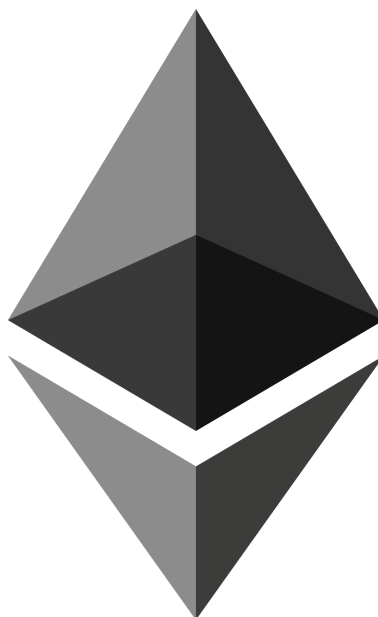# Predicting Changes in the Price of Ethereum

Alexander Nikolov, Caleb Shack, Brandon Cohen, Aaron Finkelstein

December 5, 2021

Introduction

Developments in the digital world have reconstructed the contexts in which humans perceive value, particularly with the creation of cryptocurrency. Cryptocurrency has become a frequently discussed topic in economics due to its status as the digital alternative to central bank-controlled fiat money. As more businesses begin to accept various cryptocurrencies for payment, its entryway into other industries is inevitable.

Cryptocurrencies use a distributed ledger system shared among nodes of a blockchain network to secure its transactions. This peer-to-peer network validates and clusters information on each transaction in components called "blocks." Each block is continuously closed and linked to previously filled blocks, creating a chain that acts as a public database. This design is inherently irreversible in that each block cannot be re-recorded after its close, and decentralized, in that all users of the system maintain equal ownership of it. These two qualities make blockchain hacking exceedingly difficult as new copies of the chain are created for every new block added, meaning any chain containing altered information may easily be distinguished among the identical group and be deemed invalid. Due to the security, privacy and convenience provided by blockchain technology, cryptocurrency has become a popular investment choice for businesses and consumers alike.

The most popularly exchanged and highly valued cryptocurrency is Bitcoin (BTC), based on the blockchain network that was launched in 2009 in the wake of the global financial crisis. BTC's blockchain purely exists as a means to exchange digital currency. Its massive and timely success signaled to creators that its underlying innovations could be utilized for other purposes. Such utilities are found in the Ethereum blockchain, created not only to maintain a digital payment network, but also to allow applications to be built, deployed and run without third-party interference. Its currency, Ether (ETH), powers online contractual agreement applications called smart contracts for a respective transaction cost, known as the gas price, paid to the miners that validate each transaction in the system. Ethereum rewards miners who append each block on the network and gives marginal rewards to those who validate a block slower than competitors, called uncle blocks. This issuance works differently for Ethereum than Bitcoin, which is issued diminishingly with a cap at 21 million BTC in contrast to ETH's infinite ceiling. However, an improvement proposal was recently implemented that increases the base gas fee to burn a portion of every transaction, known as EIP-1559. With enough network activity, ETH could become a deflationary currency despite its positive rate of emissions today.

Since August of 2021, a substantially large amount of ETH burned has been attributed to the sale of non-fungible tokens (NFTs), which are unique, authenticated digital files stored on the Ethereum blockchain. Such virtual assets are bought with ETH in smart contracts and change in value based on market demand. Other applications on Ethereum support the creation of wallets which are virtual containers for users to hold cryptocurrency and tokens, in addition to art marketplaces, decentralized finance platforms and social networks. These characteristics have garnered Ethereum significant interest since its inception, helping to make ETH hold the second highest market cap behind BTC.

Research Question

With Ethereum's increase in attention and common knowledge in recent years, much effort has been spent to predict and analyze future prices of ETH. Its price is especially difficult to forecast due to high volatility and lack of association with classical economic indicators, in addition to having its own blockchain-unique characteristics relating to gas, burning and issuance. This study attempts to determine the extent to which such external and internal factors influence changes in the price of ETH. External factors refer to market activity, such as the S&P 500, price of gold and silver, yields on the 1, 5, and 10-year treasuries, and industry-based indices, as well as the price of competitive coins, such as BTC and Litecoin (LTC). Internal factors evaluated include ETH created, number of transactions, and amount of gas used given in daily snapshots.

First, we will standardize our datasets into daily snapshots to obtain specific observations to use in our models. Then, we will run random forests to determine the significance of each of the collected fields' impact on the price of ETH. Once we have an understanding of how these variables explain ETH price in the past,

we will attempt to predict its future price through further random forest models and create an algorithm that maximizes profits through buying, selling and holding based on its own prediction. Additionally, we will conduct a separate analysis to discover if Twitter data has a significant influence on ETH price changes.

Our findings suggest that the price change of ETH fluctuates similarly to the price of BTC and LTC, the first two cryptocurrencies ever deployed, and is also impacted by the energy and real estate sectors. When considering internal Ethereum blockchain-based variables, transaction count and percentage of gas used per day were found to be significant indicators of price movement. We then shift our attention to forecasting *tomorrow's* change in the price of Ethereum. We find that changes in the performance of the real estate and bond market preempts changes in the price of Ethereum. Our model successfully capitalizes on this information to develop a profitable trading strategy. The volatility of Twitter data on the sentiments of tweets containing references to ETH and NFTs is noticeably visually associated with changes in the price of ETH. However, we do not have enough evidence to indicate a causal relationship between the two variables.

Literature Review

Although interest in cryptocurrencies and their future applications has only become widespread in recent years, there are a number of studies that have effectively utilized machine learning methods to predict the price movements of such tokens. The large majority of each study's target variable is the price of the respective cryptocurrency under examination, while the feature variables used can be categorized as either external or internal factors. Generally, the external factors include macroeconomic data, such as S&P 500 index, Dow Jones 30 index, Crude Oil, and Gold, among others, as well as global currency ratios, such as EUR/USD, JPY/USD and GBP/USD. The internal factors refer to blockchain information, such as trading volume, average block size, hash rate of block, difficulty of block, and miners' revenue, among others. In this section, the chronology of each paper relevant to our studies is summarized, specifically in regard to the regressors used and concluding remarks.

During the initial cryptocurrency market boom in 2017, an empirical study was conducted by Jang and Lee to determine how to model and predict the logarithmic price of BTC. They collected macroeconomic data, BTC blockchain information, and global currency ratios, to create 10, 10, and 5 regressors respectively for their model. The researchers first used linear regression to estimate the price, but found issues with multicollinearity and non-normal distributions of logarithmic volatility in BTC. After determining that a linear analysis may not be appropriate for time-series data, the authors employed a Bayesian Neural Network (BNN) to forecast BTC prices. Artificial Neural Networks, although beyond the scope of this course, provide a nonlinear method to analyze complex time series data. The Bayesian version of this model intuitively uses Bayesian theory, in that it maximizes the value of the posterior, to perform non-linear regression. This version of the model produced the lowest RMSE and MAPE (.0069 and .0180, respectively), compared to the linear and SVM methods. Ultimately, the researchers concluded that all models could predict price movements to an extent, but BNN could best describe the volatility and price of Bitcoin. The analysis did not touch upon which input factors were more or less influential on the response of BTC price and volatility.

A more recent study, published in September 2021 by Hamayel and Owada, expands the scope of currency analysis by attempting to predict the prices of BTC, ETH and LTC. The researchers strictly collected daily price data on all three cryptocurrencies between October 2020 and June 2021, including token prices at day-open, day-close, and the high and lows achieved for each day. Given the fact that no other input factors were taken into account, this study aimed to create a prediction and analysis-based model on the technical movements in price using three neural network-based machine learning models: Long Short-Term Memory (LSTM), Bidirectional LSTM (bi-LSTM), and Gated Recurrent Unit (GRU). The models were trained with 80% of the dataset, from January 2018 to October 2020, and tested on the remaining 20%, from October 2020 through June 2021. The authors concluded that the GRU outperformed the other algorithms and can be considered efficient and reliable for predictions of cryptocurrency prices. Although this study only considers price data, Hamayel and Owada mention their interest in investigating how tweets can have an effect on the price and trading volume of various cryptocurrencies, specifically using Natural Language Processing and sentiment analysis.

Sebastiao and Godinho's study from January 2021 provides direct application to this paper through its forecast of BTC, ETC, and LTC prices based on both trading and blockchain information. The researchers performed regression and classification analysis using linear and random forest models respectively, in addition to a support vector machine (SVM) model – a supervised algorithm for non-linear applications. The regression analysis focuses on the model's predictive power for the prices of each cryptocurrency, while the classification analysis produces a success rate of determining a buy or sell rating for each token. The findings conclude that random forests give a high ETH success rate (60%) for classification, while linear regression gives a slightly lower ETH success rate (54.46%) for regression. Although the results for SVM are similar to these methods, this finding demonstrates that linear and random forests can be used effectively to predict movements in cryptocurrency. Additionally, the study evaluates specific trading strategies that could be used to beat the market based on the model predictions.

