NFT Price Distribution and Causal Analysis

NFTの価格分布と因果分析

付和文抄訳

A Thesis Presented to the Faculty of
the International Christian University
for the Baccalaureate Degree

国際基督教大学教授会提出学士論文

by

RAYAPEDDI, Niket
ラヤペディ　ニケト

231765

June 2023

RAYAPEDDI, Niket
ISHIBASHI, Keisuke

TABLE OF CONTENTS

Introduction **2**

Literature Review **5**

Theoretical Framework- ……………...**8**

Data and Empirical Methodology

Data …………………………………………………………………………**11**

Empirical Methodology………………………………………………………...**15**

Spectral Embedding……………………………………………………….**15**

VARLiNGAM model ……………………………………………………**16**

Results

Price Distribution Analysis…………………………………………………….. **21**

VARLiNGAM Analysis…..…………………………………………………… **22**

Conclusion **30**

References **32**

和文抄訳..34

Introduction

In 2018, a certain collection of Non-Fungible Tokens (NFT) known as CryptoKitties collection generated up to a total of 20 million dollars off of customer spending shortly after its launch. NFTs are digital assets stored on the blockchain which have gained significant popularity since 2017 (Kapilkov, 2020). NFTs serve many purposes, such as being digital representation of physical objects and eliminating the need of intermediaries. In addition, it can also be seen as a form of investment. However, the liquidity of the NFT market and the buyer and seller behavior have raised questions, leading to the analysis of the price variable of NFT markets.

Several papers, including Mekacher et al., (2022), Vidal-Tomás (2022), and Nadini et al., (2021), discuss the properties of the emerging NFT market. While price is a topic of discussion, only few delve into its behavior over time. However, Sax (2022) stands out as they not only look at the price distribution of various NFT markets but also examine how it changes over the course of several months. Although the results are compelling, the limited time frame of the observations may diminish its significance. Furthermore, none of the papers present any explanation for the behavior behind the price distributions. This paper attempts to explain this price behavior while also considering a longer time frame of observation .

The analysis aims to get a better understanding of the dynamics of the NFT market by examining the 2020 Crypto Kitties price data. By observing changes of the price over the course of the year, one can discern how the NFT market operates in response to time. Speculations regarding price volatility of NFTs can be observed. These observations enable the NFT market to be compared to other traditional assets, such as stocks, to determine the nature of NFTs as a financial asset. Part of the analysis involves identifying factors which may contribute to change in the NFT price distribution variable. Understanding which factors impact the NFT market not only enhances our understanding of the asset but it also allows for better decision making regarding investment in NFTs. Since investment in NFTs may carry some risk, the results from the research will provide potential investors with information necessary to make informed decisions.

The purpose of the analysis is to detect any changes and trends in the price distribution over the course of the year 2020 for the CryptoKitties NFT market. NFT markets are shown to display a lognormal distribution and are not subject to significant change in the short run of a couple months (Sax, 2022). The question is whether the same statement holds true in the longer run, such as a year. The initial phase of the research involved understanding the overall changes in price distribution of the NFT market over the course of the year 2020. Results were obtained through plotting monthly distributions of the NFT prices followed by quantitative and visual comparison of each distributions. Once results were obtained and changes were identified, the next step was to determine the factors responsible for the change in price over the course of time.

The next phase of the research focused on identifying factors which have an effect on the NFT price and determining the causal relationship which may exist between both variables. Previous studies have shown that certain macroeconomic variables, such as GDP and inflation, have a significant effect on several financial assets, such as stocks (Heinlein & Lepori 2022). The question arises whether these macroeconomic factors also have an effect on the NFT market as well. By identifying which factors had a causal relationship with the NFT price variable, we gain a better understanding of what factors may have contributed to the change in price distribution during the initial phase of the research. Some of these factors included interest rate, inflation, and stock index. To test for causal relationship, weekly average price value from the 2020 dataset was extracted and used to build a VARLiNGAM (Vector Autoregressive Linear Non-Gaussian Acyclic Model) model, a statistical model used for time series analysis to capture instantaneous and lagged causal relations between variables. The year 2020 was selected as it was the year the CryptoKitties market was relevant in terms of number of sales and customers, which enabled the research to utilize a substantial amount of data. The Varligam model was used to calculate the causal effects between average price and several variables as well as identify the direction of causality.

This paper will provide an overview of the distribution and the causal analysis for the price variable of the Crypto Kitties. The literature review glances at existing literature which analyze the NFT market, the price of NFTs, and the distribution of the price variable of several markets. It will also include criticism of the previous work and discuss areas of further development. The theoretical framework will then explore the possible relationship between the independent variables in the model using macroeconomic theory. The theoretical framework is used to presume the causal relationship between the independent variables and the NFT average price. Finally, data and empirical methodology describe the dataset used in this paper along with the methodologies employed to identify change and causal relationships in the dataset. The results of the model and the following conclusions will conclude the paper

Literature Review

Mekacher et al, (2022), examined the differences in price of NFTs while considering its scarcity. The research analyzed several collections involving 1,479,020 published NFT data, between January 23 2018 and June 6 2022. These data points were collected from OpenSea, a marketplace for NFTs. Rarity of NFT within a collection is calculated as the sum of the rarity of all the individual traits. Individual trait rarity was calculated as a fraction of NFTs within a collection possessing this trait. Schaar and Kampakis (2022) also explored rarity and its contribution to price through looking specifically at 11,000 transactions with the CryptoPunks collection and made use of the hedonic regression for analysis. Both studies revealed how the rarity attribute of NFT does have a positive effect on the price. Furthermore, Schnoering & Inzirillo (2022) built on the previously mentioned research and utilizes the hedonic framework to calculate indices based on fitted hedonic regression lines, which are commonly used in real estate. Hedonic indices served as proxies for price indices such as Consumer Price Index. These indices allow for them to observe dynamics and performance of the NFT market, which possesses a strong upward trend. The aforementioned literature does an exemplary job at analyzing a particular feature, the rarity, and assessing its contribution to the price of NFTs using hedonic indices. However, this paper differs from those studies as it doesn't examine internal variables such as the characteristics of the NFT asset. Moreover, this paper focuses on the relationship between price to external variables, such as inflation, interest rate, and stock index rate, which operate outside of the NFT market.

