# The Fundamental Value of Art NFTs

Gilbert Fridgen* Roman Kräussl‡†

Orestis Papageorgiou* Alessandro Tugnetti‡

February 3, 2023

**Abstract**

This paper examines the level of speculation associated with art non-fungible tokens (NFTs), comprehends the characteristics that confer value on them and designs a profitable trading strategy based on our findings. We analyze 860,067 art NFTs that have been deployed on the Ethereum blockchain and have been involved in 317,950 sales using machine learning methods to forecast the probability of sale, the trade frequency and the average price. We find that NFTs are highly speculative assets and that their price and recurrence of sale are heavily determined by the floor and the last sales prices, independent of any fundamental value.

***Keywords—*** Non-fungible tokens (NFTs), Machine Learning, Fundamental Value, Speculation, Ethereum, Blockchain

***JEL—*** C55, G11, Z11

*SnT - Interdisciplinary Center for Security, Reliability and Trust, University of Luxembourg

†Hoover Institution, Stanford University

‡Department of Finance, University of Luxembourg

# 1 Introduction

Non-fungible tokens (NFTs) rose in prominence in 2021, amassing more than $44 billion in traded volume, attracting the attention of both institutional and retail investors (Chainalysis, 2022). In less than a year, NFTs went from being known to only a small community of blockchain enthusiasts to major news outlets devoting full-length articles to them, being described as a revolution in the art and the other industries (Wilson et al., 2021); even Collins Dictionary choosing it as the word of the year for 2021 (Collins Dictionary, 2021). Inevitably, such popularity also attracted substantial criticism about their utility and whether they can revolutionize the digital asset sector. This criticism further intensified in 2022, when the value of the majority of NFTs plummeted. Despite the hype subsiding and the trading volume falling, NFTs have recently found applications ranging from empowering business models innovation (Li & Chen, 2022) to serving as the primary component in several initial coin offerings (ICOs) used by firms to raise capital (Holden & Malani, 2022). Besides these applications, the most popular category of NFTs is their use as digital collectibles and artworks.

Probably the most influential art NFT sale came in March of 2021 with the auction house Christie's first-ever sale of a digital artwork of Beeple's "Everydays: The First 5000 Days" for $69.3 million, positioning the artist among the most valuable living artists. This, combined with other notable sales such as Pak's "The Merge", the most expensive NFT sold to date, and certain Cryptopunks selling for more than $10 million, propelled NFTs to the forefront of the art world in 2021, dominating the digital art space (Financial Times, 2021a) and generating almost as much volume as the traditional art space (Financial Times, 2021b). Even in 2022, when every other NFT category experienced a dramatic decrease in the number and value of sales, art NFTs have seen a slight increase in the average sale value (Basel, 2022).

Regardless of the growth, relatively little is understood regarding the attributes that lend value to art NFTs and influence their selling probability. Although scarcity is widely acknowledged, both by academics (Mekacher et al., 2022) and practitioners (Sothebys, 2021), as an essential trait, other factors, such as the impact of the artist, and the collection characteristics, remain vague, while academic research articles tend to focus on a few collections and attributes.

Specifically, focusing on CryptoPunks and using hedonic regression models, Kong and Lin (2021) and Schaar and Kampakis (2022) deduce that rarity plays a crucial role in determining their price. They find that in a market with a strong excess of demand, investors see NFTs with uncommon attributes as scarcer and, therefore, more expensive. Kong and Lin (2021) also observe that the price of CryptoPunks is influenced by the cryptocurrency market and that buyers assess their NFT investments in USD. Utilising the same approach and concentrating on the Decentraland collection, Goldberg et al. (2021) show that the value of parcels, plots of land with specific coordinates that comprise Decentraland's metaverse, depends on their coordinates. The authors point out that parcels with easy-to-remember coordinates or coordinates close to the centre of the metaverse tend to have higher prices.

Kireyev and Lin (2021) show that the price of the CryptoKitties NFTs declines over time as more NFTs are generated. They argue that the hedonic regression approach is not the optimal method for determining the price of NFTs and claim that models such as the gradient boosting machine (GBM) are superior since they can handle a potential selection bias much better. Using an approach based on hedonic regression and random forests, Horky et al. (2022) demonstrate that the complexity of the NFT's format (for example, whether the NFT is stored in a .mp4 or .jpeg file) and the market capitalization of the collection are strongly correlated with the price of the NFTs.

Analyzing multiple collections, Nadini et al. (2021) find a strong correlation between prices in the secondary market and those of the primary market per NFT and assert that their visual features are effective predictors of their price. Oh et al. (2022) indicate, that experienced investors systematically outperform inexperienced ones by an average of 10% on each sale since NFT collections purchased by experienced investors sell out more frequently and faster in primary markets and experience higher price appreciation in secondary markets. Finally, examining the dependence of the NFT market on external factors, Ante (2022), Dowling (2022) and Umar et al.

(2022) find a strong correlation between the NFT and cryptocurrency market, particularly with Bitcoin (BTC) and Ethereum (ETH).

The ambiguity caused by the lack of in-depth understanding and the significant correlation with the cryptocurrency market combined with the collapse of the NFT market has led many commentators to discredit the potential of NFTs in the art space, hypothesizing that their rapid growth was a consequence of excessive speculation. NFT enthusiasts, on the other hand, claim that these tend to be characteristics of most industries with a limited track record, and NFTs have many unique attributes, like provable digital scarcity, which have the potential to transform the art world.

The goal of this paper is to conduct a thorough analysis of various attributes, many of which have never been studied before, and their impact on the sale price and likelihood of sale for art NFTs. We focus on the entire art NFT market, analyzing every art collection and transaction performed on Ethereum, the blockchain with the most NFT transactions and home to many successful token offerings from firms (Gan et al., 2021), by utilising supervised machine learning methods such as K-Nearest Neighbours (KNN) and Random Forest (RF). Our sample comprises 860,067 art NFTs that have been deployed on the Ethereum blockchain and have been involved in 317,950 sales. Our results show that NFTs are highly speculative assets and that their price and likelihood of sale are heavily determined by the floor and the last sales prices, independent of any fundamental value. Based on our findings, we construct a profitable trading strategy.

The remainder of this paper is structured as follows. Section 2 discusses the data extraction and cleaning technique we used to gather and prepare the data for analysis and presents an exploratory data analysis that focuses on analyzing the speculation of the art NFT market. Section 3 describes the machine learning approach used to forecast the sale probability, sale frequency, and average price per sale of art NFTs. Section 4 discusses our findings and presents a suitable NFT trading strategy. Section 5 concludes.

