than OLS at the individual cryptocurrency level, suggesting that XGB can discover nonlinear relationships between features and returns. Third, the two best methods for forecasting returns are "XGB with all features" and "OLS with only 1-day lagged return (P1DR)" at the portfolio level, and the former is slightly better than the latter. We also show that with an additional 30 features, the XGB can generate a 7.084 % (2.429 %) average daily return with a 1.95 (0.27) Sharpe ratio on an equal (capital)-weighted portfolio. They are higher than the corresponding only-P1DR-OLS average returns by 0.8 % (0.6 %) and Sharpe ratio by 0.12 (0.06). An investor using the only-P1DR-OLS model can earn better returns by using the XGB model. Fourth, there is a size effect as small cryptocurrencies' return is much more predictable than large ones. Fifth, there is a strong nonlinear one-day reversal effect on cryptocurrency returns. We also find a seven-day reversal effect when the previous seven-day return is negative. Finally, a convex relationship between standard deviation and returns indicates that riskier cryptocurrencies have higher future returns.

Regarding variable importance, we find the following results. First, at both the individual and portfolio levels, "previous 1-day

[1] In machine learning models, 'features' are distinguishing characteristics of the predictor variables that can help analyze the dependent variable. These distinguishing characteristics are measurable.

return" is the most powerful feature compared to the combined effects of all other features. Other return-based variables, including standard deviation and "previous 7-day return", also have great predictive power. Second, the global economic climate, represented by the OECD unemployment rate, OECD inflation, and OECD industrial production growth, has high predictive power. Third, assets, including S&P 500, Hushen 300, KOSPI, gold, and PIMCO bond ETF (bond), have moderate predictive ability. Results show that cryptocurrencies are closely linked to stocks, followed by gold, commodity, and bonds. Surprisingly, oil has the weakest relationship with cryptocurrencies. Fourth, Google searches have a higher predictive power for large cryptocurrencies than small ones. Finally, the cost of mining cryptocurrencies, or production cost, represented by the three electricity-related variables, has a weak predictive power.

Our paper makes several contributions. First, we examine all cryptocurrencies available on Coinmarketcap.com. This helps to understand the return predictability of both large and small cryptocurrencies. Second, we consider input features that have not been tested before and compare the prediction power of a rich set of factors. Third, our primary machine language method, XGB, can capture interactions among input features and non-linearities in the model. It is robust to multi-collinearity, missing values, and outliers and does not require the data to conform to a particular statistical distribution. Fourth, our models generate exceptionally high out-of-sample R^2 and portfolio returns. The expanded XGB model achieves an out-of-sample R^2 of 4.855 %. We find that the high R^2 is primarily achieved through the strong one-day reversal effect on cryptocurrency returns. Fifth, our results indicate that cryptocurrencies behave more like conventional assets than fiat currencies since variables, including lagged returns, can predict future returns. In addition, as an asset, cryptocurrency is not weakly efficient.

The plan of the paper is as follows. In Section 2, we discuss the data and the variables. Section 3 presents the machine learning methods employed in the paper. Section 4 reports the results and the analysis. Section 5 provides robustness checks. The final section concludes the study.

2. Data and variables

This article uses a rich set of variables as predictive factors for cryptocurrency returns. The rationale for using such an exhaustive list of input features stems from the various findings in the literature on factors predicting cryptocurrency returns. Previous research has recognized cryptocurrency as an alternative medium of exchange, similar to conventional currencies (see Polasik et al., 2015). Bitcoin, for instance, has the potential to become an e-commerce payment method and a solid competitor to established money-transfer services (Woo et al., 2013). Schilling and Uhlig (2019) demonstrate that in an endowment economy with both the dollar and Bitcoin, the evolution of Bitcoin is correlated with the dollar. We, therefore, conjecture that conventional currency returns can predict cryptocurrency returns. Following Polasik et al. (2015), we examine the relation between cryptocurrency and the two major currencies, the U.S. dollar and the Euro.

Another strand of research focuses on internal and external factors that affect cryptocurrency prices. For instance, supply and demand for cryptocurrencies are essential determinants of their prices. Popularity (or investor attention) can measure cryptocurrency demand—increased interest in the currency results in increased demand, leading to higher prices. Liu and Tsyvinski (2021) and Polasik et al. (2015) find that popularity positively relates to cryptocurrency returns. Kristoufek (2015) finds that investor attention drives Bitcoin prices up during price explosions, and investor inattention drives Bitcoin prices down during rapid price declines. Bouoiyour and Selmi (2015) and Li and Wang (2017) find that lagged Google searches affect Bitcoin prices. Moreover, Ciaian et al. (2015) show that particular Bitcoin supply and demand factors do affect its price. They use the number of transactions and addresses to measure the size of the Bitcoin economy, which affects demand. Furthermore, they use the total number of Bitcoin to measure the total stock of Bitcoin in circulation, which affects supply. Hence, we use the circulating supply to measure the supply of a cryptocurrency and Google search to measure popularity. Specifically, we calculate the difference between Google searches for a particular cryptocurrency in any week and the average of those searches in the previous four weeks. The original Google search measure for each cryptocurrency is between 0 and 100.

Factors such as network size, cryptocurrency usage, mining cost, and difficulty can also affect prices. As cryptocurrencies operate in a peer-to-peer transaction system, consumer utility rises as network size increases, impacting prices. This is referred to as network externality (Katz and Shapiro, 1985). Accordingly, network factors can impact cryptocurrency prices. So, cryptocurrency usage constitutes a key valuation determinant, measured by trading volume (Kristoufek, 2015), the number of wallet users, or the number of transactions (Liu and Tsyvinski, 2021). Therefore, we use the number of wallet users of Bitcoin, the number of transactions in Bitcoin, and the volume of trade for each cryptocurrency to measure the size and usage. These factors can influence both the network externality effect and demand.[2]

With regard to mining costs, some scholars claim that cryptocurrencies can be regarded as commodities. Hayes (2017) finds that, as with traditional commodities, the value of cryptocurrencies is derived from their production costs. Sockin and Xiong (2020) build a model in which the value of cryptocurrencies is positively related to mining costs. Following Liu and Tsyvinski (2021), we use electricity generation in the United States and China and the average electricity price in the United States to measure mining costs. Another potential predictor for cryptocurrency returns is mining difficulty. This latter, taken as a proxy for mining cost, popularity, and system security, positively affects cryptocurrency value. Our rationales are as follows. First, mining difficulty is a proxy of mining (computational) power. The more mining power is needed, the more costly it is to mine cryptocurrencies (Hayes, 2017; Li and Wang, 2017). Moreover, the price of an asset should be positively related to production cost. Second, the more people mining, the fiercer

[2] We use the value of Bitcoin as a proxy for the cryptocurrency market because it has the highest market cap.

competition increases the mining difficulty. So, mining difficulty can measure the popularity of a cryptocurrency (Hayes, 2017). Third, mining is a process of validating and securing transactions. Therefore, greater mining difficulty increases system security, increasing a cryptocurrency's value (Li and Wang, 2017).

As noted earlier, cryptocurrencies can be regarded as a commodity and could serve as a store of value. So, like other traditional assets, cryptocurrency returns could be exposed to precious metals such as gold. For instance, Bitcoin has been compared to gold because of its limited supply and high mining cost and has been labeled as a safe haven investment. Bouri et al. (2018) find nonlinear, asymmetric, and quantiles-dependent relationships between Bitcoin price and the following lagged indexes: S&P GSCI commodity, gold price, U.S. dollar, PIMCO investment-grade bond ETF, and MSCI world stock. Bouoiyour and Selmi (2015) document a strong link between the price of Bitcoin and the lagged price of Chinese stock. Zhu et al. (2017) find that Bitcoin price is influenced by lagged gold price, the U.S. Consumer Price Index (CPI), the dollar index, Dow Jones Industrial Average, and Effective Federal Funds Rate. Li et al. (2022) find that cryptocurrency's price volatility is connected with the Chinese stock market index. We also include the Market State Dynamic Commodity Index Total Return (MSDCITR),[3] an S&P long/short commodity index, PIMCO investment-grade bond ETF, Effective Federal Funds Rate, oil price, S&P 500 index, China Hushen 300 index, and South Korea KOSPI index as input features. We use these stock indexes from different countries instead of a global stock index, such as the Morgan Stanley Capital International Index (MSCI), because we are interested in the relationship between national stock indexes and cryptocurrencies. Besides, the correlation between the MSCI return and S&P 500 return is 0.97. So, using MSCI alone is almost like using S&P 500.

Looking at the possible exposure of cryptocurrency returns to conventional financial assets and macroeconomic factors, we include three global macroeconomic variables: the OECD unemployment rate, OECD industrial production growth, and OECD inflation in consumer prices, as in Polasik et al. (2015). Inspired by Kristoufek (2015), we examine the relationship between cryptocurrency returns and financial uncertainty, measured by the CBOE volatility index (VIX). Finally, we consider cryptocurrency-specific standard deviation and skewness as cryptocurrency risk. We also include the cumulative 1, 7, and 30 days previous return to detect the momentum effect in cryptocurrency returns. A factor's current and lagged values have been used in the literature as predictors. In our study, we use the lagged value.

We collect price and volume data for 3703 cryptocurrencies from Coinmarketcap.com, covering April 28, 2013, to November 30, 2021. Following Liu and Tsyvinski (2021), we exclude cryptocurrencies with less than 1 million U.S. dollar market capitalization. Most input features are in log difference. We winsorize return and the volume at 1% and 99% based on the premise that outliers in these two variables would drive their means to very large numbers. For all input features, we also replace missing values with zeros. Moreover, we use a rich set of variables as predictive factors for cryptocurrency returns. These variables include a range of financial assets, such as equity, commodity, bonds, currency, macroeconomic indicators, and cryptocurrency-specific factors, such as blockchain mining features and search engine outputs. They are listed in Table 1, along with their descriptions, frequency, and data sources.

Table 2 presents the summary statistics. Average daily cryptocurrency returns are positive and have a large standard deviation. This is typical of the cryptocurrency market. The volatility is also evident by looking at minimum and maximum returns, − 44.728 % and 86.356 %, respectively. Returns are also highly skewed, with both negative and positive skewness. A similar observation is valid for weekly, monthly, and annual returns. The average volume has higher volatility than most other input features. Statistics for the macroeconomic variables show a large mean for the unemployment rate (6.226 %) relative, for instance, to inflation (0.272 %). VIX has the highest volatility (7.653 %) among the macroeconomic variables.[4] Additionally, the S&P500 has a high average value (0.051 %), while oil has the lowest average value (−0.093 %). Crude oil is the highest volatile asset (8.097 %), reaching a maximum level of 53.08 %. The average change in the U.S. dollar index is 0.004 % with 0.299 % volatility. Compared to that, the Euro/dollar average change and volatility are 0.001 % and 0.196 %, respectively. Finally, concerning the cryptocurrency-specific variables, electricity generation in China (0.010 %) is growing higher than in the U.S. (0.003 %). The change in the average retail price of electricity in the U.S. ranges between − 0.165 % and 0.338 %.

2.1. Model

The two prediction models we use are OLS and XGB (Chen and Guestrin, 2016). OLS is a simple linear model, and XGB is a tree-based nonlinear model. Next, we describe the XGB model in detail.

2.2. XGB model

The XGB model uses a gradient-boosting framework. Gradient boosting is an ensemble machine learning algorithm that combines a series of decision trees, where each tree predicts the residuals of the previous tree. After each iteration, i.e. after each tree is built, the residual becomes smaller, improving the model (Friedman, 2001).[5]

For a dataset with N samples and K input features $\mathscr{D} = \left\{ \left(x_i, y_i \right) \right\}_{i=1}^{N}$, XGB uses V trees to predict the final result:

[3] MSDCITR is a proprietary S&P long/short active commodity index, developed by RTM Alternatives, llc.

[4] We tested for the presence of unit roots in all macroeconomic variables, and the ADF statistics confirm the stationarity of all variables.

[5] See Chen and Guestrin (2016) for more details.

Table 1
Variable description.