Vidal-Tomás (2022) looked at the specific niche of NFTs, the play-to-earn and metaverse tokens. With the data extracted from the CoinGecko database, they were able to analyze the co-movement in the tokens with the crypto-market using both Pearson and Kendall correlation. Using the short term and long term performance of tokens, calculated through equations provided in the report, it was possible to compare the dynamics/performance of the tokens pre and post covid. The post covid long term performance of the metaverse and play-to-earn tokens was revealed to be positive indicating that covid-19 did not have a negative impact on the market. The following literature offered a significant variable to account for since the dataset used in the paper is from 2020, the midst of the pandemic. Since it is shown that covid-19 had no negative impact on the price, it rules it out as a possible factor for the negative changes in the price distribution from the first half of the research.

Nadini et al, (2021) aimed at characterizing statistical properties of the market in addition to building a network of interactions between sellers and buyers. The data was collected from NonFungible Corporation, a company which tracks historical NFT sales data. The NFTs were categorized into several categories where it was revealed that the Art, Collectible, and

Games category dominate the market in the respective order. Price, Sale, and Collection distribution for each category mostly display a power law distribution. Network of trades was considered where a trader's strength, the total number of purchases, and sales made by the trader were all found to be distributed as a power law. Similarly, the Network of NFTs revealed how the strength of NFTs, defined by the number of times NFT was sold, also possess power law distribution. The collected and cleaned panel dataset utilized in the research includes various transactions involving NFTs from various markets spanning from the year 2017 to 2021. While this paper does not aim at analyzing NFT networks, the dataset utilized in the following literature was resourceful in the extraction of transaction data from the OpeanSea platform. This dataset involved the price of NFTs sold, which can be utilized to create a histogram to analyze the price distribution of specific NFT markets. This same dataset was used to conduct causal analysis with the addition of extra macroeconomic data.

Takaai et al, (2011) look at the cross sectional price distribution of housing prices in greater Tokyo. The results revealed how the distribution followed the hypothesized pattern of a log-normal distribution, where prices were asymmetric and skewed to the right. In relation to price distribution, Sax (2022) examined how price distributions look in the whole NFT market, exploring what patterns emerge across various NFTs markets. The study revealed how price distributions were found to be very concentrated across collections and had a consistent general shape. While both studies explored the change in price distribution of different markets, they inspired the idea of conducting temporal analysis of the price distribution on a particular NFT market. Sax's examination, however, of the NFT market distribution change was limited to the span of only 3 months. This short span of time may have influenced the results which showed little to no change in the price distribution. A longer duration could have allowed for external variables to take effect resulting in a change of results. Therefore, this paper examines the change in price distribution over the course of a year, a significantly long period of time. In addition, the aforementioned literature does not investigate the reasons for the change in price distribution. Causal Analysis was conducted to explore the relationship between the NFT price variable and several other variables. This is done to gain a better understanding of the potential factors which contribute to the observed changes in price distribution.

Theoretical Framework: Asset Demand and IS-LM model

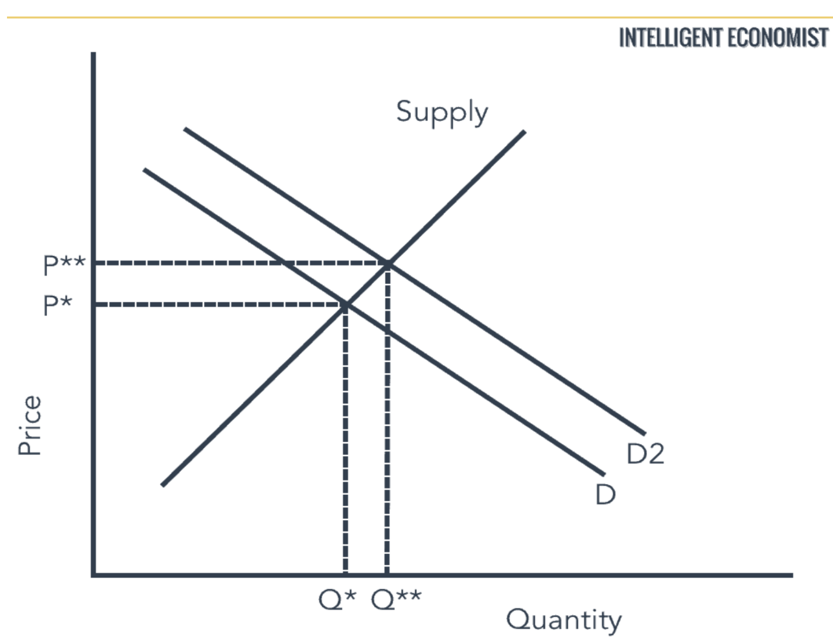


Figure 1: The Asset Demand model

Source: Theory of Asset Demand from the Intelligent Economist

The asset demand model demonstrates the relationship between a demand for an asset, specifically NFTs, and its corresponding price. The model comprises the demand and a supply curve for the asset. The asset demand curve can explain the effect NFT popularity would have on the price. According to the asset demand curve, if NFT popularity is considered as a determinant of the NFT price, an increase in the popularity will drive up the demand for the asset. This will result in the shift in the demand curve from D to D2, as shown above, raising the price of NFTs from P* to P**. In addition, the increase in popularity may also have an effect on supply of NFTs, resulting in the shift in the supply curve and consequently, an increase in the price of NFTs. This could be attributed to the increase in production and transaction volume of NFTs on the blockchain by both artists and traders.

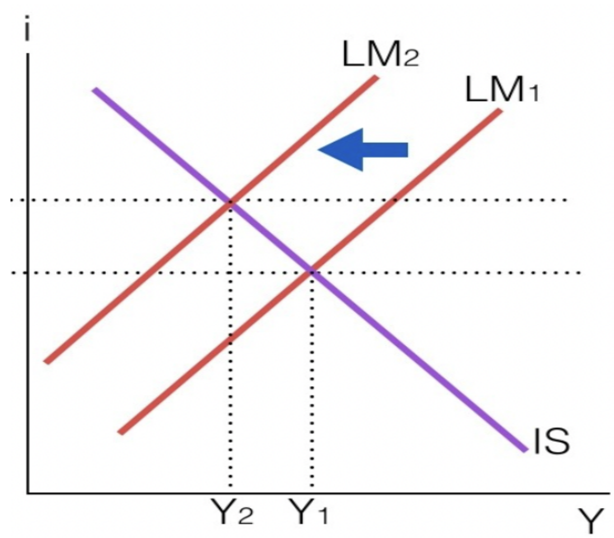


Figure 2: The IS-LM model

Source: How IS and LM modeling from Flypaper Effect

The IS-LM model is used to demonstrate the relationship between interest rates and the demand for goods which ultimately contribute to the outcome of the economy. In the context of this paper, the model can be employed to explain how change in interest rates may affect NFT prices. The model consists of the IS curve, which captures the relationship between output and interest rate, and the LM curve, which captures the relationship between interest rates and the money market. Together, the IS-LM captures the demand for money and investment in response to interest rates. An increase in interest rates will shift the LM curve to the left due to a decrease in the demand for money. This shift results in the drop of demand for NFTs and its price respectively because of the opportunity cost of holding assets that do not generate interest. In other words, the IS-LM model suggests a negative relationship between interest rates and NFT prices.