The most relevant study to this paper is based on the recent work of Kim, Bock, and Lee from November 2021. The study focuses on forecasting Ethereum prices based on 15 macroeconomic factors, 4 Ethereum-specific blockchain variables, as well as factors based on the blockchain information of other tokens, including 9 BTC variables, 9 LiteCoin (LTC) variables, and 9 Dashcoin (DASH) variables. Interpreting each of the groupings above as a single regressor, they ran both a stepwise analysis and secondary analysis, to thoroughly understand which factors are affecting ETH price. Additionally, the machine learning models ANN and SVM were both used in every model pairing. Ultimately, they found that including Ethereum information with macroeconomic factors significantly improved the models' predictive performance for ETH prices. Additionally, they deduced that blockchain information on BTC was significantly related to ETH prices, while other currencies' blockchain information was not statistically relevant.

As mentioned, the cryptocurrency industry is an extremely new space, and although there are some studies that have attempted to predict the price movements of these coins, there are many facets of this question to still uncover. The primary goal of this study is to use macroeconomic factors and blockchain information similarly to how Kim, Bock, and Lee did in their recently published research. However, the methodology differs in that random forest classification models will be utilized to predict ETH price changes, similar to how Sebastiao and Godinho conducted their model analysis, and a trading strategy will be developed from these analyses. Furthermore, we will fulfill Hamayel and Owada's goal of completing NLP and sentiment analysis on Twitter data to analyze its impact on ETH price.

Data

To predict changes in the price of ETH, we use financial market data as well as data pertaining directly to activity on the Ethereum network. All financial data was downloaded from *MarketWatch.com*, which specializes in providing historic and real-time data on a broad category of markets and asset classes. This data includes daily price levels for various cryptocurrencies, stock indices, fuels, and rare metals. To gauge current economic sentiment, we also collect data measuring the volatility of the S&P 500 as well as yields on short- and long-term treasury bonds. A full description of the financial data collected can be found in Table 1.

Table 1: Financial Market Data

Name	Description
ETH	Price level of Ethereum
BTC	Price level of Bitcoin
LTC	Price level of Litecoin
SPX	Index tracking performance of S&P500
CBOEVolatility	Index measuring expectations of volatility in the stock market
SP500Energy	Index tracking performance of major US energy companies
DJRetail	Index tracking performance of major US retailers
DJRealEstate	Index tracking performance of major US real estate companies

Name	Description
ShanghaiComp	Index tracking performance of the Shanghai Stock Exchange
CrudeOil	Price level of crude oil
NaturalGas	Price level of natural gas
Gold	Price level of gold
Silver	Price level of silver
TY1M	Yield on 1 Month US Treasury Bills
TY1Y	Yield on 1 Year US Treasury Bills
TY10Y	Yield on 10 Year US Treasury Notes
TY30Y	Yield on 30 Year US Treasury Bonds

To collect data pertaining to activity on the Ethereum network, we scraped data from each block via https://etherscan.io/blocks and aggregated it into daily snapshots. These snapshots include data on daily transaction counts, energy used to validate transactions, and amount of Ether created or burned. A full description of the data collected on the Ethereum network can be found in Table 2.

Table 2: Ethereum Network Data

Name	Description
txn_count	Number of transactions on Ethereum network per day
gas_used	Amount of Ether paid to validators on the network per day
gas_limit	Maximum amount of Ether that could be paid to validators on the network per day
eth_created	Amount of Ethereum created or burned per day
percent_gas_used	Amount of computational energy used divided by maximum amount available

Note that all market data was collected at market close, or 4PM Eastern time. The price of each cryptocurrency was collected at 5PM Eastern time, and daily snapshots aggregate data from the 24-hour period leading up to 5PM EST. Thus, for any given day, our data represents the most recent data available as of 5PM.

To collect Twitter data, 12,296 tweets from November 3rd, 2021 to December 3rd, 2021 were extracted using two separate Twitter Developer Account with filters applied to solely obtain tweets written in English from "verified" accounts that contain the word "Ethereum" or "ETH". Another 9,900 tweets that contain "NFT" were extracted from July 6th, 2021 to August 10th, 2021 using the same other parameters. Strictly "verified" accounts were utilized to streamline the tweet collection process, while still allowing for an observable effect on the ETH Twitter community through such accounts' large follower base. The ETH tweet dates were chosen at random for extraction, while the tweet dates for NFTs were chosen to correlate to a large surge in ETH price. Ethereum price data was extracted from *ethereumprice.org*, a reputable source that allows for free downloads of data on the price of ETH and associated cryptocurrencies.

Explanatory Models

Methodology

In this section, we attempt to predict how the price of ETH changes over the course of a given day by using the most recent data from financial markets and the Ethereum network. To do this, we first represent all data as a percentage change from the previous day. Then, we use random forests to see whether changes in market conditions and changes in Ethereum blockchain usage over the past day can accurately predict how the price of ETH has changed over that period.

The random forest algorithm creates each decision tree by drawing n observations with replacement from the dataset. As a result, certain observations are excluded from each tree, known as out-of-bag observations.

To reduce the correlation between decision trees, only a fraction of the regressors are considered for each split. This increases the bias of our results but lowers the variance by creating trees that are less sensitive to a small number of strong predictors. We find the optimal number of regressors that should be considered for each split by minimizing the average out-of-bag test error.

Financial Data Model

First, we attempt to predict the change in the price of ETH using only financial market data. As shown in Figure 1 below, the price of competing cryptocurrencies as well as the performance of the energy and real estate markets are the most important variables for predicting the price of ETH. Using these variables and other market indicators, we are able to explain up to 65 percent of the variance in the price changes of ETH.

Figure 1: Financial Data RF Results
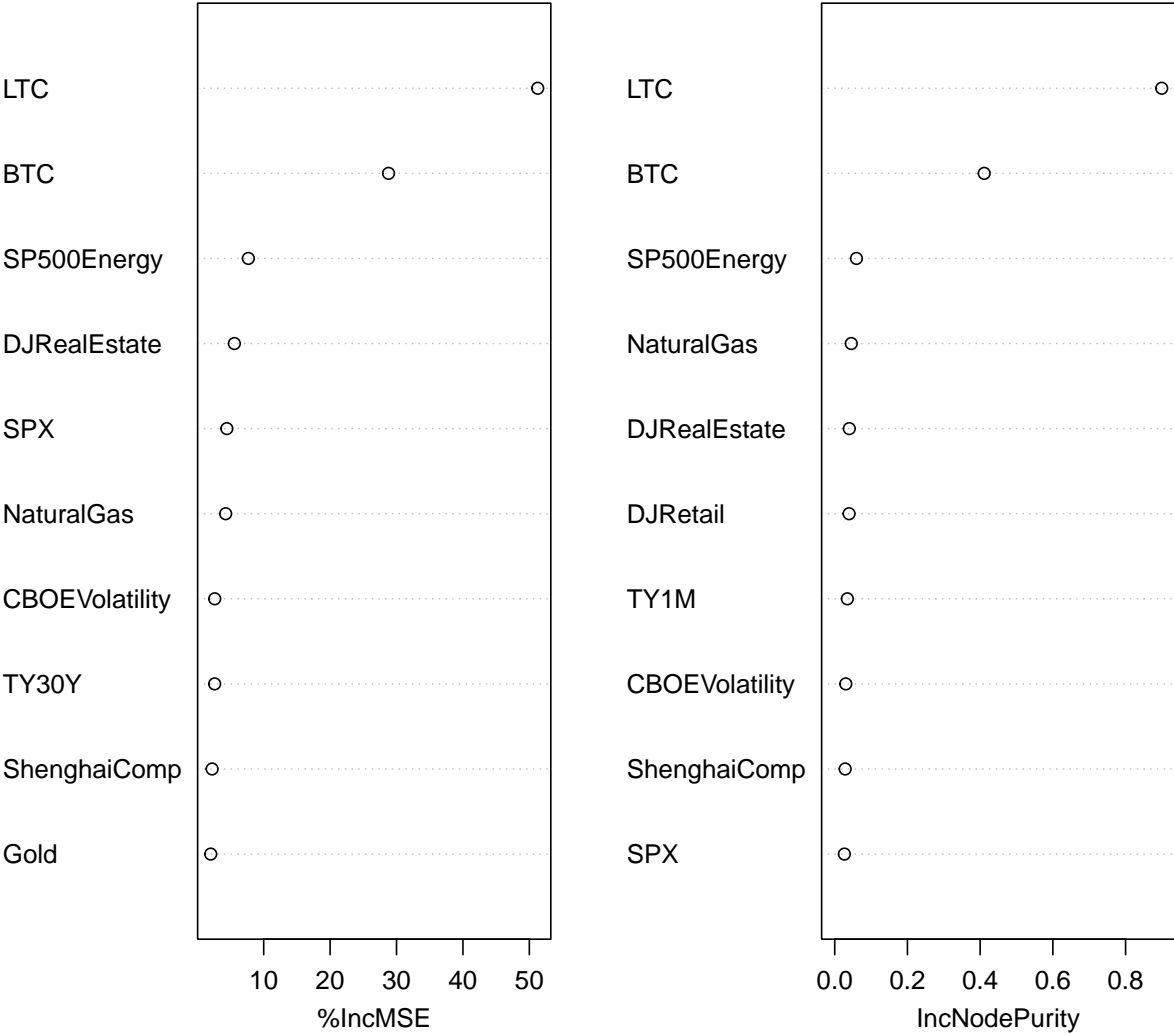
To understand how these variables impact the price of ETH, below we have included a sample decision tree

generated by the CART algorithm. Most notably, we see that the price of bitcoin and litecoin are strongly correlated with the price of ETH. BTC and LTC were the first two cryptocurrencies to enter into existence, BTC being deployed in 2009 and LTC in 2011, so it is no surprise that they represent the cryptocurrency market as a whole.