Variable name	Description	Frequency	Source
Currency-specific Target variables			
Return (%)	Individual cryptocurrency returns.	*Daily*	Coinmarketcap.com
Currency-specific Input Features			
Previous_1_Day_Return (%)[a]	Lagged returns	*Daily*	Coinmarketcap.com
SD (%)	Standard deviation of returns, previous 30 days.	*Daily*	Coinmarketcap.com
Skewness	Skewness of return, the previous 30 days.	*Daily*	Coinmarketcap.com
Volume*(%)[b]	Trading volume.	*Daily*	Coinmarketcap.com
Circulating_Supply*(%)	Cryptocurrency supply in circulation, which is estimated by dividing the market capitalization by the closing price.	*Daily*	Coinmarketcap.com
Google	The difference between Google searches for the name of a cryptocurrency in a given week and the average of those searches in the previous four weeks.	*Weekly*	Google Trends
Difficulty*(%)	Mining difficulty of Bitcoin, set to 0 for non-mineable cryptocurrencies.	*Daily*	Coinmetrics.io
Wallet_Users*(%)	Number of wallet users of Bitcoin.	*Daily*	Blockchain.info
Transaction*(%)	Number of transactions of Bitcoin.	*Daily*	Coinmetrics.io
Macroeconomic Features			
EFFR*(%)	Effective Federal Funds Rate.	*Daily*	Federal Reserve Bank of St. Louis
Unemployment_Rate (%)	OECD unemployment rate (all persons, seasonally adjusted).	*Monthly*	OECD
Industrial_Production (%)	OECD industrial production growth (previous period, seasonally adjusted).	*Monthly*	OECD
Inflation (%)	OECD inflation in CPI (previous period, all items).	*Monthly*	OECD
Uncertainty*(%)	The CBOE volatility index (VIX).	*Daily*	Federal Reserve Bank of St. Louis
Assets			
Gold*(%)	Gold price.	*Daily*	London Bullion Market Association (LBMA)
Commodity*(%)[c]	Market State Dynamic Commodity Index.	*Daily*	RTM Alternatives, llc (MSDCITR)
Bond*(%)	PIMCO investment-grade bond ETF.	*Daily*	Bloomberg
S&P_500*(%)	S&P 500 index.	*Daily*	Federal Reserve Bank of St. Louis
Hushen_300*(%)	China equity Hushen 300 index.	*Daily*	CSMAR
KOSPI*(%)	Korea equity KOSPI index.	*Daily*	Bloomberg
Oil*(%)	Crude oil price.	*Daily*	Federal Reserve Bank of St. Louis
Currency			
Dollar*(%)	The U.S. dollar index.	*Daily*	Federal Reserve Bank of St. Louis (DTWEXBGS)
Euro*(%)	Nominal effective exchange rate of the Euro.	*Daily*	European Central Bank
Mining cost			
US_Elect_Gen*(%)	Net generation of electricity of all sectors in the U.S. This variable is set to 0 for non-mineable cryptocurrencies.	*Monthly*	U.S. Energy Information Administration
Ch_Elect_Gen*(%)	Electricity generation in China. This variable is set to 0 for non-mineable cryptocurrencies.	*Monthly*	National Bureau of Statistics of China and the Price Monitoring Center, NDRC
US_Elect_Price*(%)	Average retail electricity price in the U.S. This variable is set to 0 for non-mineable cryptocurrencies.	*Monthly*	U.S. Energy Information Administration

[a] We used 1, 7, 30, 80, and 360 days lagged returns.
[b] * refers to log differences.
[c] MSDCITR is the total return on a long/short monthly commodity index based on commodities futures included in the S&P GSCI index. The S&P is the calculation agent. See https://www.spglobal.com/spdji/en/custom-indices/little-harbor-advisors/the-market-state-dynamic-commodity-index-tr/#data for more on methodology. The S&PGSCI, in contrast, is a passive long-only index. We thank RTM Alternatives, LLC for the data.

$$\widehat{y}_i = \psi(\boldsymbol{x}_i) = \sum_{v=1}^{V} f_v(\boldsymbol{x}_i; \beta_v, q_v), \tag{1}$$

where q_v stands for each tree structure that assigns a sample to a leaf index, β_v is leaf output scores, $\boldsymbol{x}_i$ is the input feature vector of sample i, and $\boldsymbol{x}$ is the set of all $\boldsymbol{x}_i$. For sample i, XGB uses the decision rule q_v to classify it into a leaf and compute the final prediction by adding up $\beta_{v,a}$ in the leaves that contain this sample ($\beta_{v,a}$ is the output score on a-th leaf in tree v).

XGB minimizes the regularized objective $\mathscr{L}(\psi)$ to build trees:

$$\mathscr{L}(\psi) = \sum_{i=1}^{N} l\left(y_i, \widehat{y}_i\right) + \sum_{v=1}^{V} \Omega(f_v), \tag{2}$$

$$\Omega(f_v) = \gamma S_v + \frac{1}{2}\lambda \|\beta_v\|^2, v = 1, \dots, V \tag{3}$$

Table 2
Summary statistics.

Variable Name	N	Mean	Median	Standard Deviation	Minimum	Maximum
Target variable						
Return (%)	1422214	0.909	-0.092	12.431	-44.728	86.356
Input Features						
Cryptocurrency						
P1DR (%)	1422214	0.737	-0.113	12.225	-44.728	86.356
P7DR (%)	1422214	4.015	-0.536	34.920	-92.914	2982.978
P30DR (%)	1422214	21.082	-1.111	133.920	-99.648	1.41E+ 04
P180DR (%)	1422214	157.069	-1.744	933.878	-100.000	1.25E+ 05
P360DR (%)	1422214	681.459	-0.961	2.35E+ 04	-100.000	9.60E+ 06
SD (%)	1422214	10.135	8.174	7.049	0.000	92.690
Skewness	1422214	0.563	0.468	1.019	-5.199	5.199
Volume (%)	1422214	41.864	-0.829	209.364	-96.729	1633.161
Circulating_Supply (%)	1422214	4.989	0.000	5562.357	-2.45E+ 06	5.844E+ 06
Google	1422214	0.037	0.000	14.598	-99.000	100.000
Difficulty (%)	1422214	0.055	0.000	0.908	-22.662	40.706
Wallet_Users (%)	1422214	0.096	0.060	0.302	0.000	36.221
Transaction (%)	1422214	0.848	-1.083	13.394	-40.549	73.830
Macroeconomic						
EFFR (%)	1422214	0.026	0.000	5.442	-77.273	146.667
Unemployment_Rate (%)	1422214	6.226	5.904	0.895	5.305	8.778
Industrial_Production (%)	1422214	0.131	0.238	2.701	-15.425	8.064
Inflation (%)	1422214	0.272	0.275	0.237	-0.415	0.782
Uncertainty (%)	1422214	0.324	0.000	7.653	-25.906	115.598
Assets						
Gold (%)	1422214	0.020	0.000	0.769	-5.846	5.267
Commodity (%)	1422214	0.019	0.000	0.779	-5.810	8.173
Bond (%)	1422214	0.005	0.000	0.380	-4.956	7.056
S&P_500 (%)	1422214	0.051	0.000	0.997	-11.984	9.383
Hushen_300 (%)	1422214	0.019	0.000	1.030	-8.748	6.715
KOSPI (%)	1422214	0.021	0.000	0.930	-8.394	8.601
Oil (%)	1422214	-0.093	0.000	8.097	-301.966	53.086
Currency						
Dollar (%)	1422214	0.004	0.000	0.299	-2.372	2.052
Euro (%)	1422214	0.001	0.000	0.196	-2.624	1.765
Mining cost						
US_Elect_Gen (%)	1422214	0.003	0.000	0.148	-0.592	0.573
Ch_Elect_Gen (%)	1422214	0.010	0.000	0.106	-0.480	0.782
US_Elect_Price (%)	1422214	0.001	0.000	0.041	-0.165	0.338

where l is a loss function that measures how good a prediction is, i.e. measures the difference between prediction $\widehat{y}_i$ and target y_i, Ω is a regularization term that prevents over-fitting, S_v is the number of leaves in the tree v. To find a solution that minimizes Eq. (2), Chen and Guestrin (2016) re-write Eq. (2) as Eq. (4):

$$\mathscr{L}^{(b)}(\psi) = \sum_{i=1}^{N} l\left(y_i, \widehat{y}_i^{(b-1)} + f_b(\boldsymbol{x}_i)\right) + \Omega(f_b), \tag{4}$$

where, $\widehat{y}_i^{(b)}$ is the prediction of sample i at the b-th iteration. Chen and Guestrin (2016) use second-order Tayler approximation to solve Eq. (4):

$$\mathscr{L}^{(b)}(\psi) \approx \sum_{i=1}^{N}\left[l(y_i, \widehat{y}_i^{(b-1)}) + g_i^{(b)} f_b(\boldsymbol{x}_i) + \frac{1}{2} h_i^{(b)} f_b^2(\boldsymbol{x}_i)\right] + \Omega(f_b), \tag{5}$$

where $g_i^{(b)} = \frac{\partial l(y_i, \widehat{y}_i^{(b)})}{\partial \widehat{y}_i^{(b)}}\Big|_{\widehat{y}_i^{(b)} = \widehat{y}_i^{(b-1)}}$ and $h_i^{(b)} = \frac{\partial^2 l(y_i, \widehat{y}_i^{(b)})}{\partial \widehat{y}_i^{(b)2}}\Big|_{\widehat{y}_i^{(b)} = \widehat{y}_i^{(b-1)}}$ are the first and second-order derivatives of the loss function, respectively. Since we are solving for the minimum value, the constant terms can be removed:

$$\widetilde{\mathscr{L}}^{(b)}(\psi) = \sum_{i=1}^{N}\left[g_i^{(b)} f_b(\boldsymbol{x}_i) + \frac{1}{2} h_i^{(b)} f_b^2(\boldsymbol{x}_i)\right] + \Omega(f_b) \tag{6}$$

Let I_a represent the sample set of leaf a. Eq. (6) can be re-written as

$$\widetilde{\mathscr{L}}^{(b)}(\psi) = \sum_{i=1}^{N}\left[g_i^{(b)}f_b(\boldsymbol{x}_i) + \frac{1}{2}h_i^{(b)}f_b^2(\boldsymbol{x}_i)\right] + \gamma S_b + \frac{1}{2}\lambda\sum_{a=1}^{S_b}\beta_{b,a}^2$$
$$= \sum_{a=1}^{S_b}\left[\left(\sum_{i\in I_a}g_i^{(b)}\right)\beta_{b,a} + \frac{1}{2}\left(\sum_{i\in I_a}h_i^{(b)} + \lambda\right)\beta_{b,a}^2\right] + \gamma S_b. \tag{7}$$

For a certain structure $q(\boldsymbol{x})$, the optimal output score at the b-th iteration of leaf a, $\beta_{b,a}^*$, and the corresponding optimal value of $\widetilde{\mathscr{L}}^{(b)}$ can be calculated as

$$\beta_{b,a}^* = -\frac{\sum_{i\in I_a}g_i^{(b)}}{\sum_{i\in I_a}h_i^{(b)} + \lambda}, \tag{8}$$

$$\widetilde{\mathscr{L}}^{(b)}(q) = -\frac{1}{2}\sum_{a=1}^{S_b}\frac{\left(\sum_{i\in I_a}g_i^{(b)}\right)^2}{\sum_{i\in I_a}h_i^{(b)} + \lambda} + \gamma S_b, \tag{9}$$

where $\widetilde{\mathscr{L}}^{(b)}(q)$ measures the quality of structure q. Since estimating all possible tree structures is difficult, a greedy algorithm is used in practice. XGB starts from a single leaf and grows the tree by adding branches. The optimal way to split a node is by maximizing $\mathscr{L}_{split}$:

$$\mathscr{L}_{split} = \frac{1}{2}\left[\frac{\left(\sum_{i\in I_L}g_i^{(b)}\right)^2}{\sum_{i\in I_L}h_i^{(b)} + \lambda} + \frac{\left(\sum_{i\in I_R}g_i^{(b)}\right)^2}{\sum_{i\in I_R}h_i^{(b)} + \lambda} - \frac{\left(\sum_{i\in I}g_i^{(b)}\right)^2}{\sum_{i\in I}h_i^{(b)} + \lambda}\right] - \gamma, \tag{10}$$

where I_L and I_R are the sample sets of the two-child nodes after splitting, and $I = I_L \cup I_R$. The gain from splitting a node into two child nodes is

$$Gain = \mathscr{L}_{split} + \gamma = \frac{1}{2}\left[\frac{\left(\sum_{i\in I_L}g_i^{(b)}\right)^2}{\sum_{i\in I_L}h_i^{(b)} + \lambda} + \frac{\left(\sum_{i\in I_R}g_i^{(b)}\right)^2}{\sum_{i\in I_R}h_i^{(b)} + \lambda} - \frac{\left(\sum_{i\in I}g_i^{(b)}\right)^2}{\sum_{i\in I}h_i^{(b)} + \lambda}\right]. \tag{11}$$

2.3. Parameters of the XGB model

Several parameters need to be set when using the XGB model. They are as follows. The number of boosting iterations, also called the number of estimators, is the number of trees V. The learning rate is between 0 and 1, which scales each tree's output score when the score is added to the current prediction. The maximum depth is the number of nodes on the longest path from the root node to the farthest leaf node in a tree. Minimum child weight is the minimum value of $\sum_{i\in I_a}h_i^{(b)}$ of a leaf. When $\sum_{i\in I_a}h_i^{(b)}$ of leaf a is smaller than the value of minimum child weight, the leaf will not be further split. Gamma (γ) is the minimum gain required to make a further split on a node. When gains from all possible node splitting are less than gamma, the node will not be further split. Gamma usually is equal to or higher than 0. "Subsample" and "Colsample_bytree" are the fraction of samples and input features randomly selected to grow each tree, respectively. If early stopping is adopted, the model stops growing new trees if the improved value of an indicator (metric name) on the test set[6] is less than a specified amount (controlled by Min Delta) for several consecutive trees (controlled by Rounds).