# 2 Data

Blockchain related analyses are often classified into two broad categories: on-chain and off-chain. On-chain research involves data retrieved directly from the blockchain's public ledger, whereas off-chain analysis utilizes data sources outside the blockchain, such as price-tracking websites. For our analysis, we rely on both categories to obtain as much data as possible on NFTs and their pricing. For the on-chain part of our analysis, we use Dune and the APIs of Alchemy and OpenSea, while for the off-chain portion, we employ Etherscan, Discord's and Twitter's API and the price oracles of Chainlink.

Using Dune, we extract every on-chain transaction involving ERC-721 and ERC-1155 tokens on six of Ethereum's largest NFT marketplaces, specifically OpenSea, Rarible, NFTX, LooksRare, Foundation, and SuperRare. The extraction period spans from the establishment of the ERC-721 standard in January 2018 through February 2022. From the retrieved transactions, we identify 14,580 collections that have sold at least one NFT. Using Etherscan and OpenSea, we identify the deployers and the artists behind each collection and gather all the available metrics on a collection level, such as floor price, the total number of owners, collection name, number of transfers, the total volume traded in ETH and the artists' characteristics like gender and the total number of artists.

In a second step, we exclude the collections whose NFTs were involved in fewer than 30 sales to establish statistical significance within collections. Following the categorization criteria of Nonfungible.com, we divide the collections into six categories: art, collectibles, game, metaverse, utility, and other. Since our research focuses on art NFTs, we omit collectibles as they are comprised of a single artwork with minor variations per NFT. We also exclude NFTs associated with games and metaverses since their price is also influenced by the in-game or metaverse mechanics. Further, we remove utility NFTs such as liquidity provider positions in Uniswap V3 and other NFTs that do not fit in any of the preceding categories, like music NFTs or those with unique features.

Based on this dataset, we continue with the off-chain data extraction on Twitter and Discord. After manually identifying the Discord Servers and Twitter handles of the NFT collections (or in the case where these were missing their artists), we retrieve all publicly available metrics such as the number of followers, retweets, likes, replies, number of messages as well as discord messages and users that where involved in the general and announcement channels. At the end of the data extraction and cleaning on a collection level, our data set comprises 531 art NFT collections containing a total of 1,460,718 NFTs.

Using the Alchemy and OpenSea APIs, we retrieve all on-chain available information on these NFTs.[1] Consequently, we further exclude the NFTs whose data are missing, those that are in video or animated SVG formats and those with wrong data formats. In a final step, we download every file associated with each NFT and converted SVG images to PNG to allow for easier image analysis. At the end of this step, our sample contains 2.15 terabytes (TB) of PNG, JPEG and GIF files of 860,067 NFTs.

From the extracted files, we determine their colour proportions and the total number of colours present by utilizing the Hue, Saturation, Value (HSV) colour model[2]. We group every colour around nine commonly used colours: black, white, grey, red, green, blue, yellow, purple, and orange. In addition, we assess the complexity of the files using Shannon's entropy and the block decomposition method (BDM), with the use of the coding theorem method (CTM), because it is regarded as one of the best approaches both from a statistical and an algorithmic point of view (Zenil, 2020). In the case of GIF files, we average the values of each frame.

So-called wash trading relates to a form of market manipulation where a small number of investors repeatedly buy and sell the same asset, generating an inorganic market activity. It creates artificial trading volume and gives the appearance that the asset is more in demand than it actually

[1] We opted for the use of both APIs because we identified that each has missing data. By utilizing both, we were able to cover most of the gaps.

[2] HSV is a mathematical abstraction that describes how colours can be represented by electronic systems. In this model, colours are represented by a cylindrical geometry whose angle represents the hue, the x-axis determines saturation, and the value depends on the y-axis.

is. Wash trading can greatly affect the price, the traded volume and the selling frequency (Cong et al., 2022). We exclude in our analysis all individual NFTs that have been involved in wash trading and all collections whose trading volume is greater than 90% due to wash trading, such as Terraforms by Mathcastles. Additionally, we eliminate NFTs whose first sale took place after 01-01-2022, since more than 75% of the total volume in January 2022 and 55% of the total volume in 2022 was a result of wash trading and identifying the genuine transactions would be complicated and risky.

Using Dune, we determine that 210,384 of the aforementioned art NFTs were involved in 317,950 sales. Based on these sales, we generate two data sets, one containing data pertaining to every sale of the extracted art NFTs, and the other comprising aggregated data per NFT, which was augmented with data from Discord and Twitter, and colour and image complexity data. In both data sets, the sale price is measured in USD, for which we utilize the exchange rates provided by Chainlink at the time of the sale for the individual NFT sales and at the end of the day on February 10, 2022, for the collection floor price. Table 1 contains the extracted variables and related descriptive statistics.

[Table 1 about here.]

Table 1 shows that the distribution of most variables is negatively skewed; for many of them, this remains around 0 up to the second quartile and then increases exponentially towards the third and fourth quartiles. Looking at, *number_of_trades*, which counts the number of sales for each NFT, one of the dependent variables that will be used during the analysis, we see that the majority of NFTs have never registered a sale. To note the first sale, it is necessary to look at the top 25% of the distribution. Furthermore, only a residual part of them (less than 25%) has undergone more than one transfer. All these are indications of an illiquid and fragmented market.

Following data extraction, we rely on exploratory data analysis (EDA) to identify key characteristics of our dataset. By analyzing the number of sales and sales volume in USD per year, we observe that art NFTs followed the overall market trend and skyrocketed in 2021. Specifically, 2018 and 2019 account for just 0.72% of the number of sales and 0.02% of the USD volume, while 2021 accounts for 93.88% of sales and 99.21% of the volume. Focusing on 2020 and 2021, we see a similar trend when examining the average monthly volume and sales price. The volume exhibits a strong uptrend for both years, however, the average price, despite initially following a similar pattern, starts plateauing after February 2021. This indicates that the volume increase in the art market in 2021 was mostly due to an increase in the number of sales and not in price.

We note that 71.9% of the examined NFTs have never been sold. Of those sold, only 22.1% were sold more than once, indicating that the art NFT market is highly illiquid. However, NFTs that are sold more than once have a higher chance of being sold a third time, with the probability increasing to 27.1%, which further increases to 30.71% for the fourth sale. Investigating further the differences between single-sold and multi-sold NFTs, we focus on the average number of days it takes for an NFT to be sold since its creation and between sales. We observe that the first sale for both cases follows a similar pattern with a considerable decline in 2021, suggesting that investors purchased newer collections during the NFTs' peak demand. However, the number of days decreases significantly for the subsequent sales in the case of multi-sold NFTs, with each of the succeeding sales taking progressively less time on average. Following this, we analyze the price differences between sales and notice that all of them follow a similar pattern, with the profit margin increasing significantly in 2021 but decreasing with every subsequent sale. Both observations indicate that multi-sold NFTs have distinct characteristics compared to single-sold NFTs, prompting us to analyze them separately.