Additionally, the increase in inflation will also result in the shift of the LM curve to the left, due to increased desire for money to keep up with increased prices. This will indicate that the central bank will need to increase interest rates to reduce money supply as a means to combat inflation. If interest rate drops as a result of high inflation, this will result in a decrease in demand for NFT and their price respectively as explained earlier. While the IS-LM model suggests a negative relationship between inflation and NFT prices, there is a possibility that inflation can be triggered by an increase in prices and demand which results in increased output.

This scenario would shift the LM curve to the right resulting in an increased demand for NFTs and a higher NFT price. Therefore the IS-LM model may also suggest a positive relationship between inflation and NFT prices.

Data and Empirical Methodology

Data

This paper focuses on the analysis of the price distribution of the panel data of the CryptoKitties NFT market throughout the whole year of 2020. As mentioned previously, The year 2020 was selected based on the market's relevance in terms of number of sales and customers, enabling the research to utilize a substantial amount of data. The price data was then retrieved from each of the 60,000 transactions of NFTs made throughout the entire year. The dataset was then split to twelve datasets for each month of the year for monthly analysis to identify changes in the price distribution as well as its shape over the course of 2020. For the sake of simplicity, summary statistics for the entire year, as opposed to monthly statistics, was calculated and provided below in table 1.

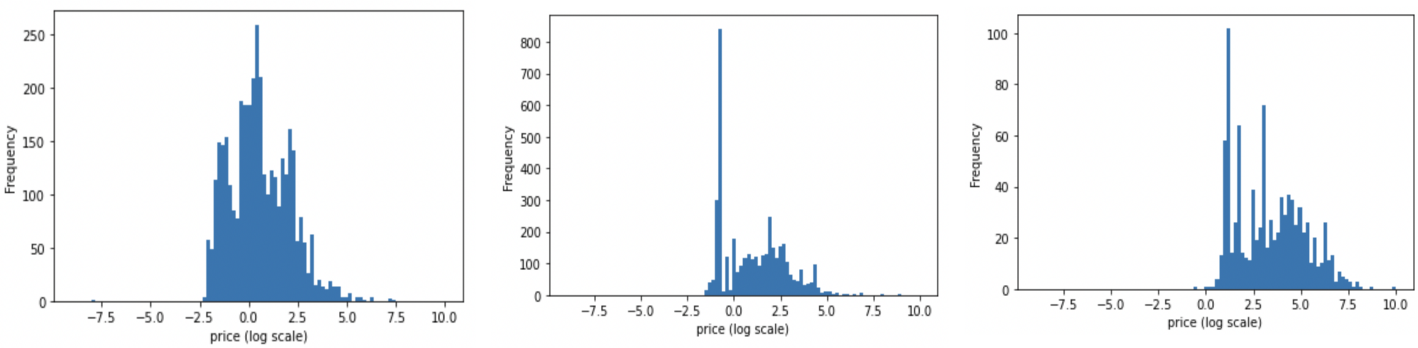
Table 1: Summary Statistics for 2020 CryptoKitties Data

	ID_token	Price_Crypto	Price_USD
count	6.000000e+04	60000.000000	59909.000000
mean	1.574406e+06	0.053928	33.470538
std	4.257854e+05	0.604975	805.670387
min	8.000000e+00	0.000002	0.000242
25%	1.333298e+06	0.002000	0.582293
50%	1.792304e+06	0.005000	1.994343
75%	1.875208e+06	0.024301	9.179875
max	1.995657e+06	85.000000	133234.525000

Source: Calculations of 2020 CryptoKitties Price data, based on panel data collected from OpenSea which includes various NFT markets spanning from 2017 to 2021

Data was collected from Opensea, a NFT marketplace. The summary statistics for the dataset is provided above in table 1 for the key variable, price. The prices for the NFT were measured in U.S. dollars. As observed in table 1, a significant amount of variance exists within the dataset in regards to the mean. The large contrast between the minimum price and the maximum price implies the diverse range of prices for the CryptoKitties NFTs. The mean for the entire dataset is shown to be significantly low, given the wide contrast between the minimum price and maximum price. The following dataset was indexed monthly to further explore the temporal changes of the price distribution. The plotted price distribution for the months of January, July, and December are given below in figure 3.

January July December



<u>Figure 3: Price Distributions of CryptoKitties Market (January, July, and December)</u>

Source: Monthly Price Distributions based on 2020 CryptoKitties Price data

All the monthly distributions were log scaled to account for the large amount of skewness present within all the distributions as shown above. Most of the monthly price distributions displayed lognormal qualities due to their right skewness, low mean, and high standard deviation. However, significant changes in the average price, shape, and frequency of the distributions were detected over the months as seen from figure 3. Further investigation on the difference between the monthly price distributions can be visualized through techniques such as spectral embedding.

In addition to analyzing change in price distribution, this paper focuses on identifying causal relationships between NFT price and several external variables. To achieve this, a VARLiNGAM model was constructed to measure the causal effects between the variables. As a significant amount of sample points was required to obtain more reliable results from the model, weekly average data was used as opposed to monthly average data. In addition, time series weekly data for the individual variables was obtained and attached to the dataset from table 1. As a part of data preprocessing, the differencing technique was used to remove any trends and seasonality present in the dataset. The summary statistics of the differenced dataset is shown below in table 2:

Table 2: Summary Statistics for 2020 Differenced Weekly Data

	average price	popularity	interest rate	stock_index	ethereum_index	cpi	bitcoin	num_of_trans
count	51.000000	51.000000	51.000000	51.000000	51.000000	51.000000	51.000000	51.000000
mean	9.922327	0.551020	0.330364	8.803922	11.196078	-0.005882	-8.205863	-20.137255
std	342.410582	7.722852	0.543305	109.124886	43.585328	0.162464	1378.551486	649.610037
min	-1680.470635	-15.000000	0.045714	-353.000000	-139.000000	-0.490000	-2443.305664	-1597.000000
25%	-10.450460	-3.000000	0.082143	-35.000000	-14.500000	0.000000	-997.841309	-230.500000
50%	0.876492	0.000000	0.090000	26.000000	9.000000	0.000000	-123.347657	-32.000000
75%	19.320881	4.000000	0.100000	66.500000	34.500000	0.000000	1023.915771	202.500000
max	1560.088478	30.000000	1.585714	187.000000	124.000000	0.672000	4193.274413	2270.000000

Source: Calculations of 2020 CryptoKitties Price data in addition to weekly data obtained from Google trends, FED, Yahoo Finance, and US Beauro of Labor Statistics

All variables in Table 2 have been retrieved from different sources and are measured in different scales.The description of each variable is as follows:

1. Average Price is calculated using weekly NFT price of Crypto Kitties from OpenSea

Market place.