2.4. Interpretation of the prediction models: SHapley Additive exPlanations (SHAP)

SHAP (SHapley Additive exPlanations) can help us interpret the results of all machine learning models. SHAP assigns an attribution ϕ_k to the k-th feature. It sums the attribution of each feature to calculate the predicted value $\widehat{y}$ for a specific observation. ϕ_k is calculated by predictive function $\psi(\cdot)$. Fig. 1 illustrates this idea. Now, assume we have a dataset that only has three input features, $\{x_{k,i}\}_{i=1}^{N}, k = 1,2,3$. For observation $i, i = 1,\dots,N$, SHAP estimates a constant predicted value $\phi_{0,i}$, and predicted values $\phi_{k,i}, k = 1,2,3$, for input feature $x_{k,i}, k = 1,2,3$, respectively. When no feature is included, the original predicted value is $\phi_{0,i}$. When $x_{1,i}$ is included, the predicted value is $\phi_{0,i} + \phi_{1,i}$. When $x_{1,i}$ and $x_{2,i}$ are included, the predicted value is $\phi_{0,i} + \phi_{1,i} + \phi_{2,i}$. When $x_{1,i}$ to $x_{3,i}$ are included, the predicted value is $\phi_{0,i} + \phi_{1,i} + \phi_{2,i} + \phi_{3,i} = \widehat{y}_i$. Let $z = (z_1, z_2, z_3)'$, where $z_k \in \{0,1\}, k = 1,2,3$. $z_k = 1$ means the k-th input feature $x_{k,i}$ is included in the model, while $z_k = 0$ means $x_{k,i}$ is omitted. Lundberg and Lee (2017) argue that SHAP is the only additive feature attribution method[7] with the three properties listed below. The first is local accuracy, which states that the total of all input feature attributions $(\phi_{1,i} + \phi_{2,i} + \phi_{3,i})$ plus the constant predicted value $(\phi_{0,i})$ equals the estimation of the original model $(\phi_{0,i} + \phi_{1,i} + \phi_{2,i} +$

[6] This test set is different from the test set mentioned elsewhere. To adopt early stopping, the original training set is divided into new training set and test set during the training process. So, this test set actually belongs to the original training set.

[7] "Features" refers to "input features". Additive feature attribution methods assign an attribution to each input feature and calculate the estimated target variable by summing up these attributions.

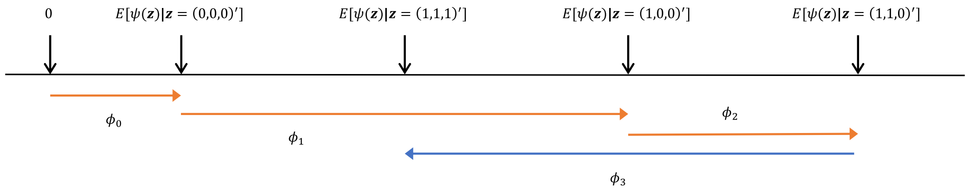

Fig. 1. How SHapley Additive exPlanations (SHAP) works. Note: SHAP is proposed by Lundberg and Lee (2017). This figure illustrates how SHAP works when there are only three input features for observation i ($x_{k,i}, k = 1, 2, 3$). SHAP sums the attribution of each feature to calculate the projected value $\widehat{y}_i$ for observation i. $\phi_{k,i}$ is calculated by predictive function $\psi(\cdot)$. When no feature is included, the original predicted value is $\phi_{0,i}$. When $x_{1,i}$ is included, the predicted value is $\phi_{0,i} + \phi_{1,i}$. When $x_{1,i}$ and $x_{2,i}$ are included, the predicted value is $\phi_{0,i} + \phi_{1,i} + \phi_{2,i}$. When $x_{1,i}$ to $x_{3,i}$ are included, the predicted value is $\phi_{0,i} + \phi_{1,i} + \phi_{2,i} + \phi_{3,i} = \widehat{y}_i$. $z = (z_1, z_2, z_3)'$. $z_k = 1$ means the k-th input feature $x_{k,i}$ is included in the model, while $z_k = 0$ means $x_{k,i}$ is not included. Please note that this figure suggests that $x_{1,i}$ is the first feature added to the model, and $x_{2,i}$ and $x_{3,i}$ are the second and third features, respectively. However, $\phi_{k,i}$ is estimated by averaging over all possible feature-inclusion sequences for nonlinear models.

$\phi_{3,i} = E[\psi(z)|z = (1,1,1)'])$. The second property is missingness, which states that if an input feature is absent, it has no predictive power for the target variable. The third is consistency, which states that when we change a model so that an input feature has a greater impact on the target variable, the predicted value of the target variable ascribed to that input feature will not decrease. Overall, SHAP quantifies how each feature contributes to predicting the model outcome. In practice, several techniques are used to approximate the SHAP value. We use the Tree SHAP in this article.

3. Model building, training, and test sets

We predict returns with lagged input features. For input features with weekly or monthly frequency, we convert them to daily frequency by assigning the value at the end of the previous week or month respectively. Return is specified as

$$r_{j,t} = \psi\left(\boldsymbol{F}_{j,t-1}, \boldsymbol{F}_{C,t-1}, \boldsymbol{F}_{C,t-2}\right) + \varepsilon_{j,t}, \tag{12}$$

where $r_{j,t}$ is the return of cryptocurrency j on day t, $\psi(\cdot)$ represents a prediction model, $\boldsymbol{F}_{j,t-1}$ represents heterogeneous features, including the lagged return, Volume, Circulating_Supply, and Google search for cryptocurrency j, $\boldsymbol{F}_{C,t-1}$ represents common features including Difficulty, Wallet_Users, Transaction, Gold, Hushen_300, KOSPI, Euro, and Ch_Elect_Gen at time $t-1$, and $\boldsymbol{F}_{C,t-2}$ represents common features, including EFFR, Uncertainty, Commodity, Bond, S&P_500, Oil, Dollar, Unemployment_Rate, Industrial_Production, Inflation,[8] US_Elect_Gen, and US_Elect_Price. We use 2-day lagged values of some variables on day $t-2$ instead of $t-1$ because the time zones with times used by these variables are behind Coordinated Universal Time (UTC). The value of $\boldsymbol{F}_{C,t}$ is the same for different cryptocurrencies. Therefore, at the beginning of UTC day t, the values of these variables at the end of UTC day $t-1$ are unavailable.

We adopt a "rolling window" model to predict. Specifically, the first training period is from the second quarter of 2013 to the fourth quarter of 2017, and the first test period is the first quarter of 2018. We then expand the training and test periods one quarter forward. We train and predict each prediction model 16 times and obtain 16 "rolling window" models. Fig. 2 shows the "rolling window" training and test periods. The hyperparameters that we tune are listed in Table 3. We use five-fold cross-validation (Hastie et al., 2009, Chapter 7) to tune the hyperparameters.

Note: This table shows the hyperparameters that are tuned.

3.1. Model evaluation

To evaluate the performance of individual return predictions, we calculate R^2:

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^{N}\left(\widehat{r}_i - r_i\right)^2}{\sum_{i=1}^{N}\left(r_i - \overline{r}\right)^2}, \tag{13}$$

where r_i is the return of sample i, $\widehat{r}_i$ is the estimated return of sample i, $\overline{r}$ is the average return of all samples, N is the total number of samples, SSE is the sum of squares of error, and SST is the total sum of squares. To evaluate the prediction performance at the portfolio level, we calculate the portfolio Sharpe ratio, Sortino ratio, maximum drawdown, and turnover. They are defined as

[8] We could not find an official description of the time zone used for variables downloaded from OECD website (Unemployment_Rate, Industrial_ Production, and Inflation). We believe that data for each country is in each country's own time zone. Since the times of some OECD members are behind UTC, we believe that the time of "all OECD members" should be behind UTC.

Training and Validation			Test
2013.Q2	…	2017.Q4	2018.Q1

Training and Validation			Test
2013.Q3	…	2018.Q1	2018.Q2

…

Training and Validation			Test
2017.Q1	…	2021.Q3	2021.Q4

Fig. 2. Model construction and prediction procedure.

Table 3
XGB Hyperparameters.

Model	Hyperparameters
XGB	Max Number of Estimators = 200, Colsample Bytree = [0.5, 0.7, 1], Subsample = [0.5, 0.7, 1], Learning Rate = [0.1, 0.01], Max Depth = [4, 8, 20, 50, 100, 200, 300], Min Child Weight = [100, 150, 200, 250, 300], Gamma = [0,0.05, 0.1, 0.15, 0.2], Rounds = 10, Min Delta = 0, Metric Name = 'rmse'

$$SharpeRatio = \frac{\overline{r}_p - \overline{r}_f}{\sigma_p}, \tag{14}$$

where $\overline{r}_p$ is the average return of portfolio p, $\overline{r}_f$ is the average risk-free rate, and σ_p is the standard deviation of the returns of portfolio p;

$$SortinoRatio = \frac{\overline{r}_p - \overline{r}_f}{\sigma_{p,down}}, \tag{15}$$

where $\sigma_{p,down}$ is the lower semi-deviation of portfolio p, T is the total number of days during the sample period, $r_{p,t}$ is portfolio return on day t;

$$MaximumDrawdown = \max_{0 \le t_1 \le t_2 \le T} \frac{NAV_{t_1} - NAV_{t_2}}{NAV_{t_1}}, \tag{16}$$

where NAV_t is net asset value of the portfolio on day t, $t_1 \in (0, T)$, $t_2 \in (t_1, T)$;

$$Turnover = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{J}\left(\left|w_{j,t+1} - w_{j,t}\right|\right), \tag{17}$$

where J is the total number of cryptocurrencies, $w_{j,t}$ represents the weight of cryptocurrency j held by a portfolio immediately after rebalancing on day t.

4. Empirical results

4.1. Predictive accuracy of individual cryptocurrencies

Table 4 lists the average R^2 of individual cryptocurrencies on the training and test samples for each prediction model. We calculate the average R^2 of the 16 "rolling window" models. We use two specifications of various sets of features to run each prediction model. First, we include all input features. Second, we remove a cryptocurrency's previous day's return (P1DR, from hereafter) and use all other features. Third, we remove all other features and keep only P1DR. In this way, we have three OLS and three XGB models. As shown later in the paper, P1DR is the most powerful feature for forecasting cryptocurrency returns.

The OLS R^2 on test data using all features, without P1DR and only P1DR, are 1.98 %, − 0.05 %, and 1.44 %, respectively. These results indicate that P1DR can linearly predict cryptocurrency performance. The XGB R^2 on test data using all features, without P1DR, and with only P1DR are 3.90 %, 0.95 %, and 4.33 %, respectively. In comparison, Yae and Tian (2022) achieve out-of-sample R^2 of 2.75 %, 1.53 %, and 2.14 % for Bitcoin, Ethereum, and Ripple, respectively. The R^2 of only-P1DR-XGB is much higher than the without-P1DR-XGB model and even higher than the all-feature-XGB model, which suggests that P1DR is a strong predictor and has more predictive power than other features. The R^2 of without-P1DR-XGB model is only 0.95%. So, the high R^2 is mainly due to P1DR. Later in Fig. 3, we find that the P1DR of large cryptocurrencies is small, mostly closer to 0, while the P1DR of small cryptocurrencies is more volatile. So, the high R^2 is mainly due to the power of P1DR of small cryptocurrencies. All R^2 of XGB are much higher than the OLS models, suggesting that XGB can detect nonlinearities and interactions in the relationship between the input features and cryptocurrency return better than OLS.

Note: This table reports the average R^2 of individual cryptocurrencies on the training and test samples for each prediction model, OLS and XGB. We calculate the average R^2 of the 16 "rolling window" models. We analyze the predictive power of P1DR by removing it and keeping all other features for each prediction model and by removing all other features and only keeping P1DR The approach

Table 4
Individual-cryptocurrency-level R^2 on the training and test samples.