To analyse the returns of the two NFT categories, we calculate the 30-day rolling mean cumulative log-returns and compare them in Figure 1a against the performance of the following six indexes: S&P 500, XAUUSD BGN Currency, PIMCO CORP US Equity, DXY Currency, CL1 COMB Commodity, and Bloomberg Galaxy Crypto Index (BGCI). We observe that the returns of both groups experienced a substantial increase in late 2020, with the returns of multi-sold NFTs surpassing those of single-sold in early 2021 and maintaining the lead thereafter. When compared to the BGCI, we see that the returns of art NFTs move in similar patterns, with both categories having higher returns than the index for most of 2021 and multi-sold NFTs recording 51.82% higher median returns in 2021, while those of single-sold were higher by 14.33%.

[Figure 1 about here.]

Fueled by these findings, we further subdivide the two categories into the top 5, 10, 20 and 50 collections by average NFT selling price. In the case of single-sold art NFTs, we observe returns that are substantially greater than those of the BGCI (with top-5 collections recording 1,65% higher median returns than the BCGI in 2021), with returns decreasing as more NFT collections, are included. On the other hand, we see that the examined subcategories for multi-sold NFTs experience higher return fluctuations, particularly in the top-5 category, with all of them outperforming the BGCI in 2021 and the top-20 NFT collections outperforming every other multi-sold subcategory, having 303.2% higher median returns than the BGCI for 2021. Comparing the subcategories between them, it is apparent that every top single-sold collection outperformed multi-sold collections and had lower fluctuations for almost the entirety of our analysis. Moreover, top single-sold subcategories started generating higher returns than the BGCI, already as early as the end of 2020, which is far earlier than when the great interest in NFTs started. Figure 1a shows that top single-sold NFTs outperform multi-sold ones, but when compared to the aggregate performance of the two main categories, the performance is inverted. This is due to the fact that there are more single-sold NFTs with lower prices in lower percentiles, though the difference in price changes at the 97th percentile, where the price of single-sold NFTs surpasses multi-sold ones.

These returns, however, diminish when risk is taken into account. Using the Sharpe ratio to calculate risk-adjusted returns, the NFT returns are essentially flat around zero, and every index outperforms all categories, as a result of the excessive volatility relative to the other categories. In order to assess the volatility of art NFTs and their correlation with the cryptocurrency market further, we conduct the standard CAPM beta-based volatility analysis. We use the daily returns of art NFTs and estimate the beta over a 180-day rolling window using BTC and ETH as benchmark assets since prior academic studies indicate a correlation between the NFT market and the cryptocurrency market.

Figure 1b shows that multi-sold NFTs, for the most part, exhibit a substantially higher degree of volatility compared to both benchmarks, in particular, BTC. Unsurprisingly the volatility peaked in 2021 along with the overall interest in the NFT market and began to decline in the second half of that year. Single-sold NFTs show a different narrative: For most periods, the returns are either uncorrelated or inversely correlated with benchmark returns. In 2021, when they become correlated, the single-sold returns are considerably less volatile than those of the benchmarks. As in the case of multi-sold NFTs, there is a higher correlation with BTC returns, although they exhibit a lower volatility than them.

Our findings show that the increased interest in both NFT categories led to higher volatility. Multi-sold NFTs are substantially more volatile than single-sold NFTs, which are predominantly uncorrelated or inversely correlated to the benchmarks. It is possible that multi-sold NFTs are considered as more speculative assets, and their demand is heavily influenced by market sentiment, while single-sold ones are viewed as collectibles. In both cases, the returns are more correlated to those of BTC, which is unexpected given that the majority of NFT sales are conducted using ETH. The findings of Kong and Lin (2021) that investors typically evaluate NFT investments in USD may provide one possible explanation for this phenomenon.

# 3 Approach

This section investigates the determinants of an art NFT's value. In a first step, we analyze the factors that determine the likelihood of an NFT being sold using the entire data set described in the previous section. In a second step, by looking only at the NFTs sold, we get a prediction of the total average value for each NFT as the product between the expected number of transfers per NFT (frequency) and the average expected price per transfer (severity).

We redesign the model applied by Henckaerts et al. (2020), who analyze exactly the frequency and severity dimensions to price the pure premium for a non-life insurance policy. Figure 2 shows that the statistical distributions of the dependent variables are very similar to those in Henckaerts et al. (2020).

[Figure 2 about here.]

Specifically, the total average value of an NFT ($\pi$) will be decomposed into:

$$\pi = \mathbb{E}(N)\,\mathbb{E}\left(\frac{P}{N}\middle|N\right) = F * S, \tag{1}$$

where $N > 0$ is the number of sales and $\frac{P}{N}$ represents the price per transfer, while $F$ and $S$, assumed independent, are the expected NFT frequency and severity, respectively.

For the predictive models, one or more of the hypotheses of the linear regression model would be too restrictive when pricing instruments like NFTs. Thus, we use in the following generalized linear models (GLM):

- **Binomial GLM** with log-link function for the probability of selling NFTs given the dichotomous nature of the variable.

- **Negative binomial GLM** with log-link function for the expected frequency $\mathbb{E}(F)$. In this case, our objective is to model the probability distribution of the number of transfers affecting the NFT over a year. As a result, the dependent variable is a counting variable. Figure 2a indicates high (over-) dispersion of the data. Therefore, it is appropriate to model it as a negative binomial distributed random variable. This choice implies some assumptions regarding claims: the probability of transferring a particular NFT is the same for every day during a given period, the occurrence of a transfer is independent of all transfers that might have occurred before for that particular digital asset, and the probability of filing two or more transfers in a sufficiently small time interval is negligible.

- **GLM** with a log-link function, that is a log-normal GLM for the expected severity $\mathbb{E}(S)$. To obtain an efficient pricing model, it is important to provide good estimates of the severity, which are definitely more difficult to evaluate than frequencies due to some problems, e.g., the number of NFTs effectively sold is very limited. The data available are few and the selling amounts are for the great majority of a limited amount, while those of a particularly high amount are a small number. These issues often make estimates from severity models less reliable compared to estimates relating to the expected frequency. The typical characteristics of loss distributions are the right skewness and the presence of a heavy right tail; so, we must choose a distribution that captures these features in such a way as to reflect as much as possible the intrinsic nature of the data. To do this, we use a Kolmogorov–Smirnov test on the average price variable, comparing it to 105 different theoretical continuous distributions. As a result, none of them seems to fit the empirical distribution of the average NFT price. For this reason, given the bell-shaped distribution of the log transformed average price in Figure 2b and knowing that in the literature, the two most used distributions are the gamma and the log-normal (de Jong and Heller, 2008), we choose the latter to model the average price per NFT.