## Figure 2: Financial Data Sample Decision Tree

LTC < −0.00265679

BTC < −0.0897258

LTC < 0.117444

SP500Energy < −0.0210522

LTC < −0.0441007

LTC < 0.0293361

DJRealEstate < −0.00345972

−0.20680   −0.12140

−0.05578   −0.01462

NaturalGas < 0.0527393

BTC < 0.0114087

0.20510

SPX < 0.00370888

0.01701   0.08241

0.02467   0.06146

0.04476   0.15710
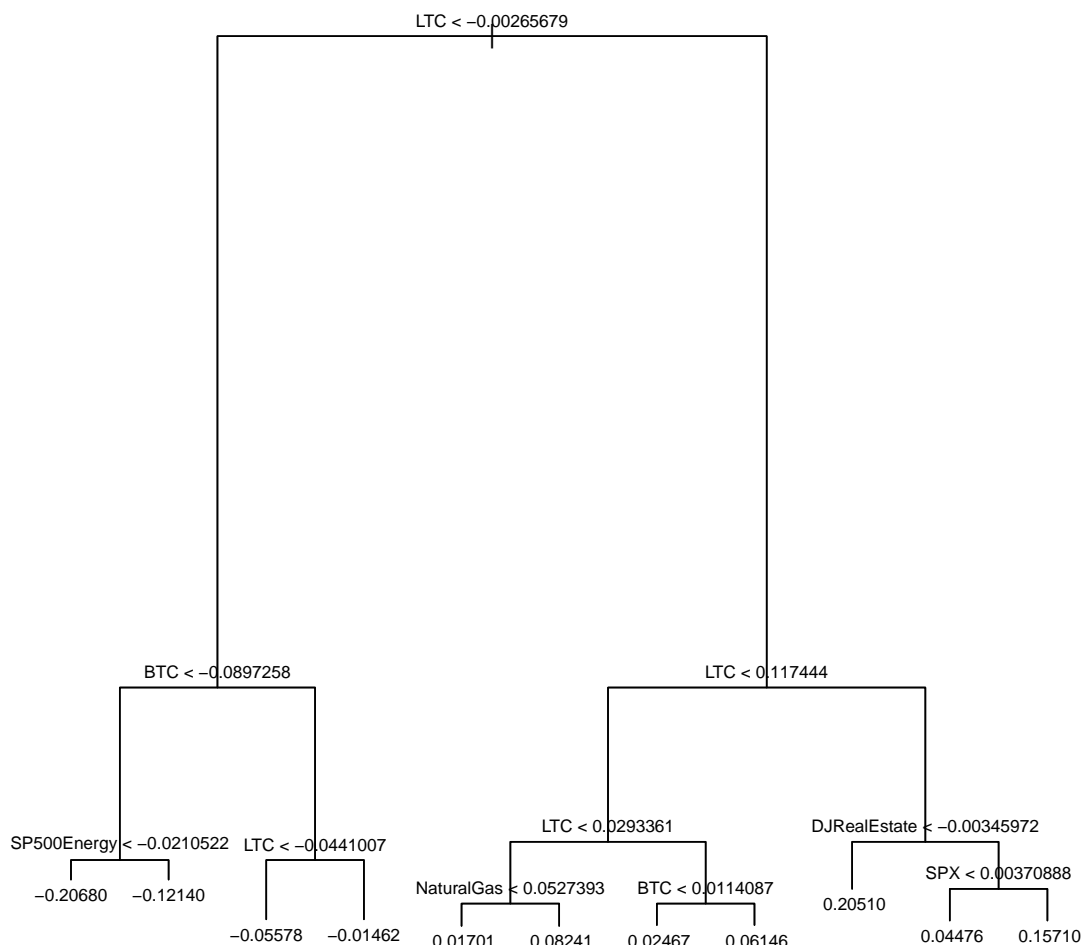
As we examine the tree further, we can see that the energy and natural gas markets also have a positive correlation with price movement in ETH. Cryptocurrency uses a lot of energy to keep alive, especially on networks that use a Proof of Work consensus mechanism that requires nodes on the network to compete with each other to solve an energy-consuming hashing problem. Ethereum is one of these networks; however, a future update to Ethereum will change the consensus mechanism to a Proof of Stake protocol where validators are randomly chosen to mine each block rather than competing for it. In theory, this will drastically reduce the energy used to keep the network alive, so, if updated successfully, ETH will become a more cost effective alternative to BTC and other cryptocurrencies when energy costs are high.

Lastly, one can see that when the real estate market is down, ETH performs well. We suspect this is due to ETH being an alternative investment for those in the real estate industry. Also, a more speculative point to make is that real estate is making its way to the "metaverse" in the form of digital plots of land for various video games, such as Sandbox and Decentraland, which may take away from money otherwise spent in real estate since these plots are purchased using ETH.

Ethereum Network Data Model

Next, we attempt to predict price changes in ETH using only internal data from the Ethereum network. In order to reduce the volatility of these data points, we constructed rolling 3-day and 5-day averages, which have been labeled accordingly. As Figure 3 below demonstrates, gas usage and transaction counts were important indicators for predicting ETH's change in price.

Figure 3: Ethereum Data RF Results
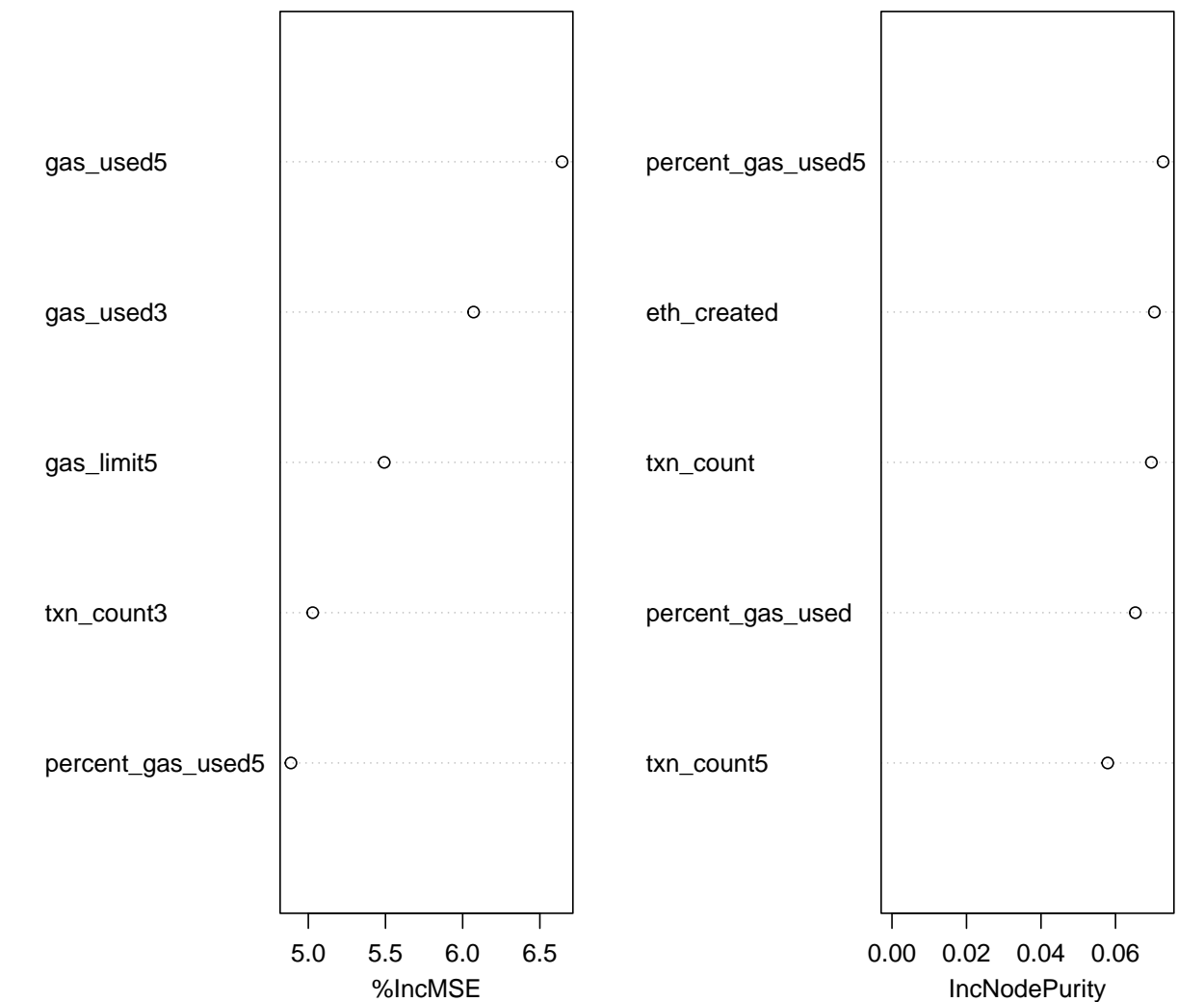
The purely internal model turned out to be not as helpful as we originally thought. As we can see from the sample decision tree below, the model fits noise when txn_count < 0.072818 since it is very rare that daily change in ETH is greater than 44%. However, the model does help us pinpoint important fields for us to consider in the combined internal, external model: percent_gas_used5 and txn_count.