	OLS All Features	OLS Without P1DR	OLS With Only P1DR	XGB All Features	XGB Without P1DR	XGB With Only P1DR
Training (%)	2.831	1.436	1.204	18.193	10.799	4.529
Test (%)	1.975	-0.045	1.439	3.897	0.952	4.327

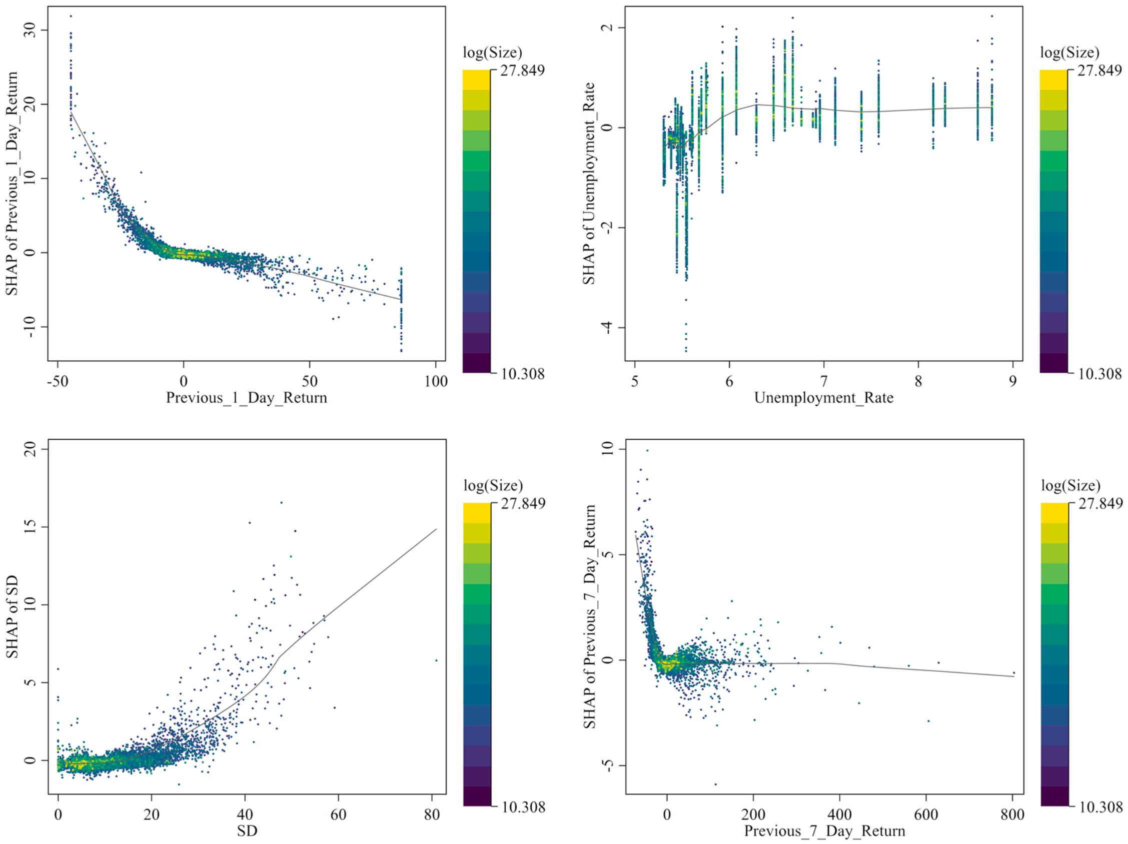

Fig. 3. The relationship between input features and cryptocurrency returns**.** Note: The figures show the nonlinear relationship between important features and the predicted cryptocurrency returns based on the all-feature-XGB model using the test dataset. Each dot represents a sample. The horizontal axis is the value of the input feature of each sample, and the vertical axis is the SHAP value, which is the predicted cryptocurrency return of each sample of each feature. We use different colors to mark cryptocurrency size to analyze how the relationship between features and returns differs across size. The line in each picture represents the trend of all samples.

yields six prediction models.

4.2. Portfolio-level performance

Next, we explore how well each model performs on a portfolio level. At the end of each day, we sort cryptocurrencies into deciles based on each prediction model's out-of-sample predicted next-day return. We also construct a long-short portfolio that goes long the top-decile portfolio and shorts the bottom-decile portfolio. We construct both equal-weighted and market-capital-weighted portfolios. We assume the transaction fee is 0.2 %.[9] If a model accurately predicts the future cryptocurrency returns, the realized returns from portfolios 1–10 should increase monotonically. Suppose one model is more accurate than the other. In that case, the long-short portfolio of the former should yield higher returns since the former assigns more high-return (low-return) cryptocurrencies to the top-decile (bottom decile). Table 5 reports the portfolio performance of each decile and the long-short strategy. Panel A lists the results using all input features. For the equal-weighted strategy, Panel A shows that both OLS and XGB perform well. Generally speaking, the

[9] The transaction fees charged by cryptocurrency exchanges are around 0.2%. Changing the transaction fees did not significantly influence our results.

Table 5
Prediction model portfolio performance.

Panel A: All input features												
	Equal-weighted						Market-capital-weighted					
	OLS			XGB			OLS			XGB		
	Mean	SD	Sharpe	Mean	SD	Sharpe	Mean	SD	Sharpe	Mean	SD	Sharpe
1 (Low)	-0.957	3.889	-0.247	-0.953	4.100	-0.234	-0.032	4.378	-0.009	-0.481	5.235	-0.093
2	-0.069	3.908	-0.019	-0.047	4.255	-0.012	0.076	4.561	0.015	-0.101	5.210	-0.020
3	0.049	4.232	0.010	0.078	4.396	0.017	-0.068	4.673	-0.016	-0.104	4.806	-0.023
4	0.075	4.394	0.016	0.077	4.618	0.015	-0.014	4.862	-0.004	0.031	4.913	0.005
5	0.142	4.464	0.031	0.129	4.638	0.027	-0.118	4.944	-0.025	-0.025	5.199	-0.006
6	0.238	4.519	0.051	0.240	4.718	0.050	-0.102	5.175	-0.021	-0.027	5.342	-0.006
7	0.313	4.562	0.067	0.378	4.817	0.077	-0.163	5.422	-0.031	0.165	5.688	0.028
8	0.603	4.625	0.129	0.499	4.946	0.100	-0.153	5.525	-0.029	0.009	5.782	0.001
9	1.103	4.645	0.236	0.917	4.838	0.188	-0.126	6.175	-0.021	0.032	6.339	0.004
10 (High)	5.274	5.125	1.028	5.734	5.307	1.079	1.255	8.965	0.139	1.501	9.075	0.165
H-L	6.231	3.622	1.719	6.687	3.508	1.905	1.287	8.906	0.144	1.983	8.959	0.221
Panel B: Without P1DR												
	Equal-weighted						Market-capital-weighted					
	OLS			XGB			OLS			XGB		
	Mean	SD	Sharpe	Mean	SD	Sharpe	Mean	SD	Sharpe	Mean	SD	Sharpe
1 (Low)	0.146	3.263	0.043	-0.052	3.798	-0.015	0.092	3.662	0.024	-0.102	4.646	-0.023
2	0.112	3.997	0.027	-0.026	4.442	-0.007	0.090	4.715	0.018	-0.097	4.884	-0.021
3	0.154	4.283	0.035	0.006	4.772	0.000	-0.041	4.834	-0.010	-0.105	5.712	-0.019
4	0.216	4.438	0.047	0.054	5.035	0.010	0.125	4.995	0.024	-0.144	5.368	-0.028
5	0.217	4.573	0.046	0.129	5.126	0.024	-0.057	5.183	-0.012	0.033	5.842	0.005
6	0.288	4.690	0.060	0.131	5.094	0.025	-0.117	5.308	-0.023	-0.053	6.388	-0.009
7	0.287	4.638	0.061	0.308	4.919	0.062	-0.085	5.612	-0.016	0.002	5.901	-0.001
8	0.492	4.661	0.104	0.509	4.835	0.104	-0.201	5.453	-0.038	-0.053	6.279	-0.009
9	1.005	4.671	0.214	0.761	5.114	0.148	-0.001	6.268	-0.001	-0.092	6.539	-0.015
10 (High)	3.891	5.109	0.761	4.633	5.324	0.869	-0.020	9.783	-0.003	0.502	9.796	0.051
H-L	3.746	3.464	1.080	4.684	3.473	1.347	-0.112	9.162	-0.013	0.603	9.361	0.064
Panel C: With Only P1DR												
	Equal-weighted						Market-capital-weighted					
	OLS			XGB			OLS			XGB		
	Mean	SD	Sharpe	Mean	SD	Sharpe	Mean	SD	Sharpe	Mean	SD	Sharpe
1 (Low)	-1.183	4.594	-0.259	-0.464	4.340	-0.108	-0.663	7.509	-0.089	-0.358	5.757	-0.063
2	0.006	4.384	0.000	0.236	4.208	0.055	0.005	5.453	0.000	0.102	5.242	0.018
3	0.213	4.235	0.049	0.249	4.228	0.058	0.144	5.154	0.027	0.062	4.668	0.012
4	0.308	4.210	0.072	0.247	4.168	0.058	0.049	4.584	0.010	-0.052	4.590	-0.013
5	0.292	4.271	0.067	0.212	4.294	0.048	-0.049	4.568	-0.012	-0.112	4.599	-0.026
6	0.329	4.297	0.075	0.268	4.396	0.060	-0.041	4.675	-0.010	-0.099	4.768	-0.022
7	0.402	4.365	0.091	0.310	4.388	0.069	-0.097	4.832	-0.021	-0.070	4.853	-0.016
8	0.443	4.447	0.098	0.349	4.557	0.075	-0.264	4.918	-0.055	-0.118	5.225	-0.024
9	0.800	4.592	0.173	0.582	4.691	0.123	-0.095	5.416	-0.019	0.071	5.620	0.012
10 (High)	5.117	5.011	1.020	4.563	5.139	0.887	1.189	7.240	0.164	0.792	6.855	0.115
H-L	6.297	3.436	1.831	5.024	3.328	1.508	1.852	8.750	0.211	1.149	6.716	0.170

realized returns increase monotonically for both models. The long-short portfolio of the OLS and XGB models earn 6.2 % and 6.7 % daily returns, respectively, which suggests that XGB performs slightly better than OLS. Panel A shows that OLS performs poorly for the market-capital-weighted strategy since the realized returns do not increase monotonically, yet the long-short strategy earns a 1.3 % daily return. XGB performs much better than OLS. The realized returns from the XGB model are monotonically increasing more than the OLS model, and the long-short portfolio achieves a 2.0 % daily return.

As mentioned before, returns on equal-weighted portfolios of OLS increase monotonically, and OLS equal- and capital-weighted portfolios earn 6.2 % and 1.3 % daily returns, respectively. These results indicate that linear models can help describe at least part of the relationship between some input features and cryptocurrency returns.[10] After trying various input feature subsets, we find that the 1-day lagged cryptocurrency return has strong predictive power. Panel B lists the results without P1DR. Panel C lists the results with only P1DR as an input feature.

Without P1DR, OLS's equal-weighted long-short portfolio performance drops to 3.7 %, and the capital-weighted portfolio generates a negative return. Also, with a negative R^2 (Table 4), these results suggest limited linear relationships between other features and

[10] In some cases, although the actual relationship may be non-linear, a linear relationship can help predict the variable on the y-axis. See the first graph in Fig. 3 as an example.

cryptocurrency performance.[11] On the other hand, the long-short portfolio performance of XGB on both equal- and capital-weighted portfolios is better than the corresponding OLS, suggesting that there are nonlinearities and interactions in the relationship between other input features and future performance.

Using only P1DR, the portfolio performances on both equal- and capital-weighted portfolios and the test R^2 (Table 4) of OLS and XGB are better than corresponding models where we use other features. This indicates that the predictive power of the previous 1-day return is stronger than all other features combined. Note that the portfolio performance of only-P1DR-XGB is worse than only-P1DR-OLS, but the R^2 of the former is higher than the latter (Table 4).

Lastly, compared to the corresponding capital-weighted portfolios, the realized returns for equal-weighted portfolios increase monotonically. This confirms that small cryptocurrencies' returns are easier to predict.

Note: At the end of each day, we sort cryptocurrencies into deciles based on each model's out-of-sample predicted next-day return. The portfolios are rebalanced every day and held for one day. This table reports the performance, net of fees, for each decile and a long-short strategy that buys the top decile and shorts the bottom decile. The table reports both the equal-weighted and market-capital-weighted portfolio performance. We assume the transaction fee is 0.2 %. Panel A lists the results using all input features. Panel B lists the results without the P1DR. Panel C lists the results with only P1DR as an input feature. Mean and SD are in percentage.

Next, we further examine the performance of the long-short portfolio of each model. We also construct portfolios based on naïve strategies, assuming investors invest in all cryptocurrencies. Table 6 reports the performance, net of fees, and turnover ratio for the long-short portfolio of each of the six methods and the naïve strategy portfolio. Specifically, it lists performance measures, including mean, median, minimum, maximum, standard deviation, Sharpe ratio, maximum drawdown, and Sortino ratio. We hold the portfolio for one day and rebalance it every day. Panel A reports the equal-weighted portfolio performance, and Panel B reports the market-capital-weighted portfolio performance.