Nadini et al. (2021) and Kireyev and Lin (2021) show that machine learning (ML) algorithms are particularly suitable for capturing the high volatility patterns typical of the NFT market, more than a simple linear model would do. We also challenge the predictive performances of GLMs with non-parametric models. We use two supervised machine learning algorithms, the k-nearest neighbors (KNN) (Cover and Hart, 1967) for the probability of selling NFTs and the random forest (Breiman, 2001) for the expected number of transfers (*i.e.*, $F$ in Equation (1)) and the expected average price per transfer (*i.e.*, $S$ in Equation (1)). Supervised learning consists of defining the input and output data, leaving the system with the task of identifying the function that associates its correct answer to each input data. The goal is to find an approximation of the function that generated the output data.

While KNN is more often used to solve classification problems, as the algorithm is based on predicting the class of new data based on the labels of the neighbors data closest to it, RF can be used for both classification and regression techniques. In particular, RF can be represented as a collection of regression trees (Breiman et al., 1984) where the bias-variance trade-off (Hastie et al., 2009) is efficiently optimized by randomly selecting a subset of variables, known as candidates, at each split to potentially be the optimal splitting variable. Specifically, at each split, $m$ variables are randomly chosen out of the total $p$ variables available. In the end, the prediction made by the random forest will be the average between the predictions made by the single trees.

# 4 Discussion of Results

Figure 3 describes the overtime trend of the monthly average floor price for each NFT considered in our analysis. We note a sharp increase in price starting from the beginning of 2021, until reaching the peak in April of the same year. We decide to introduce a structural-break point on March 11, 2021.[3] Our assumption is that the factors that drive the popularity of a digital asset in the two time periods are different.

[Figure 3 about here.]

After having obtained statistical evidence of the presence of a structural-break at that time point through the Chow (1960) test, we fit all our econometric models on three detached time series:

- the entire reference period;
- sold and not sold NFTs created between July 13, 2018 and March 11, 2021 (29.24% of the total number of observations),
- all the collections created after March 11, 2021 also including NFTs whose collection was created before that date but were sold after the cutoff until February 2022. (70.76% of the total number of observations).

[3]This date corresponds to the sale of Beeple's "Everyday: The First 5000 Days" at a Christie's auction for 69.3 million USD.

For the purpose of calculating the goodness of fit and accuracy measures, we divide the resulting data sets into two parts: a *training* set (80% of the data) on which we fit and train the econometric models and a *test* set (the remaining 20%) used to make predictions and calculate regression errors.

Table 2 presents the results of the binomial GLM regression for the three time series considered. The probability of selling an NFT is heavily influenced by the performance of the Bloomberg Galaxy Crypto Index, especially for the period before March 11, 2021. In accordance with the literature, our findings confirm the dependency relationship between cryptocurrencies and NFTs as the only assets (especially before 2021) in which it was possible to spend cryptocurrencies on Ethereum.

[Table 2 about here.]

Table 3 analyzes the coefficients of the regressors by eliminating the influence of the variable relating to the Bloomberg Galaxy Crypto Index[4]. As a result, the significant covariates are different depending on the time period taken into consideration. We note that especially for the period after March 2021, belonging to a collection between Foundation (FND), Editional, KnownOrigin, or SuperRare (*marketplace_collection*) are associated with the highest probability of being sold. We find an inverse relation that investors have towards the social networks Twitter and Discord: while the fact of having associated a Twitter account (*twitter_account*) to the NFT collection increases the chances of sale, the inverse happens for the Discord account (*discord_account*). However, when the focus on the activity of the two social networks shifts, the trend reverses. For instance, the number of tweets per account (*num_tweet*) has a negative effect on sales while the number of monthly active users on the "General" channel of the Discord server (*d_gen_users_month*) has a positive one. This indicates that investors who decide to invest in NFTs consider the presence of a Twitter account relating to the artist attractive and, thus, a valid purchase. We also find that if this becomes too long-winded and repetitive on the social network, collectors lose interest. Instead, investors have a much harder time finding NFT collections that have a Discord account attached.

[4] All details on robustness tests for the complete time series are reported in the Appendix in Table A2

Nonetheless, when they do join, they feel part of a community closer to the artist who will benefit from a better chance of selling their NFTs.

[Table 3 about here.]

In order to improve the accuracy of the categorization between sold and unsold NFTs, we decide to apply in the following a KNN model. Although it is not possible to investigate the magnitude of the single covariates on the dependent variable[5], we are able to greatly improve the accuracy of the prediction, defined as the number of correct predictions made divided by the total number of predictions made. The latter increases from 61.97%, 85.08% and 75.66%, respectively, for the complete historical series, the one before March 11, 2021, and the one after the cutoff for the binomial GLM, to 75.22%, 91.84% and 78.21%, respectively, of the KNN model. For each time series taken into consideration, the accuracy of the KNN models is always higher than that of the GLMs.[6]

Table 4 and Table 5 show the results for the negative binomial and for the log-normal GLM, respectively, for the various time series considered[7].

[Table 4 about here.]

[Table 5 about here.]

Surprisingly, we find that the Bloomberg Crypto Index plays no role in determining frequency and severity. Therefore, combining the result obtained with Table 2, we find that the Bloomberg Crypto Index is essential only in the preliminary phase of entering the NFT market. This result is helpful for investors' understanding when is the most profitable moment to convert their cryptocurrency investment into non-fungible tokens. We find that when the first NFT trade takes

[5]Supervised ML methods do not return as output the "betas" we are used to in regressions.

[6]We refer to Figure 5b later in the paper which provides additional information on the comparison of the binomial GLM and KNN models through the use of the receiver operating characteristic (ROC) curve. The ROC curve is a probability curve which represents the degree of separability between classes for a classification algorithm.

[7]All details on robustness tests for the complete time series are reported in the Appendix in Table A3 for the frequency and in Table A4 for the severity models.

place, the covariates that move their popularity on the market, in terms of the number of transfers between investors and average price, deviate from the trend of digital currencies. More precisely, being subject to ERC-721 has one of the biggest (negative) impacts on the expected frequency. This is likely due to the very nature of the protocol. The ERC-721 is used for the creation of NFTs with the pure non-fungibility feature, while the ERC-1155 is used for the creation of hybrid NFTs that can store both the non-fungibility and the fungibility features. The latter makes NFTs subject to ERC-1155 easier to transfer than ERC-721.

We find that both frequency and severity models agree that the floor and the last selling NFT prices are the main statistically relevant variables for the determination of the number of sales and the average price in all the time series taken into consideration. No variable regarding the artist's social network activity or the intrinsic characteristics of the artwork is statistically significant, i.e., seems to be relevant. Our results show that agents are moved by speculative possibilities when it comes to investing in this market, rather than by belonging to a community of investors around a single NFT author.