2. Number of Transactions (num_of_trans) is the weekly number of transactions which take place in the marketplace.
3. The popularity variable is used to measure the level of popularity of NFTs. The data for popularity was extracted using Google Trends and is scaled from 0 to 100, with 100 being the peak popularity in terms of google search.

Financial indices, such as stock index, ethereum index, and bitcoin index were collected through Yahoo Finance. These indices were included in the dataset to investigate the causal effects financial markets may have on the NFT prices.

4. Stock index measures performance of the stock market. The S&P 500 stock index was selected for this paper
5. Ethereum index measures the performance of the Ethereum cryptocurrency
6. Bitcoin index measures the performance of the Bitcoin cryptocurrency

Macroeconomic variables, such as inflation and interest rate, were also included as variables due to its possible influence over NFT prices as explained in the Theoretical Framework section.

7. The Fed Funds rate (interest rate) was used to measure the interest rate as it is the rate set by the central bank of the US and is the cost of borrowing for both businesses and consumers.

8. Consumer Price Index (CPI) was used to measure inflation as it is a popular measure of

change in price for a fixed basket of goods and services over time. Since the weekly CPI was unable to be obtained, the monthly values were substituted instead.

The 51 count value corresponds to the 51 weeks worth of data samples collected throughout the year. Once the data was properly constructed and cleaned, the dataset was normalized to account for the difference in metrics between the 8 variables. This was done as a final step of preprocessing to get more accurate and unbiased results of the causal effect from the VARLiNGAM model.

Empirical Methodology

1. Spectral Embedding

To compare 12 different separate price distributions with each other would be a daunting task as it would be difficult to measure the change over the year. To overcome this challenge, spectral embedding was employed to visualize the differences between the monthly price distributions in the year 2020 of the CryptoKitties NFT market. It is a technique used for bringing higher dimension data to a lower dimension while preserving the key properties. The technique allows for the comparison of the monthly price distributions through quantitatively measuring the similarities between the distributions and storing them in a matrix. The matrix is embedded into a lower dimension through selecting appropriate eigenvalues and eigenvectors. It enables a more effective means to analyze and interpret the changes in price distributions throughout the year as opposed to simply comparing price distributions/average prices.

The execution of spectral embedding analysis involved constructing a similarity matrix using the

$$d(p,q) = \sqrt{(p-q)^2}$$

Euclidean distance of the average monthly prices. A similarity matrix is a quantitative representation of the similarity between the different pairwise values in the matrix. In this case, it would represent the level of closeness of the monthly price distributions. To represent the monthly price distributions, the average price of each month was selected. To compare the similarity between the monthly averages, the euclidean distance was used as shown in figure 4.

<u>Figure 4: Euclidean Distance</u>

Source: Formula to compute similarity of pairwise values in similarity matrix

Through taking average price p and average price q, one can calculate the level of similarity between the two by measuring the distance between both points. Therefore, a higher value would imply a lower level of similarity. The euclidean distance was calculated between all of the monthly average values and stored in the similarity matrix. The constructed similarity matrix was then fitted into the spectral embedding model in python. The model projects the average monthly data points to a lower dimension through utilizing the appropriate eigenvalues and eigenvectors derived from the matrix. A scatter plot can then be plotted based on the embedded matrix to visualize the lower dimensional representation of the price distributions.

2. VARLiNGAM Model

The VARLiNGAM model was used to measure the causal relationship between external variables and NFT price. It is an extension of the VAR model, which is used to analyze the level of effect lagged values have on the current values. The VARLiNGAM is superior to the VAR model in that it allows for the analysis of the causal relations of both the lagged and instantaneous values. The order of the model, or the number of lagged variables to include in the model, is automatically calculated by the model during implementation. VARLiNGAM allows for a proper examination of both the degree and direction of causal effect on NFT price by looking at the effects of both the current and past values of the external variables.

The following assumptions are required when utilizing VARLiNGAM model:

1. Linearity：variables have a linear relationship with each other
2. Non-Gaussian continuous errors: Non- normal residual distribution
3. Acyclicity of contemporaneous causal relations: No cycles in the causal graph
4. no hidden common causes: Assumption of common causal effects on NFT prices

Linearity for the current variables was confirmed through creating a heatmap which revealed the variables to have a relatively close linear relationship. Whether the same could be said about the lagged variables was left unchecked. Non-Gaussian continuous errors were confirmed through creating distributions of the residuals of the model which mostly returned non normal distributions. The causal graph created from the model also displayed acyclicity meeting the third condition. Lastly, it was assumed that the common causal effects of NFT prices were covered by all the explanatory variables in the VARLiNGAM model.

The mathematical representation of the model is given below in figure 5 as follows:

$$x(t) = \sum_{\tau=0}^{k} B_{\tau} x(t-\tau) + e(t)$$

Figure 5: VARLiNGAM mathematical model

source: keuchi,, T., & Haraoka, H., Ide, M., Kurebayashi, W, Shimizu, S. *Varlingam*.VARLiNGAM - LiNGAM 1.7.1 documentation. (n.d.). https://lingam.readthedocs.io/en/latest/tutorial/var.html

where:

x(t) represents the vector of all instantaneous variables but this paper only focuses on the the equation for current NFT price

k represents the order of the model. The order given by the model is one, indicating that only the instantaneous variables and the 1 week lagged variables are included in the equation.

τ represents the time lag. As this is a VARLiNGAM model, it starts with 0 to measure the effects of instantaneous variables

$B\tau$ represents the adjacency matrix with time lag τ. It is the correlation matrix which includes the coefficient of the variables in their respective linear equations. These coefficients identify the degree and direction of the causal relationship between the variables in the linear equation. The coefficient values are calculated using either Least Squares (LS) or Maximum likelihood estimation (MLE). LS is a method which identifies the coefficient values which returns the least amount of sum of squared errors between the observed NFT weekly price data and the predicted values from the model. Similarly, MLE is a method which estimates the coefficient values which has the maximum probability of generating the NFT weekly price data if used in the model.

t represents the time instance of the variables. For example, x(t - 1) would indicate the data 1 week prior to the current data (lag 1).

e(t) represents the vector of error variables for each of the linear equations. It measures the difference between the observed data and the predicted values from the VARLiNGAM model. It is included in the model to account for any variation not accounted for by the explanatory variables.