All six models except the capital-weighted without-P1DR-OLS perform much better than the corresponding naïve model with the same weighted strategy in terms of mean, maximum, Sharpe ratio, Sortino ratio, and maximum drawdown. Again, this indicates nonlinear relationships between input features and future return, and P1DR can linearly predict cryptocurrency return. Consistent with the previous results, only P1DR OLS and all-feature-XGB perform the best. The latter performs slightly better than the former in terms of mean, maximum, and Sharpe ratio for both equal- and capital-weighted strategies. Later in this article, we show that with an additional 30 features, the XGB model earns 7.084 % (2.429 %) mean daily return with a 1.95 (0.27) Sharpe ratio on an equal (capital)-weighted portfolio. They are higher than the corresponding only-P1DR-OLS mean returns by 0.8 % (0.6 %) and Sharpe ratio by 0.12 (0.06). An investor can earn a good performance with the only-P1DR-OLS model but a better performance with this new XGB model.

The turnovers of only-P1DR-OLS and all-feature-XGB are higher than other portfolios with the same weighting strategy, suggesting that these portfolios achieve good performances through active management. Finally, equal-weighted portfolios have a higher mean, median, Sharpe ratio, Sortino ratio, and lower standard deviation and maximum drawdown than the corresponding capital-weighted portfolios. This confirms that our model predicts small cryptocurrencies better than large ones.

How long should an investor hold our portfolio? Table 7 reports daily portfolio returns for the next 1, 2, 7, 15, 30, and 60 days, net of fees. We invest in portfolios daily but can hold for longer than a day. In other words, we keep investing at the beginning of each day and check the return when the portfolio is held for a longer time. Panel A reports the equal-weighted portfolio performance, and Panel B reports the market-capital-weighted portfolio performance. The results show that for all the five methods that can predict future cryptocurrency returns, the long-short portfolio can generate positive returns for 180 days for equal-weighted portfolios and 3–30 days for capital-weighted portfolios. However, the daily return declines dramatically even for the equal-weighted portfolios. For example, for the all-feature-XGB method, an investor can earn 6.7 % daily return on the first day but can only earn 2.2 % or 0.9 % daily return if he holds the portfolio for 3 or 7 days, respectively. These results suggest that it would be better to rebalance the portfolio daily. Nevertheless, even if an investor holds the portfolio for 180 days, he would still earn 0.1 % daily, i.e., about 36.0 % annual return. The fact that good performance of equal-weighted portfolios lasts longer than capital-weighted portfolios indicates that small cryptocurrencies' returns are easier to predict.

4.3. Importance of input features

Next, we identify which feature has the most predictive power using variable importance based on the SHAP value for each input feature for the XGB model. The variable importance of a feature is estimated by taking the average of the absolute SHAP value of the test dataset. We then rank the features based on their "Mean(|SHAP|)" value. The higher the ranking, the more important the feature is. Finally, we calculate the average ranks of the 16 "rolling window" models as the final rank for each feature. When calculating the SHAP, we randomly selected 1 % of observations in the test dataset when there were more than 1,000,000 observations in the entire test dataset (from 2018. Q1 to 2021. Q4). An alternative is to randomly select 10 % of observations if there are more than 100,000 observations or use all data if there are less than 100,000 observations in the test dataset. We randomly select 1 % or 10 % observations because otherwise, the estimation will be very time-consuming, notwithstanding that the results based on more than 10,000 observations are accurate enough.

Panel A of Table 8 reports the results of the entire sample. We find that the OECD unemployment rate, OECD inflation, and OECD

[11] R^2 will be negative when SSE is bigger than SST, indicating that the estimations of the model ($\widehat{r}_i$) is worse than just taking the mean of the observations ($\overline{r}$) as an estimation. Note that the R^2 of without-P1DR-OLS is negative, which indicates that the average return of all cryptocurrencies is a better prediction than the prediction of this model.

Table 6
Performance of the long-short portfolios.

Panel A: Equal-weighted portfolios							
	Naïve	OLS All Features	OLS Without P1DR	OLS With Only P1DR	XGB All Features	XGB Without P1DR	XGB With Only P1DR
Mean (%)	0.680	6.231	3.746	6.297	6.687	4.684	5.024
Median (%)	0.988	6.070	3.609	6.178	6.523	4.477	4.849
Min (%)	-34.330	-14.736	-19.010	-6.136	-9.347	-7.949	-5.319
Max (%)	19.797	28.719	24.291	28.513	30.577	26.320	26.330
SD (%)	4.223	3.622	3.464	3.436	3.508	3.473	3.328
Sharpe Ratio	0.160	1.719	1.080	1.831	1.905	1.347	1.508
Maximum Drawdown (%)	82.459	14.736	23.590	8.345	9.347	7.949	5.920
Sortino Ratio	0.245	2.642	2.328	2.664	2.659	2.627	2.702
Turnover	0.041	0.169	0.067	0.207	0.319	0.211	0.197
Panel B: Market-capital-weighted Portfolios							
	Naïve	OLS All Features	OLS Without P1DR	OLS With Only P1DR	XGB All Features	XGB Without P1DR	XGB With Only P1DR
Mean (%)	0.147	1.287	-0.112	1.852	1.983	0.603	1.149
Median (%)	0.205	0.860	-0.570	1.879	1.683	0.178	0.818
Min (%)	-36.593	-49.763	-45.002	-53.315	-44.567	-54.005	-46.579
Max (%)	56.154	81.729	79.586	85.312	86.337	83.232	75.171
SD (%)	4.581	8.906	9.162	8.750	8.959	9.361	6.716
Sharpe Ratio	0.031	0.144	-0.013	0.211	0.221	0.064	0.170
Maximum Drawdown (%)	89.586	80.797	99.983	88.033	83.968	87.292	71.451
Sortino Ratio	0.050	0.291	-0.024	0.355	0.423	0.119	0.299
Turnover	0.038	0.683	0.252	0.719	0.837	0.556	0.737

Note: This table reports the performance, net of fees, for the long-short strategy portfolios of each model, namely OLS and XGB, and the naïve strategy. To analyze the predictive power of the feature P1DR, we remove it and keep all other features for each prediction model, and remove all other features and only keep P1DR for each prediction model. In this way, we obtain six prediction models: all-feature-OLS, only-P1DR-OLS, without-P1DR-OLS, all-feature-XGB, only-P1DR-XGB, and without-P1DR-XGB. At the end of each day, we sort cryptocurrencies into deciles based on the out-of-sample predicted next-day return for each prediction method. We also form a naïve strategy portfolio that invests in all cryptocurrencies. The portfolios are rebalanced every day and held for one day. Panel A reports the equal-weighted portfolio performance, and Panel B reports the market-capital-weighted portfolio performance. We assume the transaction fee is 0.2 %.

industrial production growth are significant, indicating that the global economic climate can influence cryptocurrency returns.

P1DR and P7DR are ranked high, suggesting that the cryptocurrency returns have short-term momentum or reversal effect. Previous 30, 180, and 360 days return rank 10, 19, and 23, respectively. The results show that the longer the time horizon of the investment strategy, the lower the predictive accuracy. The standard deviation has excellent predictive power, but skewness has low predictive power.

Assets, including S&P 500, Hushen 300, KOSPI, gold, and bond, have moderate predictive ability. Among them, cryptocurrencies are most closely related to stocks, followed by gold, commodities, and bonds. Oil has the weakest relationship with cryptocurrencies. The rankings of the three stock indexes are similar. The U.S. stock index has the strongest correlation with cryptocurrencies, followed by China and South Korea stock indexes.

The number of wallet users, transactions, and volume measure the size and usage of cryptocurrency. These factors can influence both the network externality effect and demand. Among these three factors, transactions and the number of wallet users have good predictive ability, and volume ranks 14 among all 30 input features. So, we find that the network externality effect and demand influence cryptocurrency returns. On the other hand, cryptocurrency supply (Circulating_Supply) has a weak impact on return. This may be because the supply of many cryptocurrencies is predictable, so the price already includes the supply effect.

With regard to the exchange rates, the dollar and Euro rank 9 and 18, respectively. Since cryptocurrencies are denominated in USD, it is logical that the US dollar is more important than EUR. VIX (Uncertainty) ranks 12. Surprisingly, Google search ranks 24.

Next, we use subsets of samples to check variable importance further. We run a new model to conduct each check. First, since the Google search of unpopular currencies has a lot of 0 values, we select cryptocurrencies with more than 90 % non-zero Google search values. We report the result in columns 1–3 in Panel B of Table 8. Second, we select currencies with the top 2 % cross-sectional market capital. The results are reported in columns 4–6 in Panel B of Table 8. Google search ranks 26 and 13 in the two checks, respectively. The results indicate that popularity has higher predictive power for large cryptocurrencies. Some studies find that investor attention has strong predictive power, maybe because they only studied a few big coins like Bitcoin (for example, Polasik et al., 2015).[12]

[12] We acknowledge that our use of weekly Google search data, not daily data, may weaken the predictive ability of Google searches. We use weekly data because it would be very hard to download daily data for 3703 key words for more than 8 years. Google Trends returns daily data when the timeframe is less than 9 months. It returns weekly data when the timeframe is between 9 months and 5 years. Google Trends blocks IP if a lot of downloading is detected.

Table 7
Portfolio performance by deciles and holding period.

Panel A: Equal-weighted portfolios											
OLS All Features											
	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	H-L
1 day	-0.957	-0.069	0.049	0.075	0.142	0.238	0.313	0.603	1.103	5.274	6.231
3 days	-0.383	0.052	0.076	0.095	0.115	0.150	0.157	0.200	0.372	1.525	1.908
7 days	-0.091	0.092	0.103	0.113	0.100	0.122	0.114	0.114	0.184	0.656	0.747
15 days	0.009	0.094	0.108	0.112	0.113	0.105	0.104	0.098	0.103	0.308	0.299
30 days	0.077	0.125	0.149	0.155	0.161	0.157	0.153	0.134	0.115	0.204	0.127
60 days	0.103	0.148	0.183	0.195	0.211	0.207	0.199	0.186	0.153	0.209	0.106
90 days	0.138	0.175	0.207	0.225	0.229	0.238	0.231	0.221	0.189	0.221	0.083
120 days	0.150	0.200	0.232	0.251	0.251	0.253	0.246	0.249	0.246	0.260	0.110
150 days	0.135	0.204	0.235	0.244	0.248	0.246	0.242	0.234	0.230	0.256	0.121
180 days	0.148	0.197	0.226	0.226	0.231	0.240	0.219	0.205	0.199	0.230	0.082
OLS Without P1DR											
	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	H-L
1 day	0.146	0.112	0.154	0.216	0.217	0.288	0.287	0.492	1.005	3.891	3.746
3 days	0.140	0.108	0.143	0.144	0.116	0.168	0.126	0.169	0.298	0.956	0.816
7 days	0.120	0.113	0.128	0.130	0.118	0.134	0.086	0.101	0.157	0.424	0.304
15 days	0.128	0.107	0.119	0.130	0.112	0.124	0.093	0.069	0.083	0.190	0.062
30 days	0.119	0.134	0.162	0.173	0.170	0.167	0.142	0.125	0.102	0.137	0.018
60 days	0.104	0.156	0.194	0.224	0.214	0.203	0.192	0.198	0.136	0.173	0.069
90 days	0.109	0.191	0.221	0.255	0.240	0.231	0.226	0.227	0.179	0.194	0.085
120 days	0.112	0.216	0.255	0.276	0.250	0.242	0.236	0.256	0.249	0.246	0.134
150 days	0.102	0.220	0.242	0.274	0.240	0.240	0.240	0.238	0.242	0.235	0.134
180 days	0.103	0.221	0.248	0.259	0.225	0.235	0.235	0.195	0.181	0.217	0.114
OLS With Only P1DR											
	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	H-L
1 day	-1.183	0.006	0.213	0.308	0.292	0.329	0.402	0.443	0.800	5.117	6.297
3 days	-0.579	0.061	0.132	0.174	0.161	0.152	0.153	0.163	0.299	1.628	2.206
7 days	-0.200	0.118	0.147	0.142	0.137	0.105	0.105	0.090	0.139	0.714	0.914
15 days	-0.061	0.116	0.125	0.126	0.111	0.102	0.092	0.082	0.109	0.349	0.410
30 days	0.052	0.150	0.160	0.154	0.145	0.140	0.132	0.129	0.133	0.232	0.180
60 days	0.103	0.183	0.188	0.190	0.182	0.185	0.180	0.178	0.181	0.222	0.119
90 days	0.158	0.213	0.211	0.210	0.205	0.205	0.206	0.202	0.211	0.248	0.090
120 days	0.193	0.237	0.235	0.229	0.235	0.231	0.229	0.228	0.234	0.285	0.092
150 days	0.192	0.227	0.222	0.213	0.225	0.233	0.227	0.221	0.242	0.267	0.075
180 days	0.179	0.222	0.203	0.197	0.212	0.216	0.215	0.209	0.224	0.240	0.061
XGB All Features											
	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	H-L
1 day	-0.953	-0.047	0.078	0.077	0.129	0.240	0.378	0.499	0.917	5.734	6.687
3 days	-0.414	0.001	0.035	0.011	0.083	0.157	0.199	0.253	0.358	1.811	2.225
7 days	-0.138	0.075	0.059	0.034	0.062	0.127	0.148	0.166	0.193	0.805	0.943
15 days	-0.034	0.103	0.068	0.023	0.069	0.115	0.118	0.110	0.117	0.410	0.445
30 days	0.070	0.163	0.121	0.081	0.101	0.155	0.155	0.135	0.119	0.266	0.196
60 days	0.130	0.197	0.171	0.119	0.132	0.170	0.196	0.183	0.156	0.243	0.113
90 days	0.166	0.245	0.229	0.188	0.188	0.204	0.236	0.228	0.188	0.265	0.099
120 days	0.187	0.285	0.288	0.273	0.267	0.271	0.277	0.251	0.217	0.309	0.122
150 days	0.176	0.277	0.295	0.292	0.307	0.294	0.281	0.249	0.204	0.305	0.129
180 days	0.173	0.274	0.289	0.282	0.309	0.297	0.263	0.242	0.196	0.265	0.092
XGB Without P1DR											
	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	H-L
1 day	-0.052	-0.026	0.006	0.054	0.129	0.131	0.308	0.509	0.761	4.633	4.684
3 days	-0.025	-0.036	-0.018	0.002	0.055	0.055	0.146	0.166	0.238	1.354	1.379
7 days	0.030	0.006	0.008	-0.009	0.047	0.041	0.094	0.119	0.164	0.605	0.575
15 days	0.056	0.033	0.011	-0.010	0.030	0.019	0.053	0.089	0.107	0.304	0.248
30 days	0.109	0.090	0.086	0.057	0.078	0.086	0.101	0.121	0.128	0.219	0.110
60 days	0.137	0.161	0.156	0.164	0.177	0.162	0.181	0.170	0.165	0.216	0.079
90 days	0.150	0.208	0.228	0.231	0.254	0.229	0.244	0.211	0.194	0.243	0.093
120 days	0.166	0.244	0.263	0.263	0.283	0.273	0.274	0.255	0.226	0.311	0.145
150 days	0.156	0.246	0.276	0.250	0.306	0.298	0.274	0.244	0.214	0.293	0.137
180 days	0.160	0.222	0.219	0.204	0.255	0.258	0.249	0.216	0.197	0.260	0.100
XGB With Only P1DR											
	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	H-L
1 day	-0.464	0.236	0.249	0.247	0.212	0.268	0.310	0.349	0.582	4.563	5.024
3 days	-0.260	0.118	0.141	0.093	0.059	0.049	0.079	0.136	0.227	1.452	1.712
7 days	-0.068	0.114	0.095	0.074	0.043	0.030	0.046	0.082	0.138	0.654	0.722
15 days	0.021	0.120	0.081	0.054	0.031	0.024	0.025	0.052	0.100	0.329	0.308
30 days	0.104	0.167	0.129	0.080	0.061	0.060	0.050	0.083	0.130	0.223	0.120