When the percentage of gas used increases, the transactions on the network require more computing power and thus it will cost more additional money to complete a transaction. This is one of the major pitfalls of Ethereum right now: it's too expensive to conduct low value transactions. For example, sending $5 of ETH to your friend might cost you another $80 just to pay the miners to validate your transaction. Thus, when the percentage of gas used increases, it makes sense that people would want to sell their ETH for cheaper alternatives or to get out of the market entirely.

Figure 4: Ethereum Data Sample Decision Tree
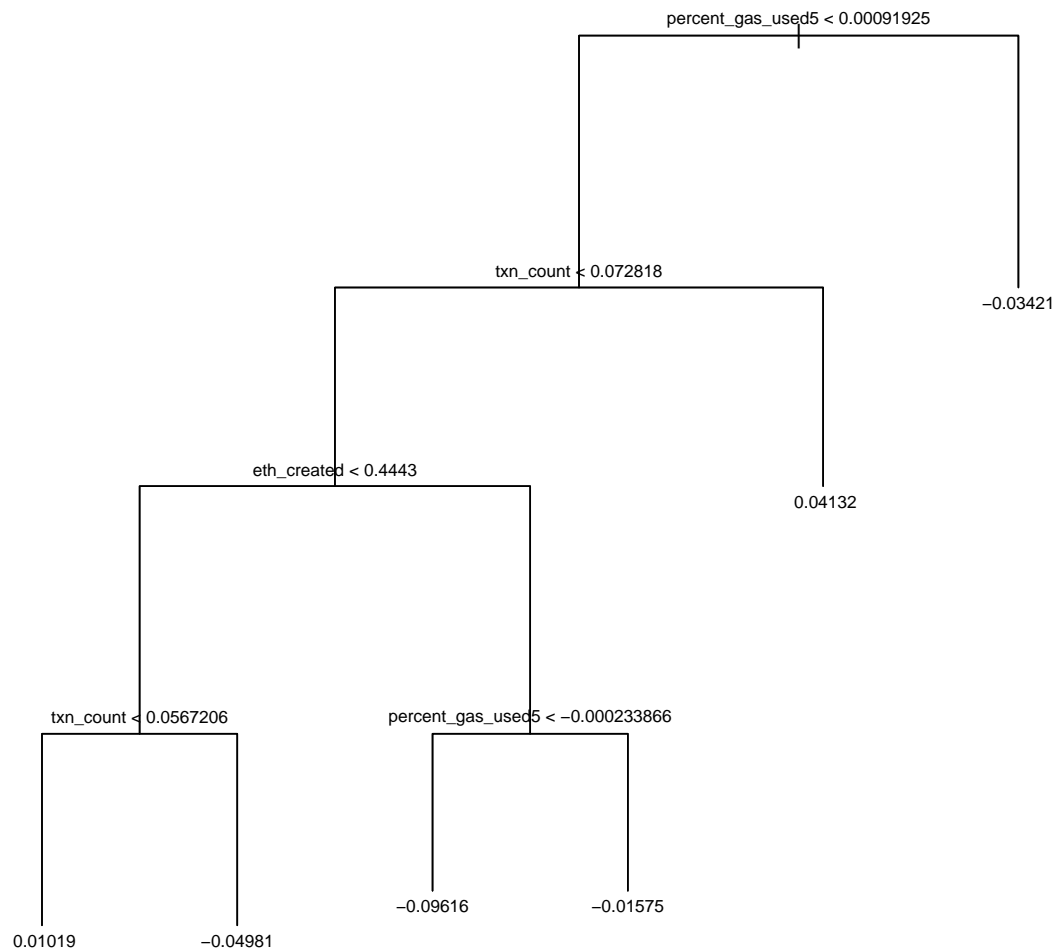
Lastly, when network transactions increase, we can see a very obvious increase in the price and vice versa when network activity decreases. Figure 5 below demonstrates this well.
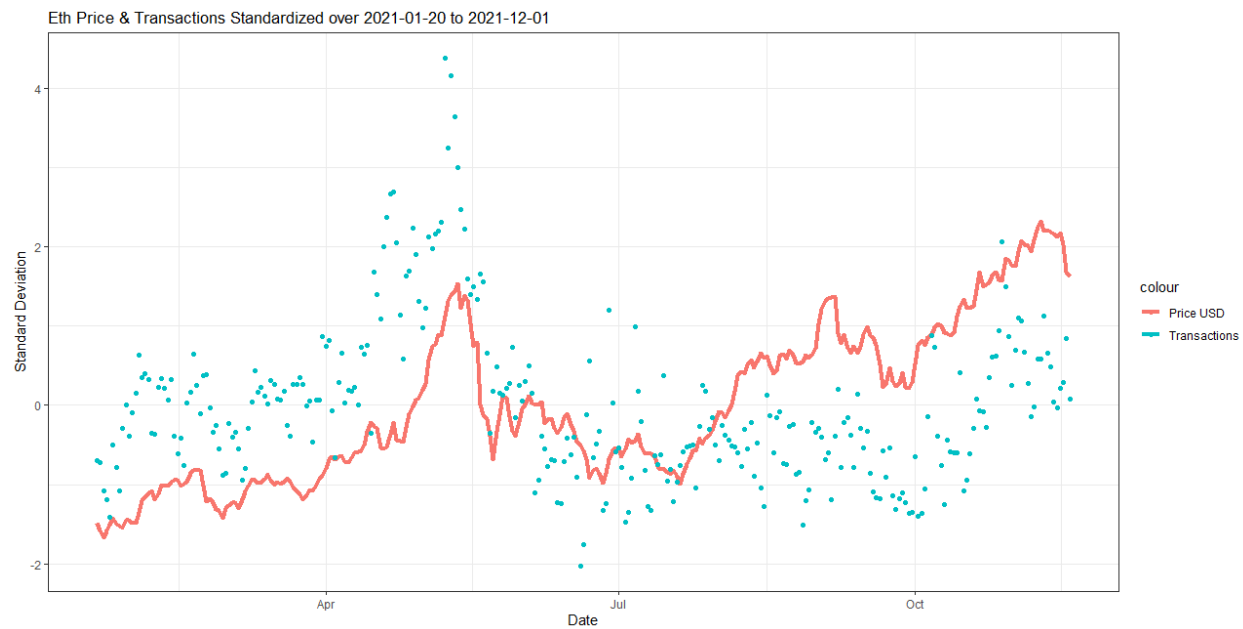

Figure 1: Figure 5: Daily Transactions and Price of Ethereum

Combined Model

Finally, we use both datasets combined to predict changes in the price of Ethereum. Again, the performance of competing cryptocurrencies and the performance of the energy sector are important for predicting changes in the price of ethereum. However, we also see that the amount of gas used and the transaction count although less influential, are still important in explaining ETH prices.