To confirm the latter result and to improve the predictive performance of the expected frequency and severity, we choose to use a supervised ML model: the random forest (RF). Also, in this case, this method does not provide for the calculation of the magnitudes of the effects of the covariates on the dependent variable. For this reason, we report in Figure 4 the graphs relating to the variable importance of the two models.[8]

[Figure 4 about here.]

We note the large difference in terms of the number of important variables, which separates Figure 4a and Figure 4b. This is consistent with the existing literature in non-life insurance policy modeling (Henckaerts et al., 2020): Severity is typically much harder to predict than the frequency,

[8]The importance of a variable in a RF is determined by the mean decrease in impurity (or Gini importance) mechanism. This measures the improvement in the split criterion at each split in each tree, and is calculated by summing the importance attributed to the splitting variable across all the trees in the forest for the average number of transfers and the average price per NFT considered.

resulting in fewer determinant variables for the former with respect to the latter. The last transaction price (*last_price*) and the floor price are confirmed to be the preponderant variables in determining both the average number of transfers and their average price. We find that the frequency dimension is mainly affected by these two variables, adding evidence to the assumption that speculation is the main reason why investors enter the NFT market.

Figure 5a illustrates the comparison of frequency - severity models with respect to the predictive performance using the mean squared error (MSE) calculated on the test set as an out-of-sample testing measure. We note that we are able to achieve better performance, particularly in predicting severity. In the prediction of expected frequency, we observe a test MSE mean percentage reduction of 31.70% (with the largest improvement being for the model using data after the time cutoff, which dropped from 98.75 for the GLM to 8.37 for the RF) across the three time series considered. In the prediction of expected severity, this reduction is even more substantial at 99.95%.

[Figure 5 about here.]

Figure 5b shows the ROC curves belonging to the binomial GLM and KNN for each of the tested data sets. Generally, the curve for a binary classification problem plots the true positive rate as a function of the false positive rate. The points of the curve are obtained by sweeping the classification threshold from the most positive classification value to the most negative. For a completely random classification, the area under the curve (AUC) corresponds to the dotted red line. We find that for each time series, the accuracy of the KNN models is consistently better than that of the GLMs.

It is not possible for the RF technique to establish the magnitude of the effect of the independent variables on the dependent one. However, we can investigate the marginal effect of one feature (given the others) on the predicted outcome of our machine learning model using the partial dependence plot (PDP) in Friedman, 2001). Figure 6a shows the PDP function of the ”*last_price*” variable for the frequency and severity RF.

[Figure 6 about here.]

By bootstrapping our data 50 times , we generate the gray bands visible in Figure 6a. These individual conditional expectations (ICE) allow us to evaluate in a more granular way the effect of "*last_price*" on the average price, taking into account a different set of data at each trial. The average of these ICEs is the PDP marked with a solid red line, while the bootstrap standard errors (SE) are also highlighted in the figure.

These results confirm our previous findings in Figure 4: the estimate of the effect of the last price on the average price in severity brings less uncertainty (in terms of bootstrap SE) than frequency, given the high importance that this variable plays in the former than the latter. We find that both the average number of sales and the average price per NFT are positively influenced by the increase in the last recorded price per token. In particular, it seems that the sudden growth for both values stops at the value of 20,000 USD. Indeed, after this amount, both measures appear to have a slower trend upwards to around USD 80,000, when then a further bump up is visible for the average number of sales. All this is plausible: the least expensive NFTs tend to be the ones that are transacted more often because they find a demand to satisfy given their price and the high interest in these digital assets during the period under consideration. When these become too expensive (around 20,000 USD) they become very difficult to sell, and their price rises with less speed.

In a final step, we aim to build a profitable NFT trading strategy. Since our findings indicate that the price (floor and last) is the only factor that influences the future price and probability of sale of art NFTs, we rely entirely on market conditions to identify possible strategies. The impact of past volatility is widely acknowledged as a crucial factor behind stock, and option returns (Ang et al., 2009; Goyal & Saretto, 2009; Lee et al., 1994). To formulate our strategy, we follow the approach in Bollen and Whaley (2004), focusing on art NFT collections that are most actively traded and have an established track record to prevent cases where collections with a limited number of sales outperform the market but are unable to repeat this performance when the sales numbers increase. Therefore, we focus on collections with more than 60 sales during a two months period. Within these collections and for each of the 24 months of our analysis, we determine the top-5, top-10 and top-20 collections with respect to the past month's highest and lowest volatility and evaluate the performance of the selected collections for the subsequent month. Essentially under this strategy, investors would look into the volatility of art NFT collections of the past month and buy NFTs from the collections with the highest/lowest volatility rebalancing their NFT portfolio every month. Our approach does not take into consideration the gas fees required for each transaction on a blockchain since they can vary greatly on any given day. However, we expect it to be realistic, given that fees are fixed per transaction and do not fluctuate depending on the value of the transaction as in traditional markets.

To assess the performance of our strategy, we calculate the 90-day rolling mean cumulative log-returns. The extended rolling window allows us to smooth out some of the extreme market fluctuations that emerged in 2022. For 2022, the data set is constructed based on the NFT sales included in the 531 collections previously extracted as described in Section 2. Table 6 shows our results and provides detailed information about the performance of our strategy.

[Table 6 about here.]

Table 6 shows that in 2021, single-sold and multi-sold NFTs exhibit comparable return patterns, beginning to exceed the BGCI in July 2020 and May 2020, respectively, which was reversed in March 2022. We find that the strategy based on selecting the top-5 NFTs of the previous month displays substantial return volatility, and underperforms the BGCI in terms of returns. Comparing the top-10 and top-20 strategies, their returns appear to be similar in 2021, with the latter having lower volatility under each and every scenario. In 2022, however, the volatility pattern shifts, with the volatility of the top-20 strategy being equivalent to that of the top-10 strategy. When comparing the high and low return scenarios, we see that the former provides better returns for single-sold NFTs with higher volatility. Looking into the performance of the strategy between the two years reveals that in most circumstances, it outperforms BGCI in 2021, which is the diametric opposite of its performance in 2022. The sole strategy that outperforms BGCI for both years is based on buying single-sold art NFTs from the top-10 collections that had the highest volatility in the previous month. Surprisingly, the result about the existence of a positive relationship between idiosyncratic risk and expected returns when investors are holding imperfectly diversified portfolios is consistent with the literature on stock investments (Campbell et al., 2001; Fu, 2009).

# 5 Conclusions

We analyze the value determinants and level of speculation associated with art NFTs. Utilizing the largest, to the best of our knowledge, data set used and employing both technical indicators and machine learning methods, we conclude that art NFTs are highly speculative assets with little additional diversification benefits. We find that the most influential variables in determining the frequency of transactions and the average sale price of each NFT are the collection's floor price as well as the most recent selling price.