Σ (sigma notation) refers to the summation of all the terms in the linear equations. This includes the all the instantaneous lagged values with their respective coefficient values in addition to the error terms.

The VARLiNGAM model is retrieved from the lingam library and is executed in python. Before running the model, an indexed weekly dataset was imported into the platform (Google Colab). The data was then fitted to the VARLiNGAM model. The VARLiNGAM model includes weekly NFT prices, popularity, interest rate, stock index, ethereum index, cpi, bitcoin index, number of transactions, and their respective lagged versions as variables. Executing the model will identify the order of the model and then calculate the respective coefficient values for the causal effect between variables using MLE and LS.

From the implemented model, the causal order can be retrieved which comprises a list of variables which indicates the direction of the causal effect based on the order of the variable. In addition, the numerical causal effects can be displayed from the model in the form of an adjacency matrix. As stated previously, the values in the adjacency matrix correspond to the coefficients of the external variables which indicate the degree and direction of the causal effect. While the adjacency matrix systematically stores the causal effects, it is difficult to visualize how the values correspond to the causal effect which exists between the variables. A heatmap can be generated to gain a better visual understanding of the adjacency matrices. Additionally, the VARLiNGAM model provides the option to display a causal graph to visually represent the existing causal relationships in the model.

The causal graph is a graph representation of the adjacency matrix where each node represents the variables in the model, both the instantaneous and lagged variables. The edges between the nodes imply a causal relationship existing between the respective variables. The edges between nodes are directed which indicate the direction of the causal effect. Moreover, the edges contain weights which correspond to the causal effect coefficients from the adjacency matrices. Overall, the causal graph generated by the VARLiNGAM model enables a more comprehensive visual representation of the causal dynamics which are in play between the variables. This allows this paper to reach a conclusion as to which factors indeed have a causal effect on the NFT price variable.

As a means of testing the validity of the results, the VARLiNGAM provides a method to test the independence of errors from the model through a p test. The returned results will include a matrix with all the p values corresponding to all the error terms. If a value is less than 0.05, it would indicate a rejection of the independence of error, violating the model assumption. In addition, bootstrapping is also available to measure the probability that the results from the adjacency matrices and causal graphs are not random. Bootstrapping involves repeatedly estimating the causal effects with different samples drawn from the same dataset. The probability retrieved from the bootstrap samples indicate the likelihood of drawing the original coefficient from the resampled dataset.

<u>Results</u>

1. Price Distribution Analysis:

To capture the price distribution quantitatively, the average monthly price was used since it represents the central tendency of individual distributions. The similarity between the monthly distributions was calculated using the euclidean distance between the average monthly price of all the distributions. The respective euclidean distances were stored in a similarity matrix which was then fitted to a spectral embedding model to project the data points to a lower dimension. The lower dimension representation of the values was visualized through a scatter plot shown below in figure 6:

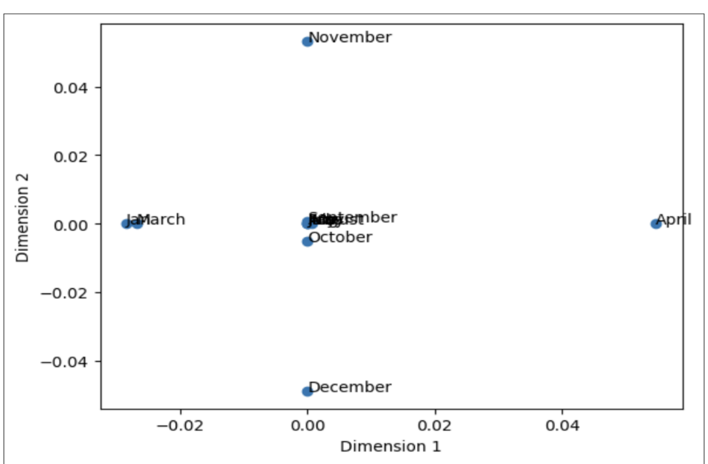


Figure 6: Scatter Plot of Spectral Embedded Average Monthly Prices

Source: lower dimension visual representation of monthly average price from spectral embedding analysis

The resulting scatter plot displays the different average monthly prices in a lower dimension and spaced to each other based on how similar they are to one another. The graph returned 5 different cluster groups as shown above. The formed clusters are months which are relatively close to one another in terms of time as shown through the January cluster and the October cluster. This would indicate periods of stability of the price distribution behavior. There were instances where the cluster group only consisted of 1 month, such as with November, April, and December clusters. This would indicate existing monthly outliers of price distributions. Overall, the formation of the different clusters spread out in the lower dimension imply significant changes in the price distribution over the year.

While significant changes were detected, it was important to consider the fact that average monthly prices were used as representations of monthly price distribution. Average prices may indicate the central tendency of the distribution, it does not capture the spread or the shape of the distribution. In addition, the scale of the x and y axis in figure 6 is unclear as it is in a lower dimension. It can be interpreted that x and y are some measure of similarity. Therefore the closer the clusters, the more similar they are. Furthermore, it is unclear as to what the possible reasons are that cause the formation of clusters. While the specific reasons for the changes in 2020 are unknown, possible factors which have a causal effect on the price values are retrieved in the VARLiNGAM model.

2. VARLiNGAM Model:

After constructing and implementing the VARLiNGAM, the causal order, adjacency matrix, and causal graph were retrieved for causal analysis. The returned causal order, a list of variables, is reported below in figure 7:

[CPI, Popularity, Average Price, Stock Index, Interest Rate, Number of Transactions, Bitcoin index, Ethereum Index]

Figure 7: VARLiNGAM Causal Order

Source: Called method from VARLiNGAM model utilizing 2020 CryptoKitties Price data

As mentioned previously, the order of the list corresponds to the direction of causal effect. In other words, a variable appearing ahead of the list (the left) is likely to affect the variables appearing after it (the right). Based on the results, we can deduce that the weekly average price variable is likely to be affected by CPI (inflation) and Popularity. The intuition behind the result holds based on what was explained in the theoretical framework regarding how inflation and popularity may shift price value. However, the causal order also reveals that the average price causes the rest of the variables in the model. This goes against some of the intuition proposed in the theoretical framework section since it was assumed that interest rate, number of transactions, bitcoin index, and ethereum index would have an effect on NFT price. It seems that the causal order list only reveals the ordering of the instantaneous variables as opposed to the lagged variables. While the causal order reveals rather interesting results, how this causal order is decided is unknown. In addition, the following adjacency matrices and the causal graph don't seem to be aligned with the figure 6, putting the credibility of the method into question.