Figure 6: Combined RF Results
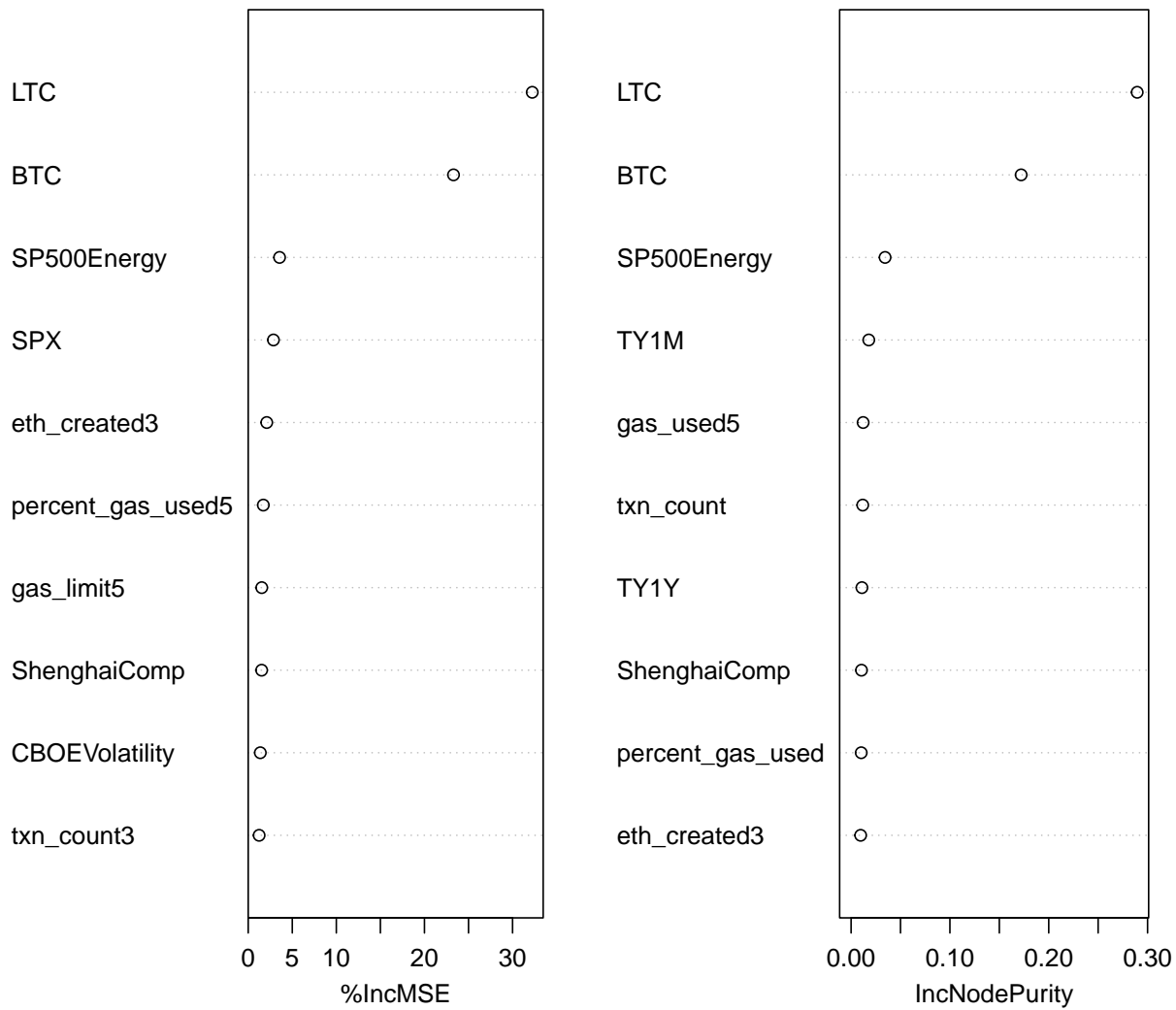
Predicting Tomorrow's Ethereum Price

Methodology

In the previous section, we found that changes in the price of other cryptocurrencies and changes in the performance of the energy and real estate markets were useful variables for explaining changes in the price of ETH. While these models are useful for explaining what drives the price of ETH, they provide little guidance on predicting ETH's future price. In this section, we turn out attention towards forecasting *tomorrow's* ETH price.

To do this, we construct a new Bernoulli variable 'Buy' which evaluates to one if the price of ETH increases by more than one percent the following day. For example, if ETH is trading at $1,000 Monday at 5PM and increases to $1,020 by Tuesday at 5PM, this indicates we should have bought ETH on Monday. For our Monday observation, 'Buy' evaluates to 1.

Using gradient boosting, we attempt to classify each day as 'Buy' or 'Not Buy' using the most recent data from financial markets and the Ethereum network. We use data from January 2021 through September 2021 in order to train our gradient boosting algorithm, and then test the model on data from October through November. In this sense, we can imagine ourselves as investors sitting at the end of September, looking to create a model using past historical data that we can use to forecast prices moving forward.

To create our gradient boosting algorithm, we must optimize over three separate tuning parameters: learning rate, individual tree size, and total number of sequential trees. To do this, we select values for the learning rate and tree size and find the number of trees that minimizes the test MSE of our model. We repeat this process for many possible learning rates and tree sizes and choose the parameters that result in the lowest possible test MSE. Finally, we use these optimal parameters and train the model with the entire set of data from January through September, giving us our final model.

Results

Using the model developed with the methodology described above, we are able to predict the change in ETH's price tomorrow using market data from today. To evaluate the performance of our model, we define the following trading strategy:

1) At market close (5PM), feed the model data from the past 24 hours. This includes financial market data and a summary of daily activity over the Ethereum network.

2) The model will return the probability that today is a 'Buy' day. If the probability is greater than 50 percent, purchase $100 worth of ETH at its 5PM price, and sell your holdings tomorrow at 5PM. If the probability is less than 50 percent, do nothing.

3) Repeat the process tomorrow at 5PM.

Using this procedure, we calculate our earnings and losses for each day. Over the entirety of October and November, this trading strategy results in a profit of $26. The following figure illustrates the decisions of the model over the course of October and November (note that market data from China was not available during the beginning of October, hence the starting date of October 11). Green arrows indicate the beginning of a 'Buy' period, and red arrows indicate the beginning of 'Hold' periods, where the model recommends not trading.
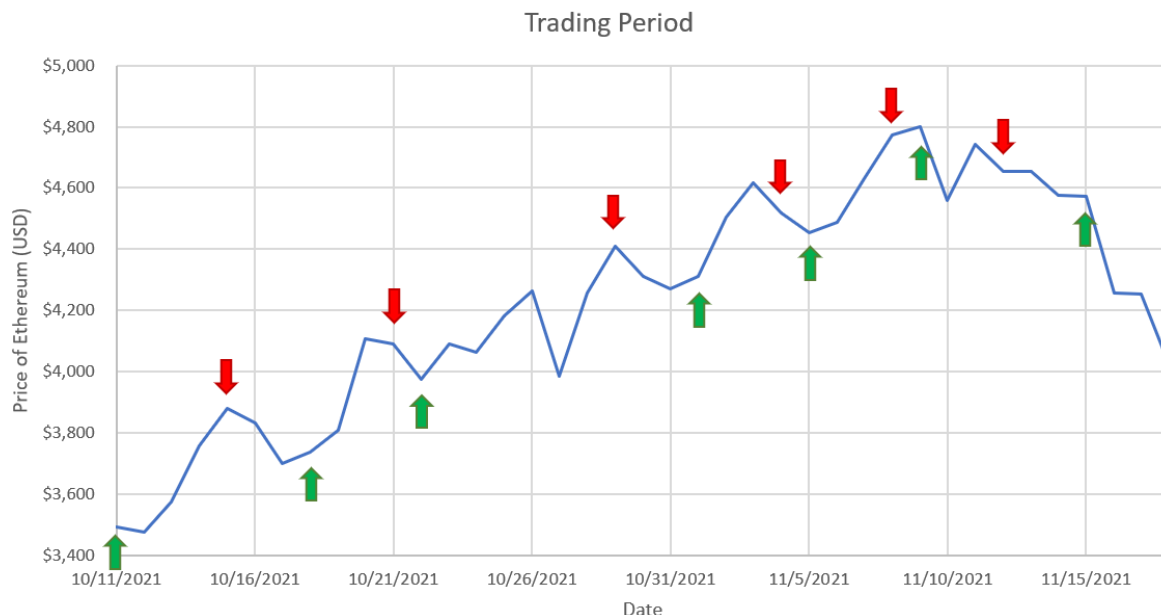
Figure 2: Figure 7: Decisions of Trading Algorithm

As we can see, the model performed well over the course of October when ETH experienced fairly steady gains. Over this period, the model successfully sent 'Buy' signals when the price was rising and accurately forecasted price declines. However, the model performed poorly towards the end of November, where markets were generally troubled by worrisome job reports. Also in November, the Securities and Exchange Commission rejected a proposal for a BTC exchange traded fund, causing falls in the price of many cryptocurrencies. With that said, given our strategy of only investing $100 each day, the losses in November were limited, and our model was able to deliver a positive net return of $26.

Finally, to understand what factors were driving the the trading decisions of our model, we plot the relative importance of each variable to the gradient boosting model (shown in Figure 8 below). We find that that the performance of the real estate market as well as changes in long-term yields were the most important variables driving the decisions of our trading algorithm. This provides further support for the idea that ETH serves as an alternative investment when traditional markets such as real estate are performing poorly. As for interest rates, large proponents for cryptocurrency, such as Elon Musk, have suggested that historically low interest rates set by the Federal Reserve Bank have driven demand for cryptocurrency. In a low-interest rate environment, investors are likely to seek out higher-risk securities like cryptocurrency that offer a higher yield. However, it appears there is a short delay between these changing market conditions and changes in the price of ETH, and our model appears to be capitalizing on this delay.