Based on these findings, we implement a trading strategy that monthly rebalances an NFT portfolio using the NFTs with the highest and lowest volatility of the preceding month. This strategy generates returns that surpass the BGCI in 2021 but severely underperform the crypto market in 2022. We show that buying single-sold NFTs from the top-10 collections that displayed high volatility in the past month yields higher returns than BGCI for both years.

On a more technical level, we demonstrate that the examined machine learning techniques (KNN and RF) produced more accurate results than the classical econometric models, allowing us to achieve an average 31.68% improvement in prediction accuracy.

Overall, our research validates the beliefs of NFT sceptics, indicating that the artistic merit of art NFTs is limited, and the driving force behind them seems to be primarily speculative. From a more general perspective, our findings appear to validate a prevalent belief in the crypto space as a whole, that due to the unregulated nature of the industry, cryptocurrency-related endeavours inevitably attract swindlers during bull markets that do more harm than good both in terms of reputation and finances. However, this does not imply that art NFTs have no artistic worth. With the hype subsiding and most bad actors out of the space, NFT enthusiasts could potentially stir the art NFT space into more creative paths that would benefit both artists and enthusiasts alike.

# References

Adams, R. B., Kräussl, R., Navone, M., & Verwijmeren, P. (2021). Gendered Prices. *The Review of Financial Studies*, *34*(8), 3789–3839.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2009). High idiosyncratic volatility and low returns: International and further U.S. evidence$. *Journal of Financial Economics*.

Ante, L. (2022). The non-fungible token (NFT) market and its relationship with Bitcoin and Ethereum. *FinTech*, *1*(3), 216–224.

Basel, A. (2022). Collectors are spending more, not less, on NFT art, according to the Art Market Report. Retrieved December 14, 2022, from https://www.artbasel.com/stories/art-market-report-2022-collectors-cryptocurrency-nfts

Bollen, N. P. B., & Whaley, R. E. (2004). Does Net Buying Pressure Affect the Shape of Implied Volatility Functions? *The Journal of Finance*, *59*(2), 711–753.

Breiman, L., Friedman, J., Stone, C., & Olshen, R. (1984). Classification and regression trees. *Brooks, Cole Publishing*.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32.

Campbell, J. Y., Lettau, M., Malkiel, B. G., & Xu, Y. (2001). Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk. *The Journal of Finance*, *56*(1), 1–43.

Chainalysis. (2022). The Chainalysis 2021 NFT market report.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, *28*(3), 591–605.

Collins Dictionary. (2021). Get your crypto at the ready: NFTs are big in 2021. Retrieved December 7, 2022, from https://blog.collinsdictionary.com/language-lovers/get-your-crypto-at-the-ready-nfts-are-big-in-2021/

Cong, L. W., Li, X., Tang, K., & Yang. (2022). Crypto Wash Trading. *National Bureau of Economic Research (NBER) Working Paper Series*, (30783).

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, *13*(1), 21–27.

de Jong, P., & Heller, G. Z. (2008). Generalized Linear Models for Insurance Data. *Cambridge Books*.

Dowling, M. (2022). Is non-fungible token pricing driven by cryptocurrencies? *Finance Research Letters*, *44*.

Financial Times. (2021a). How digital art dominated 2021.

Financial Times. (2021b). How NFTs became a $40bn market in 2021.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232.

Fu, F. (2009). Idiosyncratic risk and the cross-section of expected stock returns. *Journal of Financial Economics*, *91*(1), 24–37.

Gan, J. (, Tsoukalas, G., & Netessine, S. (2021). Initial Coin Offerings, Speculation, and Asset Tokenization. *Management Science*, *67*(2), 914–931.

Goldberg, M., Kugler, P., & Schär, F. (2021). Land Valuation in the Metaverse: Location matters. *SSRN Scholarly Paper*, 3932189.

Goyal, A., & Saretto, A. (2009). Cross-section of option returns and volatility. *Journal of Financial Economics*, *94*(2), 310–326.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. *Springer Series in Statistics*.

Henckaerts, R., Côté, M.-P., Antonio, K., & Verbelen, R. (2020). Boosting insights in insurance tariff plans with tree-based machine learning methods. *North American Actuarial Journal*, *25*(2), 255–285.

Holden, R., & Malani, A. (2022). An Examination of Velocity and Initial Coin Offerings. *Management Science*, *68*(12), 8515–9218.

Horky, F., Rachel, C., & Fidrmuc, J. (2022). Price Determinants of Non-fungible Tokens in the Digital Art Market. *Finance Research Letters*, 103007.

Kireyev, P., & Lin, R. (2021). Infinite but Rare: Valuation and Pricing in Marketplaces for Blockchain-Based Nonfungible Tokens. *INSEAD Working Paper Series*, *2020*(58).

Kong, D.-R., & Lin, T.-C. (2021). Alternative Investments in the Fintech Era: The Risk and Return of Non-fungible Token (NFT). *SSRN Journal, ID 3914085*.

Lee, C. M. C., Ready, M. J., & Seguin, P. J. (1994). Volume, Volatility, and New York Stock Exchange Trading Halts. *The Journal of Finance*, *49*(1), 183–214.

Li, S., & Chen, Y. (2022). How non-fungible tokens empower business model innovation. *Business Horizons*.

Mekacher, A., Bracci, A., Nadini, M., Martino, M., Alessandretti, L., Aiello, L. M., & Baronchelli, A. (2022). Heterogeneous rarity patterns drive price dynamics in NFT collections. *Scientific Reports*, *12*(1), 13890.

Nadini, M., Alessandretti, L., Di Giacinto, F., Martino, M., Aiello, L. M., & Baronchelli, A. (2021). Mapping the NFT revolution: Market trends, trade networks, and visual features. *Scientific Reports*, *11*(1), 20902.

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, *135*(3), 370–384.

Oh, S., Rosen, S., & Zhang, A. L. (2022). Investor Experience Matters: Evidence from Generative Art Collections on the Blockchain. *BFI Working Paper Series*, *2022*(116).

Schaar, L., & Kampakis, S. (2022). Non-Fungible Tokens as an Alternative Investment: Evidence from CryptoPunks. *The Journal of The British Blockchain Association*.

Sothebys. (2021). NFTs: Redefining Digital Ownership and Scarcity. Retrieved December 7, 2022, from https://www.sothebys.com/en/articles/nfts-redefining-digital-ownership-and-scarcity

Umar, Z., Gubareva, M., Teplova, T., & Tran, D. K. (2022). Covid-19 impact on NFTs and major asset classes interrelations: Insights from the wavelet coherence analysis. *Finance Research Letters*, 102725.

Wilson, K. B., Karg, A., & Ghaderi, H. (2021). Prospecting non-fungible tokens in the digital economy: Stakeholders and ecosystem, risk and opportunity. *Business Horizons*, *65*(5), 657–670.