Adjacency matrices are also generated from the VARLiNGAM model to capture the causal effect between the variables. More specifically, the coefficients in the adjacency matrices indicate the degree of the existing causal relationship. Each column and row corresponds to a variable where the intersection corresponds to causal effect between the column and row variable. The assignment of variables to the columns and rows is the same as the dataset from figure 4 (First Column/Row = Average price, Last Column/Row = number of transactions). The results are displayed in figure 8 and 9.

<u>Figure 8: VARLiNGAM Heatmap of Adjacency Matrix # 1 (X→Y)</u>

Source: Called Heatmap of Adjacency matrix from VARLiNGAM model utilizing 2020 CryptoKitties Price data

The heatmap for the first matrix in figure 8 displays the causal relationship between all the instantaneous variables only. Unexpectedly, the returned heat map consisted of values close to or equal to zero. This implies that either there is either no causal relationship existing between the variables or that the causal effect is significantly low. While the coefficient values signify the degree of causal effect, it did not indicate its direction. In other words, it is unclear which variables in the pair are affecting the other. Interestingly, a causal effect was detected between average price and popularity (0.354) as well as stock index and ethereum index (0.528) as per the adjacency matrix, indicating a positive relationship between both pair variables. However, while a positive relationship exists, it is not highly correlated as shown from the light blue color respectively on the heatmap from figure 7. In other words, while the popularity variable does play a factor in causing change in price, it does not seem to have a strong effect.

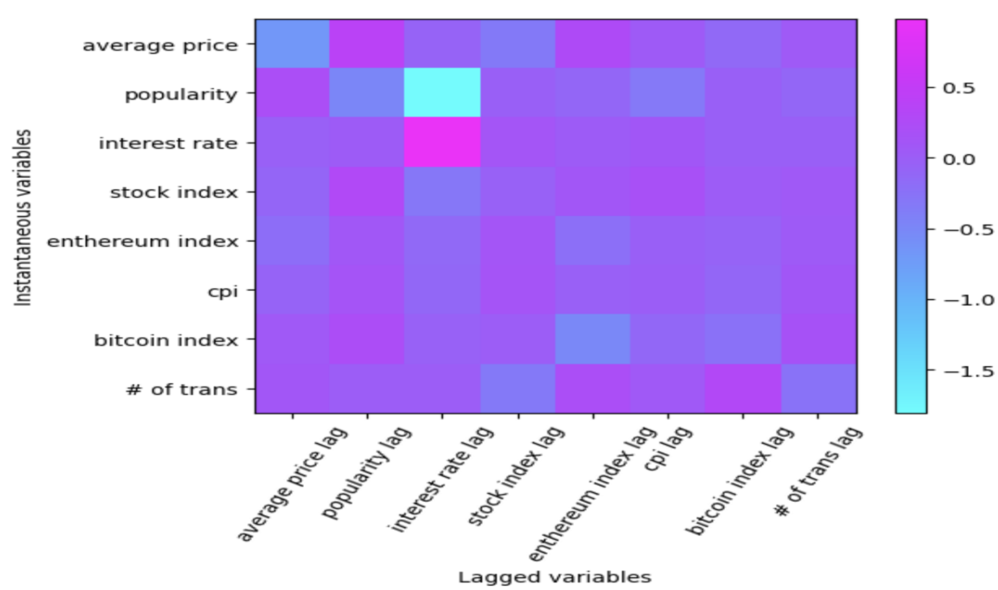


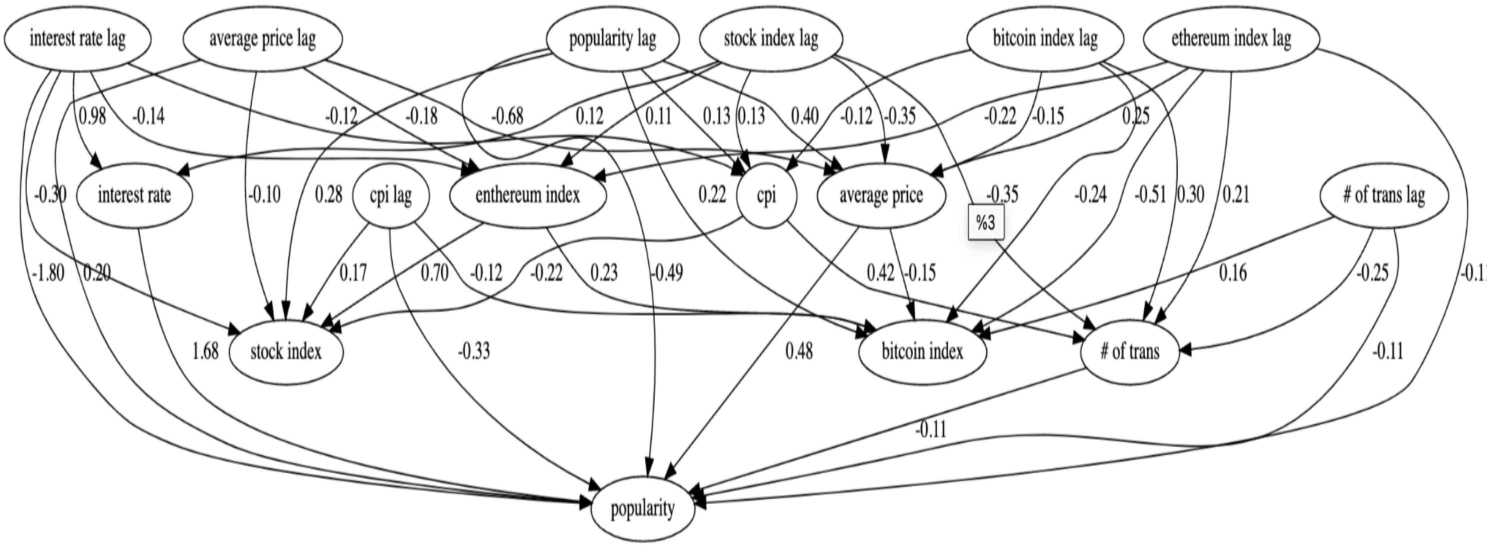
Figure 9: VARLiNGAM Heatmap of Adjacency Matrix #2 (X(t-1)→Y)

Source: Called Heatmap of Adjacency matrix from VARLiNGAM model utilizing 2020 CryptoKitties Price data