Figure 8: Relative Importance of Variables in Trading Algorithm
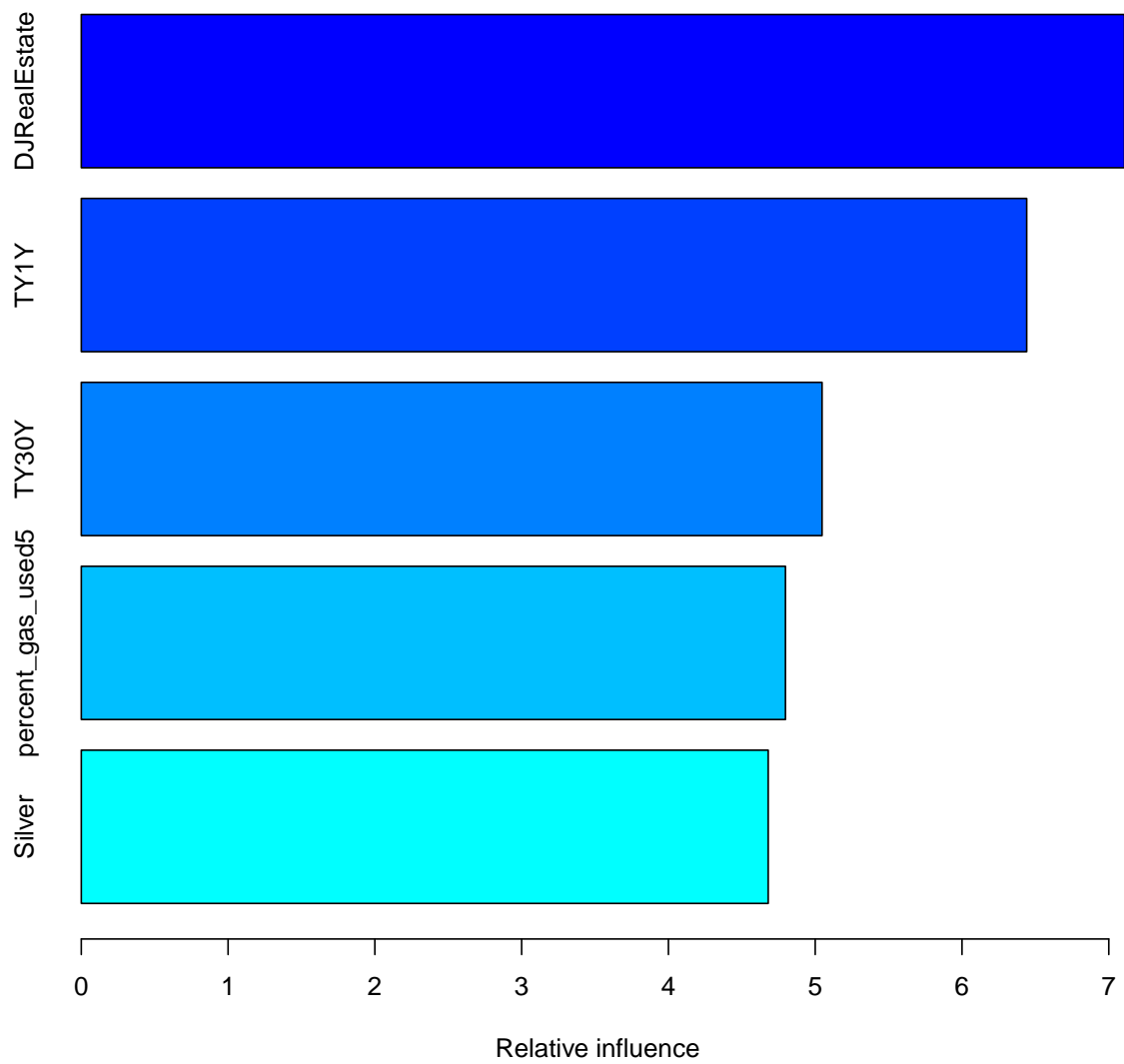
Twitter Data Analysis

We conducted an additional analysis for our report to understand how data on Twitter is correlated with the price of ETH, and if any sensible conclusions may be determined from these results.

In order to find the words most associated with ETH, we created a corpus on the textual data contained in each tweet and a data-frame matrix as the foundation for our frequency plot. We find that, barring the word "ethereum", the most common words in the tweets include "crypto", "bitcoin", and "nft", as well as more reactionary words, such as "will", "just", and "now". As observable from conclusions drawn with the preceding models, changes in BTC price have an effect on the price of ETH as supported by this textual analysis. However, the meaning of the remaining terms have yet to be explored in this analysis and the previous literature reviewed.

Figure 9: Words Frequently Associated with ETH
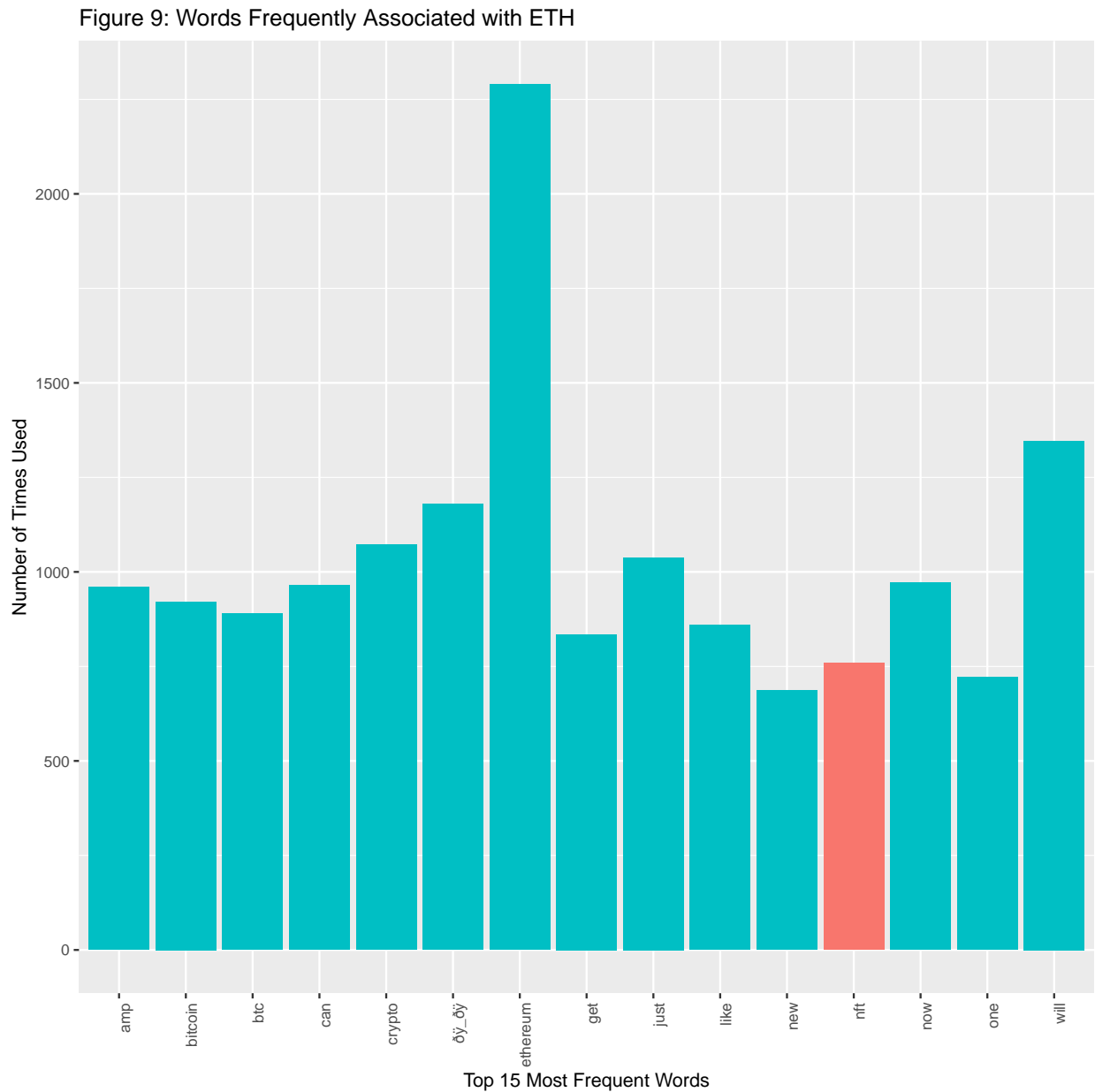
To understand the results above, further analysis was conducted using sentiment data. Each word within the tweets was assigned a "positive" or "negative" connotation from a sentiment dictionary included in R. After aggregating the scores for all the words in each tweet and adding the scores, the daily sentiment data on tweets including "ETH" was calculated. The results most clearly displayed that these tweets contained more positive sentiments than negative sentiments throughout the data collection period, by an average margin of approximately 50%. This feature could potentially be attributed to the fact that only "verified" users' tweets were collected for analysis. Many "verified" accounts are associated with businesses and are therefore more likely to be promotional in nature than degrading. Also, the considerable hype surrounding ETH and other cryptocurrencies at the time of collection would presumably create more positivity than negativity in consumers' sentiments in general.