Zenil, H. (2020). A Review of Methods for Estimating Algorithmic Complexity: Options, Challenges, and New Directions. *Entropy*, *22*(6).

**Table 1. Descriptive Statistics.** This table presents the descriptive statistics for the numerical variables related to the 860,067 observations in our NFT database. Nominal and ordinal categorical variables are reported first, while continuous variables follow. For the former, the number of unique occurrences, the mode and the frequency of the mode are represented, while the latter are summarized by their mean, standard deviation and the quartiles of the statistical distribution.

| Variable | Unique | Mode | Mode Freq. | Mean | Standard Dev. | Min | 25% | 50% | 75% | Max |
|---|---|---|---|---|---|---|---|---|---|---|
| *year_collection_creation* | 5 | 2021 | 680662 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| *month_collection_creation* | 12 | 9 | 201455 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| *year_last_sale* | 6 | 2021 | 627592 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| *month_last_sale* | 12 | 1 | 658818 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| *day_last_sale* | 31 | 1 | 634512 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| *hour_last_sale* | 24 | 0 | 638156 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| *nft_type* | 2 | erc721 | 778622 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| *platform_last_sale* | 7 | 0 | 627593 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| *sale_occurrence* | NaN | NaN | NaN | 0.28 | 0.45 | 0 | 0 | 0 | 1 | 1 |
| *twitter_account* | NaN | NaN | NaN | 0.97 | 0.17 | 0 | 1 | 1 | 1 | 1 |
| *discord_account* | NaN | NaN | NaN | 0.55 | 0.50 | 0 | 0 | 1 | 1 | 1 |
| *marketplace_collection* | NaN | NaN | NaN | 0.33 | 0.47 | 0 | 0 | 0 | 1 | 1 |
| *verified* | NaN | NaN | NaN | 0.27 | 0.44 | 0 | 0 | 0 | 1 | 1 |
| *num_of_colors* | NaN | NaN | NaN | 6.54 | 2.44 | 1 | 4 | 7 | 9 | 9 |
| *green_%* | NaN | NaN | NaN | 0.07 | 0.16 | 0 | 0 | 0 | 0.06 | 1 |
| *yellow_%* | NaN | NaN | NaN | 0.03 | 0.10 | 0 | 0 | 0 | 0.02 | 1 |
| *purple_%* | NaN | NaN | NaN | 0.04 | 0.12 | 0 | 0 | 0 | 0.02 | 1 |
| *black_%* | NaN | NaN | NaN | 0.21 | 0.30 | 0 | 0 | 0.05 | 0.29 | 1 |
| *white_%* | NaN | NaN | NaN | 0.13 | 0.25 | 0 | 0 | 0.01 | 0.13 | 1 |
| *gray_%* | NaN | NaN | NaN | 0.11 | 0.20 | 0 | 0 | 0.01 | 0.11 | 1 |
| *blue_%* | NaN | NaN | NaN | 0.14 | 0.22 | 0 | 0 | 0.03 | 0.18 | 1 |
| *orange_%* | NaN | NaN | NaN | 0.13 | 0.21 | 0 | 0 | 0.03 | 0.16 | 1 |
| *red_%* | NaN | NaN | NaN | 0.13 | 0.20 | 0 | 0 | 0.04 | 0.17 | 1 |
| *number_of_trades* | NaN | NaN | NaN | 0.41 | 3.52 | 0 | 0 | 0 | 1 | 897 |
| *log_avgprice* | NaN | NaN | NaN | 0.82 | 1.35 | 0 | 0 | 0 | 2.18 | 6.38 |
| *d_ann_messages_month* | NaN | NaN | NaN | 6.71 | 12 | 0 | 0 | 0 | 10 | 63 |
| *d_ann_users_month* | NaN | NaN | NaN | 0.76 | 1.22 | 0 | 0 | 0 | 2 | 13 |
| *norm_shannon_entropy* | NaN | NaN | NaN | 0.26 | 0.13 | 0 | 0.16 | 0.30 | 0.36 | 0.52 |
| *d_gen_messages_month* | NaN | NaN | NaN | 2851 | 9110 | 0 | 0 | 0 | 1450 | 64984 |
| *d_gen_users_month* | NaN | NaN | NaN | 123.12 | 289.86 | 0 | 0 | 0 | 101 | 1997 |
| *timediff* | NaN | NaN | NaN | 47.24 | 153.12 | 0 | 0 | 0 | 2.18 | 1306.70 |
| *total_price* | NaN | NaN | NaN | 1359.45 | 18258.95 | 0 | 0 | 0 | 162.37 | 2397320 |
| *avgprice* | NaN | NaN | NaN | 984.39 | 14067.94 | 0 | 0 | 0 | 149.67 | 2397320 |
| *floor_price* | NaN | NaN | NaN | 874.19 | 13694.19 | 0 | 0 | 0 | 122.60 | 2397320 |
| *last_price* | NaN | NaN | NaN | 1066.30 | 14874.08 | 0 | 0 | 0 | 142.67 | 2397320 |
| *listed_count* | NaN | NaN | NaN | 431.65 | 1142.84 | 0 | 3 | 27 | 155 | 45918 |
| *quote_count_month* | NaN | NaN | NaN | 142 | 296 | 0 | 13 | 51 | 185 | 6139 |
| *retweet_count_month* | NaN | NaN | NaN | 988.64 | 4953.95 | 0 | 48 | 263 | 820 | 124875 |
| *reply_count_month* | NaN | NaN | NaN | 667.59 | 1088.98 | 0 | 28 | 168 | 1083 | 17379 |
| *following_count* | NaN | NaN | NaN | 1109.30 | 2021.15 | 0 | 120 | 489 | 854 | 17650 |
| *num_tweets* | NaN | NaN | NaN | 2592 | 4855.60 | 0 | 174 | 668 | 2397 | 70961 |
| *like_count_month* | NaN | NaN | NaN | 6233.58 | 20480 | 0 | 444 | 1190 | 7713 | 417039 |
| *floor_price_collection* | NaN | NaN | NaN | 1703.34 | 16719.65 | 0 | 22.69 | 126.40 | 388.92 | 631995 |
| *followers_count* | NaN | NaN | NaN | 114353.58 | 396975.35 | 0 | 2641 | 8042 | 255317 | 16922128 |
| *BGCI* | NaN | NaN | NaN | 2139.77 | 1042.72 | 2191 | 973 | 218 | 1414 | 3870 |
| *BDM* | NaN | NaN | NaN | 64736.10 | 139913.40 | 39.37 | 8325.38 | 24846.97 | 60066.57 | 2295468.97 |