The heatmap of the second adjacency matrix in figure 9 displays the causal relationship between the instantaneous variables and the lagged variables. In this case, the columns correspond to the lagged values where the rows represent the instantaneous variables. The returned heat map consisted mostly of a variety of values in the range of -1 to 1. In contrast to the heatmap from figure 8, there were very few pair values which consisted of 0, implying that causal effects of some degree were identified for most pair variables. Additionally, most of the values from the heatmap were below 0.1 and above 0, indicating a positive yet low level causal relationship. The first row was the point of interest as it displayed the relationship between the instantaneous NFT price values to the lagged variables. Based on the color and values, it can be understood that lag average price, lag interest rate, lag stock index, and lag bitcoin index have negative causal relationships with the target variable. The other variables in the row displayed a positive correlation (lag popularity, lag ethereum index, and lag cpi). Similarly to figure 8, the coefficient values signified the degree of causal effect between the variables, but it did not indicate the direction.

To visualize the direction of the causal relationships between the variables, a causal graph was constructed through the VARLiNGAM model. The causal graph, including both the instantaneous and lagged variables, is included below in figure 10.

Figure 10: VARLiNGAM Causal Graph



Source: Called method from VARLiNGAM model utilizing 2020 CryptoKitties Price data

As mentioned previously, the variables in the model are represented as nodes with the directed edges between the nodes indicating the causal relationship between the variables. The nodes with “xxx lag” correspond to the first order lag variables, accounting for the one week delay of all the instantaneous variables. The weights on the edges correspond to the causal effect coefficients as displayed on the adjacency matrices. The causal graph revealed the variables which had a causal effect on the average price node representing the weekly average NFT prices. Average price lag , popularity lag, stock index lag, bitcoin index lag, as well as ethereum index lag were all shown to be directed towards average price. The remaining lagged variables (interest lag, number of transactions lag, and cpi lag) were shown not to have any causal relation with average price. Interestingly, none of the instantaneous variables were shown to have any causal effect on average price due to the absence of a directed arrows.

Of the variables shown to be affecting average price, only the lags of popularity and ethereum have a positive causal effect on average price as indicated by the positive weights on the respective edges. The remaining lags, such as stock, bitcoin, and average price, were shown to have a negative causal effect on average price. The weights on the edges between the external variables and average price were between 0.1 and 0.4 with the exception of average price lag which had a degree of 0.68 causal effect on average price. The range of 0.1 ~ 0.4 indicated a low degree of causal effect on the target variable.

An unexpected result was how the edges were shown to be coming out of the average price node, directed towards popularity and bitcoin index. This would imply that the average price of NFTs has a causal effect on the respective variables and not the other way around. While the significance and validity of the model was confirmed, as shown through figures 11 and 12, there were inconsistencies between the weights of the causal graph and the corresponding values in the adjacency matrices/heatmap. Some weights either slightly differ or significantly differ from the values putting the causal graph into question.

Independence of errors was checked to determine the validity of the model through obtaining the p values of each corresponding pair variable. Using a VARLiNGAM method, a matrix was returned, containing all the p values, as displayed in figure 11:

```
[[0.    0.068 0.614 0.878 0.948 0.929 0.633 0.594]
 [0.068 0.    0.803 0.893 0.367 0.859 0.116 0.895]
 [0.614 0.803 0.    0.878 0.128 0.001 0.524 0.628]
 [0.878 0.893 0.878 0.    0.619 0.24  0.007 0.191]
 [0.948 0.367 0.128 0.619 0.    0.841 0.621 0.581]
 [0.929 0.859 0.001 0.24  0.841 0.    0.805 0.356]
 [0.633 0.116 0.524 0.007 0.621 0.805 0.    0.118]
 [0.594 0.895 0.628 0.191 0.581 0.356 0.118 0.   ]]
```

Figure 11: VARLiNGAM model Independence of errors

Source: Called method from VARLiNGAM model utilizing 2020 CryptoKitties Price data

The returned matrix of the error terms between the instantaneous variables shown in Figure 7. It mostly consisted of values above 0.05 indicating that most of the error terms for the respective pair variables are independent. This would imply that most of the error terms are uncorrelated with each other which is an indicator of an efficient Varlignam model. However the p values between stock and bitcoin indices (0.007) in addition to interest rate and cpi (0.001) are shown to be less than 0.05, rejecting the null hypothesis that the variables have independent error terms. The rejection of these p values would suggest that the model does not account for factors which may lead to biased estimates. However, this was somewhat to be expected given that some of the independent variables are known to be correlated with one another, such as cpi and interest rate. This correlation between independent variables, known as multicollinearity, was accepted due to some of the independent variables serving as control variables. The diagonal values are 0 since the error terms between the same variables would be zero. Overall, the condition of independent error terms was met by most of the pair variables leading to more reliable estimates of parameters displayed in the causal graph in figure 10.

In addition to testing p values, bootstrapping was utilized to estimate the probability of the error terms from the adjacency matrices and causal graphs not being random. The results from the bootstrap is displayed under figure 12 below:

```
Probability of B1:
 [[1.   1.   0.86 1.   0.97 0.99 1.   1.  ]
 [0.99 1.   1.   1.   0.99 1.   1.   0.99]
 [0.95 0.92 1.   1.   0.96 1.   0.91 0.92]
 [1.   1.   1.   0.63 0.95 1.   0.7  0.96]
 [0.99 0.93 0.95 1.   1.   1.   1.   0.73]
 [1.   1.   1.   1.   1.   0.08 1.   1.  ]
 [0.96 1.   0.95 0.92 1.   0.98 1.   1.  ]
 [0.99 0.96 0.95 1.   1.   0.97 1.   1.  ]]
```

Figure 12: VARLiNGAM Model: Probability of Causal Effect

Source: Called method from VARLiNGAM model utilizing 2020 CryptoKitties Price data

The values in the matrix represent the probability of the causal effect coefficients generated in the second adjacency matrix heat from figure 9 in addition to the weights from the causal graph (figure 10). The probability was measured in the scale between 0 to 1 with 1 corresponding to a 100 percent. As displayed above, most of the values were shown to be above 0.95(95%) indicating that most of the causal effect coefficients generated in the adjacency matrix and causal graph were not random. The first row of the matrix, the area of interest, corresponds to the probability of non-randomness of the coefficients between the price variable and the lagged independent variables. From this row, only the probability of non randomness between average price and interest rate was shown to be below 0.95 (0.86). The results from the bootstrapping revealed that most of the coefficients in the VARLiNGAM model were significant implying that the generated model is valid.