Figure 10: Daily ETH Tweet Sentiment and Price (November 2021)
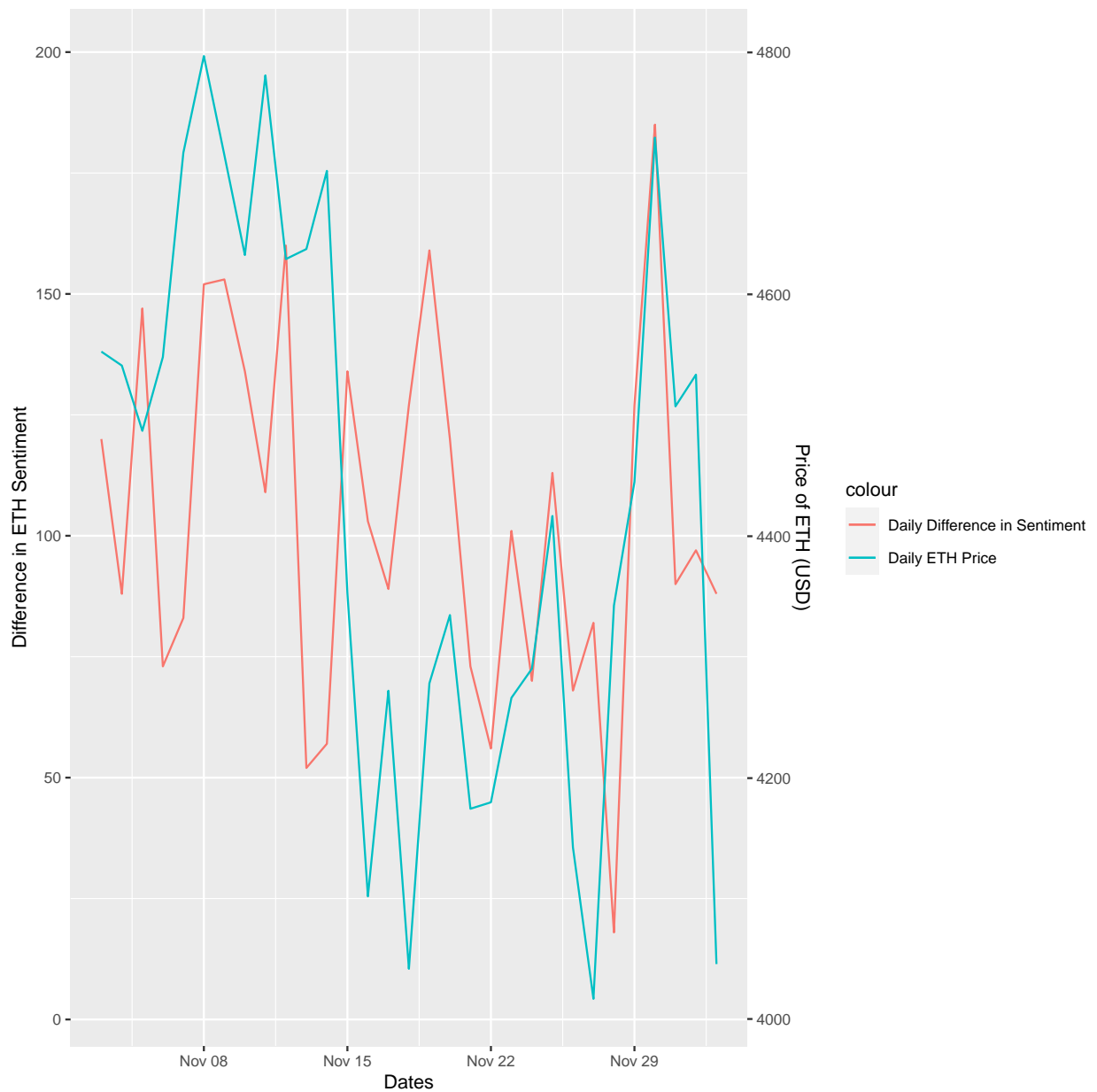
More insightful results can be seen (in Figure 10 above) with the tweet sentiments' relation to the price of ETH in the same period. The difference in positive and negative ETH sentiment is displayed and appears to be largely associated with given spikes or declines in ETH price. For example, the increase in price to $4800 on November 8th is in sync with significant increase in sentiments on the same day. The crash in price that occurred after on November 14th was quickly followed by a large reduction in positive sentiments. This trend similarly occurs following the ETH boom at the end of November. Thus, the reactionary terms observed in our frequency plot above are supported through this trend and can be interpreted as a call to take action on ETH at a particular instant. Additionally, there is substantial variability of positive sentiment in the data, which may indicate that "verified" users are more likely to tweet positively about ETH in boom periods and not exhibit any sentiment when the price begins to decline. This could potentially be due to a substantial increase in positive headlines concerning ETH in times of boom, which in turn creates similar sentiment through replies. This follows the general reactionary thesis, as such tweets and conversation would naturally follow any metrics observed in previous days.

Although there are many factors not immediately observable in this analysis, there is some evidence to suggest a relationship between ETH sentiment identified in tweets and the price of ETH during the same day. Higher frequency data measured in smaller units of time such as hours or minute intervals could be more helpful in drawing a causal relationship between these two variables.

NFT Tweet Analysis

To explore how NFTs are associated with the price of Ethereum, Twitter data was collected for a period of 34 days (July 6 - August 10, 2021) in which the ETH price displayed a significant incline. Filters were similarly applied for English tweets from "verified" accounts, but this time "NFT" was used as a keyword and "giveaways" were removed from the data. This ensured that promotional spam was not included in the limited supply of tweets, due to its significant presence on Twitter and relative unlikelihood of affecting ETH price. After cleaning the data, the number of NFT tweets per day was plotted against the price of ETH, along with a sentiment analysis (shown in Figure 11 below).

The number of NFT tweets appears to be extremely volatile, moving through four distinguishable cycles with lows of 150 to highs of 400 NFT tweets per day. Demonstrating no significant association with the price of ETH, the number of NFT tweets per day seems to be influenced by other factors; perhaps the specific day of the week has an effect on tweet volume, with more users having time to tweet and buy NFTs on weekends.

In contrast, the change in sentiment data on NFTs (found in Figure 12 below) suggests that this asset may be somewhat associated with the price of ETH. The pattern observed in the graph indicates a strong change in sentiment as the price of ETH changes; however, the positive association presented in the ETH sentiment analysis above is not as strong here. This follows a more volatile pattern, where even a slight increase or decrease in the price of ETH may be followed by widely changing sentiments, as observed on July 12th and 26th. The higher volatility can potentially be explained by NFT Twitter data occurring in smaller magnitudes than ETH Twitter data in a similar time range. However, a more interesting point demonstrated in this graph, is that the ETH price seems to almost be reactionary to the NFT sentiment! An example can be observed on July 18th, where a large increase in NFT sentiment is followed by gains in ETH price, starting on July 19th. Since NFTs are one of, if not, the main asset on the ETH blockchain, this data makes sense and could prove to be valuable when predicting ETH price.