(*continued on next page*)

Table 7 (*continued*)

Panel A: Equal-weighted portfolios											
OLS All Features											
	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	H-L
60 days	0.140	0.204	0.168	0.137	0.102	0.093	0.095	0.113	0.175	0.216	0.077
90 days	0.186	0.229	0.193	0.168	0.146	0.132	0.139	0.157	0.204	0.243	0.057
120 days	0.214	0.256	0.230	0.209	0.205	0.195	0.188	0.204	0.240	0.285	0.071
150 days	0.206	0.242	0.236	0.216	0.226	0.217	0.208	0.212	0.231	0.272	0.066
180 days	0.194	0.222	0.219	0.200	0.205	0.212	0.203	0.201	0.214	0.241	0.047
Panel B: Market-capital-weighted Portfolios											
OLS All Features											
	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	H-L
1 day	-0.032	0.076	-0.068	-0.014	-0.118	-0.102	-0.163	-0.153	-0.126	1.255	1.287
3 days	0.067	0.123	0.070	0.076	0.022	-0.003	-0.009	-0.147	-0.118	0.200	0.134
7 days	0.122	0.121	0.107	0.101	0.047	0.037	-0.002	-0.115	-0.108	0.010	-0.112
15 days	0.129	0.105	0.111	0.111	0.104	0.061	0.017	-0.036	-0.074	0.021	-0.108
30 days	0.107	0.122	0.146	0.166	0.160	0.124	0.078	0.020	-0.027	0.059	-0.048
60 days	0.081	0.152	0.187	0.208	0.223	0.176	0.127	0.064	0.021	0.034	-0.047
90 days	0.100	0.177	0.237	0.236	0.231	0.200	0.174	0.122	0.076	-0.005	-0.105
120 days	0.100	0.211	0.263	0.270	0.278	0.235	0.222	0.163	0.120	0.028	-0.072
150 days	0.104	0.225	0.255	0.269	0.280	0.253	0.228	0.175	0.120	0.019	-0.085
180 days	0.122	0.234	0.255	0.284	0.275	0.258	0.223	0.177	0.123	0.012	-0.110
OLS Without P1DR											
	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	H-L
1 day	0.092	0.090	-0.041	0.125	-0.057	-0.117	-0.085	-0.201	-0.001	-0.020	-0.112
3 days	0.111	0.095	0.032	0.106	0.008	0.003	-0.044	-0.142	-0.040	-0.133	-0.243
7 days	0.089	0.105	0.093	0.091	0.028	0.077	-0.042	-0.131	-0.032	-0.030	-0.119
15 days	0.090	0.099	0.106	0.117	0.073	0.097	-0.002	-0.113	-0.096	0.021	-0.070
30 days	0.068	0.130	0.166	0.175	0.140	0.132	0.040	-0.007	-0.059	0.125	0.057
60 days	0.068	0.173	0.200	0.234	0.200	0.172	0.094	0.071	-0.020	0.078	0.010
90 days	0.087	0.226	0.260	0.267	0.216	0.182	0.124	0.133	0.055	-0.045	-0.131
120 days	0.108	0.249	0.284	0.316	0.255	0.215	0.169	0.159	0.104	0.016	-0.093
150 days	0.127	0.239	0.274	0.306	0.279	0.245	0.199	0.163	0.099	0.017	-0.110
180 days	0.151	0.246	0.277	0.302	0.268	0.235	0.213	0.165	0.098	0.000	-0.151
OLS With Only P1DR											
	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	H-L
1 day	-0.663	0.005	0.144	0.049	-0.049	-0.041	-0.097	-0.264	-0.095	1.189	1.852
3 days	-0.130	0.059	0.115	0.117	0.103	0.043	0.059	-0.077	-0.054	0.231	0.363
7 days	-0.018	0.108	0.150	0.145	0.129	0.054	0.051	-0.036	-0.045	0.057	0.076
15 days	0.024	0.127	0.146	0.151	0.160	0.090	0.072	0.009	0.003	-0.001	-0.025
30 days	0.104	0.128	0.156	0.151	0.201	0.151	0.099	0.068	0.051	0.015	-0.089
60 days	0.109	0.158	0.173	0.179	0.210	0.170	0.147	0.125	0.122	0.054	-0.056
90 days	0.131	0.194	0.212	0.214	0.243	0.222	0.162	0.158	0.156	0.061	-0.071
120 days	0.143	0.220	0.232	0.241	0.275	0.257	0.204	0.197	0.189	0.100	-0.044
150 days	0.121	0.211	0.230	0.247	0.281	0.277	0.219	0.208	0.193	0.101	-0.020
180 days	0.131	0.218	0.234	0.241	0.290	0.279	0.238	0.213	0.200	0.113	-0.018
XGB All Features											
	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	H-L
1 day	-0.481	-0.101	-0.104	0.031	-0.025	-0.027	0.165	0.009	0.032	1.501	1.983
3 days	-0.153	-0.012	0.003	0.064	-0.002	0.064	0.144	0.057	0.150	0.552	0.705
7 days	-0.042	0.023	0.027	0.040	0.052	0.088	0.140	0.038	0.083	0.210	0.252
15 days	0.023	0.068	0.078	0.080	0.081	0.108	0.119	0.034	0.031	0.143	0.120
30 days	0.073	0.126	0.097	0.101	0.116	0.157	0.123	0.072	0.051	0.124	0.051
60 days	0.133	0.125	0.120	0.130	0.134	0.182	0.140	0.125	0.074	0.108	-0.026
90 days	0.158	0.173	0.178	0.177	0.176	0.224	0.179	0.172	0.113	0.064	-0.094
120 days	0.178	0.206	0.245	0.256	0.241	0.280	0.235	0.215	0.145	0.106	-0.072
150 days	0.176	0.230	0.270	0.293	0.288	0.311	0.256	0.236	0.159	0.104	-0.073
180 days	0.184	0.263	0.299	0.318	0.318	0.335	0.285	0.246	0.166	0.109	-0.075
XGB Without P1DR											
	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	H-L
1 day	-0.102	-0.097	-0.105	-0.144	0.033	-0.053	0.002	-0.053	-0.092	0.502	0.603
3 days	-0.037	-0.073	-0.018	-0.062	0.096	-0.060	0.064	0.006	-0.041	0.123	0.159
7 days	0.030	-0.014	0.035	-0.007	0.029	-0.054	0.039	-0.009	-0.003	0.028	-0.002
15 days	0.050	0.037	0.030	0.021	0.025	-0.049	0.025	0.014	0.020	0.078	0.028
30 days	0.103	0.065	0.052	0.065	0.078	0.014	0.048	0.021	0.035	0.096	-0.008
60 days	0.141	0.128	0.085	0.113	0.117	0.077	0.119	0.081	0.083	0.081	-0.061
90 days	0.164	0.181	0.166	0.171	0.185	0.142	0.196	0.153	0.121	-0.004	-0.168
120 days	0.198	0.214	0.203	0.194	0.229	0.209	0.244	0.177	0.135	0.085	-0.113
150 days	0.211	0.206	0.215	0.192	0.255	0.231	0.248	0.198	0.146	0.085	-0.126
180 days	0.233	0.186	0.170	0.159	0.247	0.208	0.256	0.202	0.139	0.078	-0.155

(*continued on next page*)

Table 7 (*continued*)

Panel A: Equal-weighted portfolios											
OLS All Features											
	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	H-L
XGB With Only P1DR											
	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	H-L
1 day	-0.358	0.102	0.062	-0.052	-0.112	-0.099	-0.070	-0.118	0.071	0.792	1.149
3 days	-0.016	0.076	0.096	0.025	-0.016	0.008	0.061	0.011	0.077	0.094	0.112
7 days	0.060	0.114	0.111	0.070	0.014	0.019	0.046	0.026	0.041	0.053	-0.006
15 days	0.120	0.148	0.115	0.074	0.037	0.067	0.053	0.036	0.056	0.025	-0.095
30 days	0.162	0.182	0.139	0.092	0.090	0.090	0.075	0.065	0.092	0.009	-0.153
60 days	0.171	0.193	0.164	0.147	0.124	0.129	0.100	0.096	0.148	0.075	-0.096
90 days	0.198	0.208	0.182	0.178	0.165	0.162	0.131	0.143	0.172	0.087	-0.111
120 days	0.212	0.228	0.213	0.217	0.213	0.215	0.179	0.194	0.205	0.119	-0.093
150 days	0.203	0.236	0.222	0.239	0.238	0.238	0.204	0.205	0.211	0.128	-0.075
180 days	0.205	0.242	0.231	0.236	0.238	0.228	0.203	0.206	0.224	0.143	-0.063

Note: We have two prediction models, OLS and XGB. To analyze the predictive power of the P1DR, we remove it and keep all other features for each prediction model, and remove all other features and only keep P1DR for each prediction model. In this way, we obtain six prediction models: all-feature-OLS, only-P1DR-OLS, without-P1DR-OLS, all-feature-XGB, only-P1DR-XGB, without-P1DR-XGB. At the end of each day, we sort crypto-currencies into deciles based on the out-of-sample predicted next-day return for each prediction method. We invest in portfolios daily but hold them for longer than a day. In other words, we keep investing at the beginning of each day and check what we would earn if the portfolio is held longer. This table reports the daily portfolio returns, net of fees, for the next 1, 2, 7, 15, 30, 60, 90, 120, 150, and 180 days. Panel A reports the equal-weighted portfolio performance, and Panel B reports the market-capital-weighted portfolio performance. We assume the transaction fee is 0.2 %. All numbers are in percentage.

Third, for non-mineable currencies, mining-related variables, i.e., mining difficulty and the three electricity-related variables, are set to 0. This leads to the possibility of underestimating these variables using the whole sample. We select mineable cryptocurrencies and report the results in the last three columns of Panel B of Table 8. We find that U.S. electricity generation, U.S. electricity price, China electricity generation, and difficulty rank 15, 22, 29, 19, respectively. These results show that production cost, represented by the three electricity-related variables, holds low predictive power. Mining difficulty, a proxy for mining cost, popularity, and system security, does not have much predictive ability. We also investigate if the feature importance is stable. Table 9 reports the number of importance rankings achieved by each feature in the 16 "rolling window" predictions. We find that feature importance is fairly stable.