Conclusion

The overall objective of the research is to gain a better understanding of the dynamics of the NFT market by analyzing the 2020 Crypto Kitties price data. This paper analyzes these dynamics through examining the monthly change in price distribution of NFTs in addition to identifying causal factors which affect the price variable.

Price Distribution analysis was conducted to detect change in the price distributions to gain an understanding of the stableness of the price variable and the market. Spectral Embedding was conducted with the monthly average price values to compare the different price distributions in a lower dimension. The results, as per the scatter plot of the embedded data, consisted of clusters of monthly average prices, which represent periods of time. This indicated that significant price changes between these periods of time existed. Based on the results, this paper has achieved its goal in detecting changes in price behavior over the course of a year.

Causal analysis was conducted to identify factors which cause change in the price variable of NFTs and possibly contribute to the change in price distribution throughout the year. Based on the results, the historic variables (lagged variables) of average price, popularity, stock index, ethereum index, and bitcoin index were all shown to have a significant direct effect on instantaneous price variables as per the VARLiNGAM model. This would indicate that the behavior of price is influenced not by immediate changes of external factors but to the changes in past values. In other words, the immediate variables are shown to have a delayed effect on price. It also demonstrates the direct effect other financial markets, such as stock and other cryptocurrencies, have on certain NFT assets. Unlike other traditional financial assets, it is revealed that macroeconomic variables, such as inflation and interest rate do not affect the price of NFTs. This is possibly due to NFTs decentralized nature, as it is consumed outside of traditional financial markets, being independent of the same factors as other assets. For this reason, the IS-LM model from figure 2 would not apply to NFT assets. While significant insights were retrieved from the VARLiNGAM model, inconsistencies between the adjacency matrices and causal graph existed. However, the overall model was proven to be valid through model valuation techniques and bootstrapping.

With these conclusions, this paper has achieved its objective in better understanding the dynamics of the price variable in the NFT market as well as how the price variable is affected by external factors to a certain extent. In addition, the nature of NFTs as a financial asset and how it compares to traditional assets was also better understood. As NFTs are still a relatively new market and are shown to be price volatile, they carry a significant amount of risk. The findings from this paper may allow for improved decision making regarding investment in NFTs. However, it is important to keep in mind that the findings may be exclusive to the year 2020 or even the specific NFT market (CryptoKitties) . Therefore, further research has to be conducted to ensure these dynamics also take place in the following years and with different NFT markets. But given that these factors are accounted for in addition to the information presented in this paper, it gives potential investors a better glimpse of the NFT market and the necessary knowledge to make more informed decisions regarding NFT related transactions.

List of References

Heinlein, R., & Lepori, G. M. (2021, August 14). *Do financial markets respond to macroeconomic surprises? evidence from the UK - empirical economics*. SpringerLink. https://link.springer.com/article/10.1007/s00181-021-02108-1

Ikeuchi,, T., & Haraoka, H., Ide, M., Kurebayashi, W, Shimizu, S. *Varlingam*. VARLiNGAM - LiNGAM 1.7.1 documentation. (n.d.). https://lingam.readthedocs.io/en/latest/tutorial/var.html

Mekacher, A., Bracci, A., Nadini, M., Martino, M., Alessandretti, L., Aiello, L. M., & Baronchelli, A. (2022, October 16). *Heterogeneous rarity patterns drive price dynamics in NFT collections*. nature.com. https://www.nature.com/articles/s41598-022-17922-5

Nadini, M., Alessandretti, L., Giacinto, F. D., Martino, M., Aiello, L. M., & Baronchelli, A. (2021, October 22). *Mapping the NFT revolution: market trends, trade networks, and visual features*. nature.com. https://www.nature.com/articles/s41598-021-00053-8

Takaai, O., Takayuki, M., Chihiro, S. D., & Tsutomu, W. (2011, March 11). *Evolution of House Distribution*. DP. https://www.rieti.go.jp/jp/publications/dp/11e019.pdf

Sax, G. (2022, March 18). *Exploring NFT Price Distribution Across Collections*. medium.com. https://medium.com/@goblinsax/exploring-nft-price-distribution-across-collections-782193240f2c

Schaar, L., & Kampakis, S. (2022, January 18). *Non-fungible Tokens as an Alternative Investment:Evidence from CryptoPunks*. scholar.googleusercontent.com. https://scholar.googleusercontent.com/scholar?q=cache:wMGNE8eazjMJ:scholar.google.com/++journal+of+economic+perspective+nfts&hl=en&as_sdt=0,5

Schnoering, H., & Inzirillo, H. (2022, March 7). *Constructing a NFT Price Index and Applications*. https://arxiv.org/abs/2202.08966

Vasan, K., Janosov, M., & Barabási, A. L. (2022, February 17). *Quantifying NFT-driven networks in crypto art*. nature.com. https://www.nature.com/articles/s41598-022-05146-6

Vidal-Tomás, D. (2022, June 10). *The new crypto niche: NFTs, play-to-earn, and metaverse tokens*. sciencedirect. https://www.sciencedirect.com/science/article/abs/pii/S1544612322000630

和文抄訳

この論文の全体的の目的は、2020年のCryptokitties NFTの価格データを分析することにより、NFT市場のダイナミクスとNFT価格変数の挙動を理解することである。分析の内容としては、価格分布の月次変化の調査し、価格変数に影響与える因果関係を指定した。

価格分布分析には、NFT価格変数と市場の安定性を理解するために2020年のデータの中に価格分布の変化の検出を行われた。スペクトラル埋め込みと言う可視化技術を用いて、月次価格分布を表す月次平均価格を異なる月次平均価格と低次元で比較することを行われた。埋め込まれたデータの散布図にはクラスタは存在し、これは時間帯の間に価格の変動が存在してると示してる。

因果分析には、NFT価格変数の変化を引き起す要因を特定した。VARLiNGAMモデルにより、どの変数が価格変数に重要と直接的な影響をもつかを示すことができる。結果に基づいて、平均価格、人気度、株価指数、イーサリアム指数、ビットコイン指数の過去

の変数が影響を持つと示された。NFT価格挙動は外部要因の即時的な変化ではなく、過去の値の変化に影響を与えられる。

これらの結果により、NFTの価格変数と市場のダイナミクスの理解を達成した。特に、どんな外部要因に影響を受けているかが理解できた。この論文の結果は、NFTに投資するために関する情報を提供し、意思決定に役立つ可能性がある。ただし、この結果は2020年のCryptokitties NFT市場に特定するので、次の年や異なる市場にも発生するかを確認する研究も必要です。