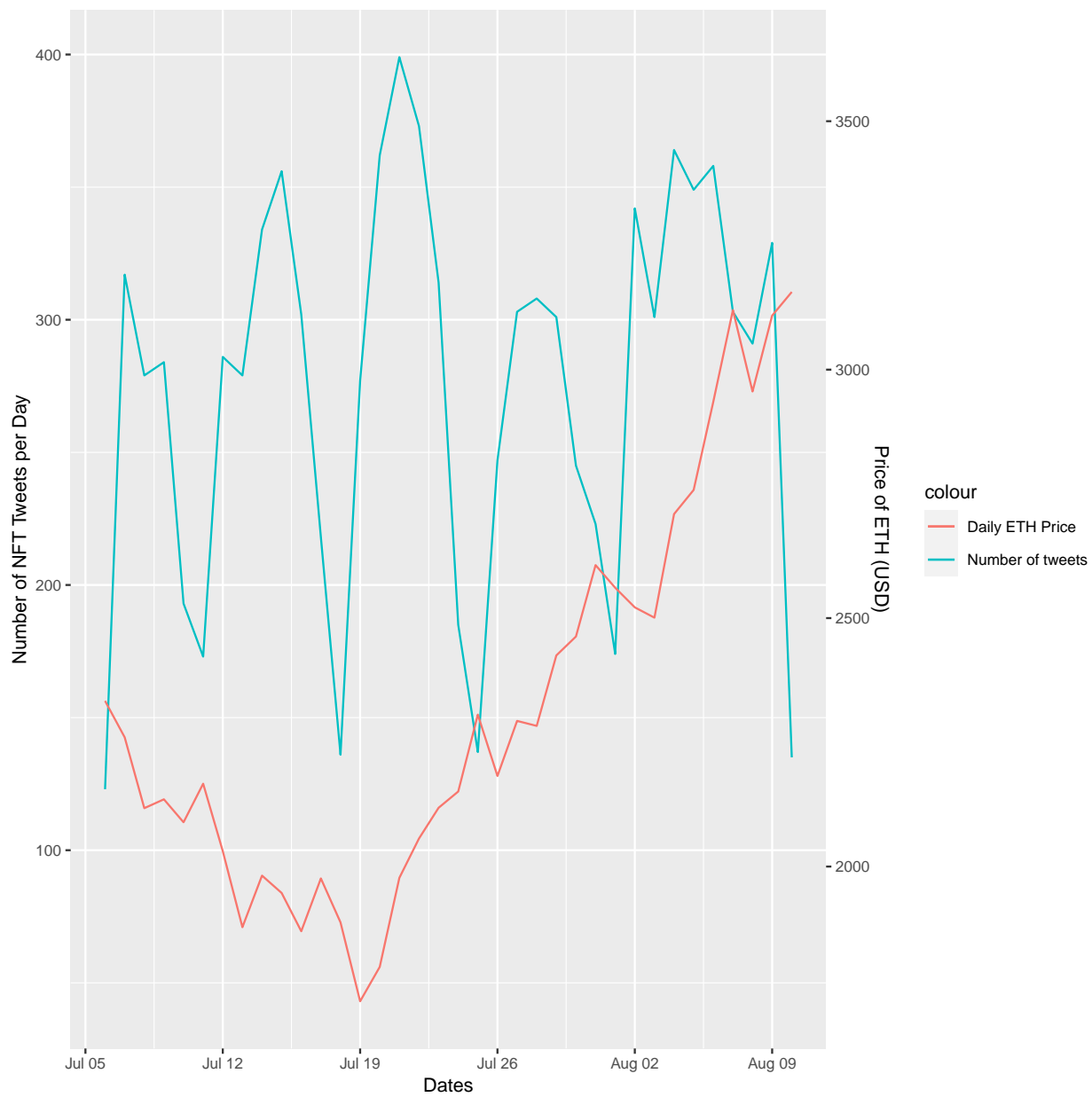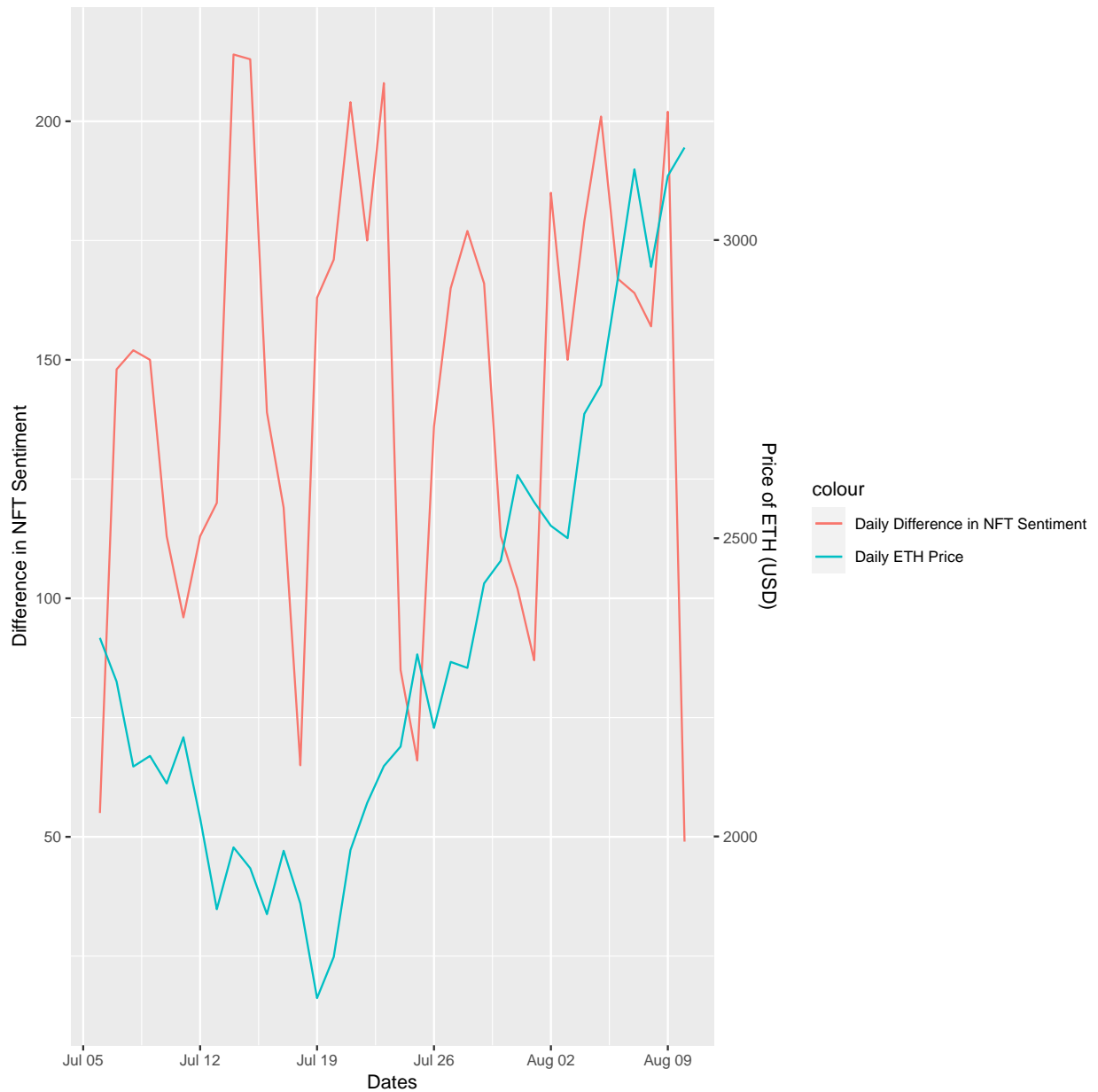**Figure 11:** Daily NFT Tweet Number and ETH Price (July 5 - August 10, 2021)

Figure 12: Daily NFT Tweet Sentiment and ETH Price (July 5 - August 10, 2021)

Ultimately, the results imply there may be an association between the price of ETH and NFT sentiment, supporting the frequency plot's results that the term "nft" is important to ETH. Similar to the ETH data above, NFTs' newfound popularity on social media platforms and unique qualitative application to the Ethereum blockchain could be valid reasons for associating this topic to future ETH price changes, especially as its market size continues to grow in time.

Conclusion

In this report, we attempt to predict changes in the price of ETH using financial market data and data directly related to activity on the Ethereum network. Our results indicate that the performance of competing coins like BTC and LTC, as well as the performance of the US energy and real estate markets, explain a large portion of the variance in the price changes of ETH. These results indicate that the price of Ethereum is largely driven by the same factors that affect the prices of BTC and other cryptocurrencies, such as uncertainty over the regulatory landscape in the cryptocurrency space as well as fears over the volatility of these currencies. With that said, Ethereum stands out for its potential to drastically reduce the amount of energy needed to support its network, making it a popular investment choice when energy costs are high or other traditional investment spaces like real estate are underperforming. When it comes to forecasting tomorrow's price of Ethereum, we find that performance in the real estate and bond markets actually preempts changes in Ethereum's price. We speculate that falling yields in the bond market drive investors to more speculative assets such as Ethereum. Due to a slight lag between these two events, we are able to develop a successful trading algorithm that appears to accurately predict when the price of Ethereum is likely to increase. However, we are limited by a small dataset and are unable to test how reliable this trading strategy is over a prolonged period. Research into how public sentiment and investors' views of Ethereum change under different market scenarios provides a fruitful avenue for future exploration.

Bibliography

H. Jang; Jaewook Lee, "An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information." *The Institute of Electrical and Electronics Engineers* (2018)

M.J.Hamayel; A.Y. Owda, "A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms." *AI* (2021)

Helder Sebastião; Pedro Godinho, "Forecasting and trading cryptocurrencies with machine learning under changing market conditions." *Financial Innovation, Springer;Southwestern University of Finance and Economics* (2021)

Myeong Jun Kim; Nguyen Phuc Canh; Sung Y. Park, "Causal relationship among cryptocurrencies: A conditional quantile approach." *Finance Research Letters* (2020)

Acknowledgements