Altogether, these results indicate that cryptocurrencies behave more like conventional assets and are not weakly efficient since variables, including previous returns, can predict future returns. In addition, the price of cryptocurrency is not affected by production costs.

4.4. The relationship between input features and cryptocurrency return

Next, we explore nonlinear relationships between input features and cryptocurrency returns. Fig. 3 shows the nonlinear relationship between the four most important features and the predicted cryptocurrency returns based on the all-feature-XGB model using the test dataset. Each dot represents a sample. The horizontal axis is the value of the input feature of each sample, and the vertical axis is the SHAP value, which is the predicted cryptocurrency return of each sample of each feature. Since we have found that returns on small cryptocurrencies are easier to predict, we use different colors to mark cryptocurrency sizes. Our aim is to analyze how the relationship between features and returns differs across cryptocurrency sizes. The line in each picture represents the trend of all samples.

We find that cryptocurrencies have a one-day reversal effect. Moreover, this short-term reversal effect is not linear. Near 0, the effect is almost non-existent, and when the previous 1-day returns are large, there is an obvious short-term reversal effect, especially when the returns are negative. We also find that the P1DR of large cryptocurrencies fluctuates around 0, while the P1DR of small cryptocurrencies is more volatile, so the returns of small cryptocurrencies are more predictable. This finding is consistent with Zaremba et al. (2021), who find short-term reversal phenomena in small cryptocurrencies using linear regressions. They claim that the short-term reversal effect is due to illiquidity.

The results also show a 7-day reversal effect when the previous 7-day return is negative. Similar to P1DR, P7DR of large cryptocurrencies fluctuates around 0, while the P7DR of small cryptocurrencies is more volatile, so the returns of small cryptocurrencies are more predictable. There is a convex relationship between SD and return. Riskier cryptocurrencies have higher future returns. We find that the SD of large cryptocurrencies is close to 0, while the SD of small cryptocurrencies is larger, making the returns of small cryptocurrencies more predictable. Finally, when the unemployment rate is between 5.5 and 6.3, its relationship with the returns is generally positive.

4.5. Additional input features

This section adds 30 more features and re-estimates the all-feature-XGB model. We do not use these 30 features in our main analysis because they are highly correlated with at least one of our existing features. For highly correlated features, the feature importance will

Table 8
Feature importance.

Panel A		
Entire Sample		
Features	SHAP	Rank
P1DR	0.723	1
Unemployment_Rate	0.415	2
SD	0.318	3
P7DR	0.294	4
Inflation	0.227	5
Transaction	0.215	6
Industrial_Production	0.180	7
Wallet_Users	0.161	8
Dollar	0.139	9
P30DR	0.132	10
S&P_500	0.122	11
Uncertainty	0.122	12
Hushen_300	0.118	13
Volume	0.113	14
Gold	0.112	15
KOSPI	0.107	16
Commodity	0.101	17
Euro	0.095	18
P180DR	0.094	19
Bond	0.089	20
Skewness	0.084	21
Oil	0.078	22
P360DR	0.057	23
Google	0.052	24
Circulating_Supply	0.048	25
EFFR	0.021	26
US_Elect_Gen	0.012	27
US_Elect_Price	0.010	28
Difficulty	0.010	29
Ch_Elect_Gen	0.009	30

Panel B								
Non-zero Google Search Value			Top 2% in Size			Mineable		
Features	SHAP	Rank	Features	SHAP	Rank	Features	SHAP	Rank
P1DR	0.506	1	Hushen_300	0.172	1	P1DR	0.701	1
Unemployment_Rate	0.366	2	P7DR	0.117	2	SD	0.438	2
Inflation	0.198	3	Inflation	0.116	3	Unemployment_Rate	0.438	3
Transaction	0.187	4	Industrial_Production	0.116	4	P7DR	0.305	4
P7DR	0.173	5	KOSPI	0.109	5	Dollar	0.152	5
Industrial_Production	0.169	6	Unemployment_Rate	0.108	6	Industrial_Production	0.148	6
SD	0.156	7	Gold	0.108	7	Transaction	0.136	7
Wallet_Users	0.152	8	Transaction	0.104	8	P30DR	0.135	8
S&P_500	0.122	9	S&P_500	0.091	9	Volume	0.126	9
Dollar	0.120	10	Dollar	0.081	10	Inflation	0.123	10
KOSPI	0.115	11	Commodity	0.070	11	S&P_500	0.116	11
P30DR	0.111	12	P1DR	0.069	12	P180DR	0.114	12
Gold	0.110	13	Google	0.068	13	Hushen_300	0.112	13
Uncertainty	0.106	14	Bond	0.062	14	Gold	0.112	14
Hushen_300	0.100	15	Euro	0.059	15	US_Elect_Gen	0.108	15
Euro	0.096	16	P30DR	0.055	16	KOSPI	0.107	16
Commodity	0.089	17	Uncertainty	0.052	17	Bond	0.104	17
Bond	0.089	18	P180DR	0.050	18	Uncertainty	0.104	18
Oil	0.072	19	Oil	0.046	19	Difficulty	0.089	19
Volume	0.068	20	Volume	0.045	20	Euro	0.088	20
P180DR	0.061	21	Skewness	0.038	21	Wallet_Users	0.082	21
Skewness	0.039	22	SD	0.036	22	US_Elect_Price	0.079	22
EFFR	0.032	23	Wallet_Users	0.032	23	Oil	0.078	23
P360DR	0.030	24	P360DR	0.031	24	Commodity	0.072	24
Circulating_Supply	0.026	25	Circulating_Supply	0.019	25	Skewness	0.070	25
Google	0.017	26	US_Elect_Gen	0.014	26	P360DR	0.066	26
Difficulty	0.007	27	Ch_Elect_Gen	0.009	27	Google	0.061	27
US_Elect_Gen	0.005	28	EFFR	0.008	28	Circulating_Supply	0.052	28
Ch_Elect_Gen	0.004	29	US_Elect_Price	0.008	29	Ch_Elect_Gen	0.051	29
US_Elect_Price	0.003	30	Difficulty	0.007	30	EFFR	0.035	30

Note: This table lists variable importance based on the SHAP value for each input feature for the all-feature-XGB model. The importance of a feature is calculated by taking the average of the absolute SHAP of the test dataset. We then rank the features based on their "Mean(|SHAP|)" value. The higher the ranking, the more important the feature. Finally, we calculate the average ranks of the 16 "rolling window" predictions as the final rank for each feature. Panel A lists the results of the entire sample. We then run new models using subsets of samples in Panel B. The first three columns in Panel B list results for cryptocurrencies with more than 90 % of the weeks with non-zero Google search value, and columns 4–6 list results of cryptocurrencies with the top 2 % cross-sectional market cap. The last three columns list results for the subsamples of mineable cryptocurrencies.

be underestimated. The new features are the cumulative lagged return of a cryptocurrency of the previous 3, 15, 60, 90, 120, 150, 210, 240, 270, 300, 330, 540, 720, 900, and 1080 days; and kurtosis of return during the previous 30 days. In addition, we include growth rates of the following variables- the confirmation time of Bitcoin; the active number of Bitcoin addresses; the block count of Bitcoin; the growth rate of hash rate of Bitcoin; the growth rate of miner revenue of Bitcoin; the number of transfers of Bitcoin; adjusted U.S. dollar transfer amount of Bitcoin; silver and platinum prices; Morgan Stanley Capital International Index; Straits Times Industrial Index; FTSE 100 Index; Nikkei 225 Index; and the Dow Jones Industrial Average Index. Note that the following variables: Ch_Elect_Gen, US_Elect_Gen, US_Elect_Price, Unemployment_Rate, Industrial_Production, and Inflation are monthly data. At the beginning of each month, last month's data for these variables are unavailable. So, we use data in months − 2, − 3, − 3, − 3, − 4, and − 2, respectively, according to the availability of each of them.

The R^2 for the training and test dataset of the new model are equal to 21.318 % and 4.855 %, respectively. The equal (capital)-weighted long-short portfolios earn 7.084 % (2.429 %) daily returns with a 1.95 (0.27) Sharpe ratio. The returns are higher than the corresponding only-P1DR-OLS returns by 0.8 % (0.6 %), and the Sharpe ratio is higher by 0.12 (0.06). An investor can earn a good performance with the only-P1DR-OLS model but a better performance with this new XGB model.

4.6. Other machine learning methods

We also use LASSO (Tibshirani, 1996), Ridge (Hoerl and Kennard, 1970), RF (Breiman, 2001), GB (Friedman, 2001), and FFNN (McCulloch and Pitts, 1943; Cybenko, 1989; Hornik et al., 1989) to predict cryptocurrency returns. In particular, LASSO and Ridge are penalized linear models. LASSO sets many covariate coefficients to zero, so it could be considered a variable-selection technique. Ridge only sets coefficients close to zero. It can be considered a shrinking technique that keeps the coefficients from getting too big. RF, GB, and XGB are non-linear tree-based models. RF grows many trees (decision-making-diagram) independently. GB and XGB grow trees one at a time in a sequential manner, with each new tree helping to rectify errors in the previous trees. So, GB and XGB can be more accurate than RF and detect intricate data patterns. XGB is a specific implementation of GB. XGB uses second-order gradients of the loss function (Eq. 2) and provides more accurate estimates. FFNN is a non-linear neural network model. Because there are likely non-linear relationships in the data, non-linear models are expected to be better than linear models. And because of the advantages of the XBG model, it can perform better than RF and GB. FFNN is also a powerful model for handling economic data. For example, Gu et al. (2020) find that the best techniques to predict U.S. stock returns are FFNN and, to a lesser extent, tree-based models. It turns out that FFNN performs worse than tree-based models in our study.

The average R^2 of each of these methods is listed in Table 10. The data used are the same as in Table 4. All the main 30 features are used. The average test R^2 of XGB is the best among these methods. The average test R^2 of other tree-based methods - RF and GB - are slightly worse than that of XGB. The average test R^2 of LASSO, Ridge, and FFNN is worse than XGB. These results indicate that there are non-linear relationships in the data and non-linear models can capture them better than linear models. LASSO performs worse than Ridge, indicating that many input features are useful, and we cannot just select a few. FFNN turns out to perform worse than tree-based methods.

5. Robustness

Cryptocurrencies are very volatile. Next, we examine the robustness of models in Table 4 across time and different market situations. In Panel A of Table 11, we list the test R^2 of each model before and after the COVID-19 outbreak. Panel B lists the test R^2 in high and low market fear and uncertainty. Specifically, we contrast the values of the VIX index with their historical medians. We describe the quarter as being in the low (high) fear stage if the index is lower (higher) than the historical median. Finally, we compare S&P 500 index values in panel C with their antecedent medians. One quarter is characterized as being in the up (down) stage if the S&P 500 index is higher (lower) than the historical median.

We find the results to be stable. The ranking of the models' predictive power is similar to that reported in Table 4. For example, all-feature-XGB and only-P1DR-XGB have the highest R^2 in most cases. Without-P1DR-OLS model always produces the lowest R^2. All models perform better in the post-2020 (high S&P 500) period than in the pre-2020 (low S&P 500) period. This is probably because the post-2020 (high S&P 500) period looms over more recent quarters with more small cryptocurrencies, and therefore, easier to be predicted.

6. Conclusions

This study uses machine learning models to forecast returns for 3703 cryptocurrencies. Based on daily data, we find several interesting results. First, the estimated models have high out-of-sample R^2 and portfolio returns. The expanded XGB model achieves an out-of-sample R^2 of 4.855 %. An investor can earn 6.297 % (1.852 %) on an equal (capital)-weighted portfolio using the OLS model with 1-day lagged returns, which generates the second-highest portfolio returns among the six prediction methods in Table 4. In addition, an investor can earn more using an expanded version of the XGB, where we include all input features. We show that the expanded XGB model generates 7.105 % and 2.165 % daily returns on equal- and capital-weighted portfolios. Second, there is a strong

Table 9
Feature importance stability.

Features	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
P1DR	14	1	1																											
Unemployment Rate	1	6	4	2	1	1			1																					
SD		3	4	2	2	2						2		1																
P7DR		2	2	6		1	2	2			1																			
Inflation	1	2		1	1			1		2		1					2		3	1			1							
Transaction		1	2	1	1	1	1	2	1	1		1	1		2		1													
Industrial_Production		1	2		1	2	3	1	1	1					1			2	1											
Wallet_Users				1	3	1	1	1			1		1	1	1			1	1	1		1		1						
Dollar					3	1	1	2	2	1					3	1		2												
P30DR						1	2		4		1	1	2	2		2	1													
S&P_500			1	1	1	1	2			2	2		1			1		1	1	1			1							
Uncertainty						1		1	3	1		2	2		1	1	2	1				1								
Hushen_300				2	1			1	1		1		1	2		1	1			2	2			1						
Volume						1	1	1			1	3				1	1	2	2	1	1				1					
Gold						1	1	1	1		3			1	2			1	1	2	1	1								
KOSPI					1	2	1				1		1	2	1		1	2	2			1			1					
Commodity								1	1	3	1		1	1	1	1		1		2	1	1		1						
Euro										2	1	2	2	2			5				1	1								
P180DR								1			2	1		3	1		1				2	4		1						
Bond					1			1	1				1		2	2	1		1	2	2		1	1						
Skewness										1	1		1		1	2		1	3	2	3		1							
Oil							1			1			2	1		2		1				2	3	1	2					
P360DR												2							1	1	1	2	1	6	1	1				
Google																2		1			1	2	3	1	5	1				
Circulating_Supply										1											1		2		4	4	3	1		
EFFR												1								1			3	2		4	1	1	1	2
US_Elect_Gen																										3	5	4	2	2
US_Elect_Price																										1	5	3	6	1
Difficulty																								1	1	1	2	4	2	5
Ch_Elect_Gen																									1	1		2	7	5

Note: This table reports the number of importance rankings achieved by each feature in the 16 "rolling window" predictions. The variable importance is calculated based on the SHAP value for each input feature for the all-feature-XGB model. The importance of a feature is calculated by taking the average of the absolute SHAP of the test dataset. We then rank the features based on their “Mean(|SHAP|)” value. The higher the ranking, the more important the feature.

Table 10
Individual-cryptocurrency-level R^2 on the training and test samples for other machine learning methods.

	LASSO	Ridge	RF	GB	FFNN
Training (%)	0.002	2.830	6.782	16.327	10.322
Test (%)	-0.490	1.976	3.814	3.699	1.814

Note: This table reports the average R^2 of individual cryptocurrencies on the training and test samples for each of the machine learning methods, which are Least Absolute Shrinkage and Selection Operator (LASSO), Ridge, Radom Forest (RF), Gradient Boosting (GB), and Feed Forward Neural Network (FFNN). We calculate the average R^2 of the 16 "rolling window" models. The data used in models of this table is the same as that in models of Table 4. We use all the main 30 features to construct the models.

Table 11
Individual-cryptocurrency-level R^2 for different sub-samples.

	OLS All Features	OLS Without P1DR	OLS With Only P1DR	XGB All Features	XGB Without P1DR	XGB With Only P1DR
Panel A: Before and after COVID-19						
Pre-2020 (%)	1.434	-0.471	0.979	2.479	-0.338	3.044
Post-2020 (%)	2.516	0.380	1.898	5.315	2.242	5.609
Panel B: Different fear and uncertainty stage						
Low VIX (%)	2.260	-0.021	1.727	4.275	0.848	4.189
High VIX (%)	1.880	-0.054	1.342	3.772	0.986	4.373
Panel C: Up and down market						
Low S&P 500 (%)	0.058	-1.444	0.503	0.501	-2.712	2.244
High S&P 500 (%)	2.418	0.277	1.654	4.681	1.797	4.807

Note: This table shows the out-of-sample R^2 for machine-learning models in Table 4. We investigate whether these models are stable across time and market situations. Panel A lists the R^2 before and after the COVID-19 outbreak. Panel B lists the R^2 in high and low market fear and uncertainty. Specifically, we contrast the values of the VIX index with their historical medians. We describe the quarter as being in the low (high) fear stage if the index is lower (higher) than the historical median. In Panel C, we compare S&P 500 index values with their antecedent medians. One quarter is characterized as being in the up (down) stage if the S&P 500 index is higher (lower) than the historical median.

nonlinear one-day reversal effect on cryptocurrency returns. The 1-day lagged return is the most powerful feature for predicting returns, which is more potent than all other features combined. We also find a seven-day reversal effect when the previous seven-day return is negative. Third, small cryptocurrencies' returns are more predictable than large ones. This is mainly because small cryptocurrencies have a much stronger one-day reversal effect than large cryptocurrencies. Finally, a convex relationship between standard deviation and returns in volatility dynamics indicates that riskier cryptocurrencies have higher future returns.

Regarding variable importance, the following results are found. Macroeconomic variables, the OECD unemployment rate, OECD inflation, and OECD industrial production growth are important for forecasting cryptocurrency returns. In contrast, assets, including the S&P 500, Hushen 300, KOSPI, gold, and bond, are less effective for forecasting cryptocurrency returns. One of the significant results is that cryptocurrency returns correlate with stock returns and, to a lesser extent, gold, commodity, and bonds. While this is very interesting, it also points to lower diversification benefits from adding cryptocurrencies to a portfolio composed of stocks and bonds. Google searches, measuring the popularity of the cryptocurrencies or investor attention, were most effective for forecasting returns for larger cryptocurrencies. It points to the notion that smaller currencies are often neglected or do not get attention. In addition, the cost of mining cryptocurrencies, or production cost, represented by the three electricity-related variables, has weak predictive power. The above results indicate that cryptocurrencies behave like conventional assets than currency since variables such as lagged return, global economic climate, and investor attention can all predict cryptocurrency returns. In addition, as an asset, cryptocurrency is not weakly efficient.

Our paper has several insights into the trading and regulation of the cryptocurrency market. The ongoing political debate on regulating cryptocurrencies is a cause for concern. In addition, the possible use of cryptocurrencies to avoid sanctions and money laundering adds volatility to the market. However, interest in cryptocurrencies as financial assets remains strong with explosive growth in the DeFI, NFT, and monetizing Blockchain innovations (Yousaf et al., 2022). As hedge funds, institutional investors, and mutual funds look for cryptocurrency-related ETFs and assets, market participants continue to find ways to forecast cryptocurrency returns. The model used in this paper and the results suggest that machine learning can generate good cryptocurrency forecasts. This can provide insights to crypto investors, traders, and financial analysts for portfolio construction and risk management. Future studies with alternative features and higher frequency data can offer further insights.

CRediT authorship contribution statement

Yujun Liu: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Zhongfei Li**: Conceptualization, Writing – review & editing, Supervision, Funding acquisition. **Ramzi Nekhili**: Conceptualization, Writing – original draft, Writing – review & editing. **Jahangir Sultan**: Conceptualization, Writing – original draft, Writing – review & editing.

Declarations of interest

None.

Data availability

The authors do not have permission to share data.

References

Akyildirim, E., Goncu, A., Sensoy, A., 2021. Prediction of cryptocurrency returns using machine learning. Ann. Oper. Res. 297, 3–36.
Bouoiyour, J., Selmi, R., 2015. What does bitcoin look like. Ann. Econ. Financ. 16 (2), 449–492.
Bouri, E., Gupta, R., Lahiani, A., Shahbaz, M., 2018. Testing for asymmetric nonlinear short- and long-run relationships between bitcoin, aggregate commodity and gold prices. Resour. Policy 57, 224–235.
Bouri, E., Christou, C., Gupta, R., 2022. Forecasting returns of major cryptocurrencies: evidence from regime-switching factor models. Financ. Res. Lett. 49, 103193.
Breiman, L., 2001. Random forests. Mach. Learn 45 (1), 5–32.
Chen, T.Q., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, pp. 785–794.
Chowdhury, R., Rahman, M.A., Rahman, M.S., Mahdy, M.R.C., 2020. An approach to predict and forecast the price of constituents and index of cryptocurrency using machine learning. Phys. A: Stat. Mech. Appl. 551, 124569.
Ciaian, P., Rajcaniova, M., Kancs, 2015. The economics of BitCoin price formation. Appl. Econ. 48 (19), 1799–1815.
Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. Math. Control, Signals Syst. 2 (4), 303–314.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29 (5), 1189–1232.
Gu, S.H., Kelly, B., Xiu, D.C., 2020. Empirical asset pricing via machine learning. Rev. Financ. Stud. 33 (5), 2223–2273.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second ed. Springer, New York.
Hayes, A.S., 2017. Cryptocurrency value formation: an empirical study leading to a cost of production model for valuing bitcoin. Telemat. Inform. 34 (7), 1308–1321.
Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12 (1), 55–67.
Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Netw. 2 (5), 359–366.
Indera, N.I., Yassin, I.M., Zabidi, A., Rizman, Z.I., 2018. Nonlinear autoregressive with exogeneous input (NARX) bitcoin price prediction model using PSO-optimized parameters and moving average technical indicators. J. Fundam. Appl. Sci. 9 (3S), 791–818.
Jiang, K., Zeng, L., Song, J., Liu, Y., 2022. Forecasting value-at-risk of cryptocurrencies using the time-varying mixture-accelerating generalized autoregressive score model. Res. Int. Bus. Financ. 61, 101634.
Katz, M., Shapiro, C., 1985. Network externalities, competition, and compatibility. Am. Econ. Rev. 75 (3), 424–440.
Kristoufek, L., 2015. What are the main drivers of the bitcoin price? Evidence from wavelet coherence analysis. PLOS One 10 (4), e0123923.
Lahmiri, S., Bekiros, S., 2019. Cryptocurrency forecasting with deep learning chaotic neural networks. Chaos Solitons Fractals 118, 35–40.
Li, X., Wang, C.A., 2017. The technology and economic determinants of cryptocurrency exchange rates: the case of Bitcoin. Decis. Support Syst. 95, 49–60.
Li, X.Y., Gan, K., Zhou, Q., 2022. Dynamic volatility connectedness among cryptocurrencies and China's financial assets in standard times and during the COVID-19 pandemic. Financ. Res. Lett., 103476
Liu, M.X., Li, G.W., Li, J.P., Zhu, X.Q., Yao, Y.H., 2021. Forecasting the price of Bitcoin using deep learning. Financ. Res. Lett. 40, 101755.
Liu, Y., Tsyvinski, A., 2021. Risks and returns of cryptocurrency. Rev. Financ. Stud. 34 (6), 2689–2727.
Lundberg, S., Lee, S.-I., 2017. A unified approach to interpreting model predictions. ArXiv.
McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. 5 (4), 115–133.
Orte, F., Mira, J., Sánchez, M.J., Solana, P., 2023. A random forest-based model for crypto asset forecasts in futures markets with out-of-sample prediction. Res. Int. Bus. Financ. 64, 101829.
Polasik, M., Piotrowska, A.I., Wisniewski, T.P., Kotkowski, R., Lightfoot, G., 2015. Price fluctuations and the use of bitcoin: an empirical inquiry. Int. J. Electron. Commer. 20 (1), 9–49.
Ren, Y.S., Ma, C.Q., Kong, X.L., Baltas, K., Zureigat, Q., 2022. Past, present, and future of the application of machine learning in cryptocurrency research. Res. Int. Bus. Financ 63, 101799.
Schilling, L., Uhlig, H., 2019. Some simple bitcoin economics. J. Monet. Econ. 106, 16–26.
Sockin, M., Xiong, W., 2020. A model of cryptocurrencies. Working Paper. Available at NBER 26816.
Sun, X.L., Liu, M.X., Sima, Z., 2020. A novel cryptocurrency price trend forecasting model based on Light GBM. Financ. Res. Lett. 32, 101084.
Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. J. R. Stat. Soc.: Ser. B (Methodol.) 58 (1), 267–288.
Wang, Y.Q., Wang, C.F., Sensoy, A., Yao, S.Y., Cheng, F.Y., 2022. Can investors' informed trading predict cryptocurrency returns? Evidence from machine learning. Res. Int. Bus. Financ. 62, 101683.
Woo, D., Gordon, I., Iaralov, V., 2013. Bitcoin: a first assessment. FX and Rates, Global. Bank of America, Merrill Lynch.
Yae, J., Tian, G.Z., 2022. Out-of-sample forecasting of cryptocurrency returns: a comprehensive comparison of predictors and algorithms. Phys. A: Stat. Mech. Appl. 598, 127379.
Yousaf, I., Nekhili, R., Gubareva, M., 2022. Linkages between DeFi assets and conventional currencies: evidence from the COVID-19 pandemic. Int. Rev. Financ. Anal. 81, 102082.
Zaremba, A., Bilgin, M.H., Long, H., Mercik, A., Szczygielski, J.J., 2021. Up or down? Short-term reversal, momentum, and liquidity effects in cryptocurrency markets. Int. Rev. Financ. Anal. 78, 101908.
Zhang, Z.R., Dai, H.-N., Zhou, J.H., Mondal, S.K., García, M.M., Wang, H., 2021. Forecasting cryptocurrency price using convolutional neural networks with weighted and attentive memory channels. Expert Syst. Appl. 183, 115378.
Zhu, Y.C., Dickinson, D., Li, J.J., 2017. Analysis on the influence factors of Bitcoin's price based on VEC model. Financ. Innov. 3, 1.