**Table 2. Binomial GLM regression results**. This table presents the coefficients of the GLM regression relative to the probability of an NFT being sold divided by the observed time series: from July 2018 to 10th February 2022, from July 2018 to 11th March 2021 and from 12th March 2021 to 10th February 2022. The number of observations is 860,067, and the dependent variable is *sale_occurrence*, which assumes 0 value for non-sold NFT and 1 vice versa. Variable selection was performed using the Variance Inflation Factor (VIF) method. At the bottom, four goodness of fit measures are reported: the Bayesian information criterion (BIC), the Akaike Information Criterion (AIC), the Deviance and the Pearson $\chi^2$. Standard errors are clustered at the firm level and reported in parentheses. ***, **, and * correspond to statistical significance at the 1%, 5%, and 10% level, respectively.

| | Dependent Variable: *sale_occurrence* | | |
|---|---|---|---|
| Variable | July 01, 2018 - February 10, 2022 | July 01, 2018 - March 11, 2021 | March 12, 2021 - February 10, 2022 |
| *const* | -0.0000*** | -0.0000*** | -0.0000 |
| | (0.0000) | (0.0000) | (0.0000) |
| *BDM* | 0.0842*** | 0.0146 | 0.1472*** |
| | (0.0095) | (0.0159) | (0.0135) |
| *BGCI* | 0.7350*** | 1.1313*** | -0.1805*** |
| | (0.0149) | (0.0726) | (0.0118) |
| *black_%* | -0.0790*** | -0.0159 | -0.0150 |
| | (0.0093) | (0.0170) | (0.0129) |
| *blue_%* | 0.0592*** | -0.0023 | 0.0811*** |
| | (0.0092) | (0.0181) | (0.0124) |
| *d_ann_messages_month* | | 0.2379*** | |
| | | (0.0366) | |
| *d_ann_users_month* | -0.0608*** | | 0.1533*** |
| | (0.0221) | | (0.0247) |
| *d_gen_users_month* | 0.4007*** | -0.1615*** | 0.5034*** |
| | (0.0213) | (0.0372) | (0.0229) |
| *discord_account* | -0.2134*** | -0.2670*** | -0.3549*** |
| | (0.0149) | (0.0423) | (0.0193) |
| *floor_price_collection* | 0.3535*** | 0.3326*** | 0.4646*** |
| | (0.0102) | (0.0435) | (0.0130) |
| *following_count* | 0.0403** | 0.0687* | 0.0912*** |
| | (0.0163) | (0.0393) | (0.0221) |
| *gray_%* | 0.0494*** | 0.0052 | 0.0701*** |
| | (0.0093) | (0.0170) | (0.0130) |
| *green_%* | 0.0645*** | 0.0009 | 0.0320*** |
| | (0.0087) | (0.0176) | (0.0116) |
| *like_count_month* | | -0.2271*** | |
| | | (0.0458) | |
| *listed_count* | -0.3940*** | 0.4779*** | -0.6474*** |
| | (0.0132) | (0.0361) | (0.0164) |
| *marketplace_collection* | 0.1319*** | 0.5969*** | 0.6419*** |
| | (0.0173) | (0.0446) | (0.0203) |
| *nft_type_erc721* | 0.1127*** | -0.1468*** | -0.0332*** |
| | (0.0146) | (0.0356) | (0.0109) |
| *norm_shannon_entropy* | -0.1501*** | -0.0116 | -0.2871*** |
| | (0.0109) | (0.0177) | (0.0162) |
| *num_of_colors* | 0.0438*** | -0.0148 | 0.0596*** |
| | (0.0114) | (0.0190) | (0.0158) |
| *num_tweets* | -0.0987*** | | -0.5478*** |
| | (0.0209) | | (0.0283) |
| *orange_%* | -0.0270*** | -0.0421** | 0.0354*** |
| | (0.0100) | (0.0196) | (0.0134) |
| *purple_%* | -0.0032 | 0.0177 | -0.0169 |
| | (0.0087) | (0.0171) | (0.0116) |
| *quote_count_month* | -0.0942*** | | -0.0451 |
| | (0.0162) | | (0.0306) |
| *red_%* | 0.0831*** | 0.0162 | 0.1350*** |
| | (0.0090) | (0.0170) | (0.0122) |
| *reply_count_month* | | | 0.1071*** |
| | | | (0.0326) |
| *twitter_account* | 0.1745*** | -0.1951*** | 0.2899*** |
| | (0.0091) | (0.0229) | (0.0114) |
| *verified* | 0.2149*** | | 0.3857*** |
| | (0.0164) | | (0.0200) |
| *white_%* | -0.0118 | 0.0038 | 0.0001 |
| | (0.0087) | (0.0159) | (0.0118) |
| *year_collection_creation_2019* | 0.3968*** | 0.4789*** | -0.1392 |
| | (0.0162) | (0.0476) | (108.4365) |
| *year_collection_creation_2020* | -0.0207 | 0.0689 | -0.0011 |
| | (0.0128) | (0.0456) | (78.9488) |
| *year_collection_creation_2021* | -0.4542*** | -0.8158*** | -5.6663 |
| | (0.0245) | (0.0911) | (143.6943) |
| *year_collection_creation_2022* | -0.1822*** | | -4.3944 |
| | (0.0127) | | (108.3092) |
| *yellow_%* | 0.0124 | 0.0192 | 0.0108 |
| | (0.0091) | (0.0191) | (0.0120) |
| BIC | -802413.29 | -190143.77 | -579368.86 |
| AIC | 100499.86 | 27939.71 | 58336.88 |
| Deviance | 10044 | 27888 | 58277 |
| Pearson $\chi^2$ | 80528.45 | 25749.63 | 56180.61 |

**Note:** Robust Standard Errors in parenthesis.
$***\rho < 0.01$, $**\rho < 0.05$, $*\rho < 0.1$

**Table 3. Binomial GLM regression results, *BGCI* Excluded**. This table presents the coefficients of the GLM regression relative to the probability of an NFT being sold, by excluding the variable relating to the Bloomberg Galaxy Crypto Index (*BGCI*). The regressions are divided by the observed time series: from July 2018 to 10th February 2022, from July 2018 to 11th March 2021 and from 12th March 2021 to 10th February 2022. The number of observations is 860,067, and the dependent variable is *sale_occurrence*, which assumes 0 value for non-sold NFT and 1 vice versa. Variable selection was performed using the Variance Inflation Factor (VIF) method. At the bottom, four goodness of fit measures are reported: the Bayesian information criterion (BIC), the Akaike Information Criterion (AIC), the Deviance and the Pearson $\chi^2$. Standard errors are clustered at the firm level and reported in parentheses. ***, **, and * correspond to statistical significance at the 1%, 5%, and 10% level, respectively.

| | Dependent Variable: *sale_occurrence* | | |
|---|---|---|---|
| Variable | July 01, 2018 - February 10, 2022 | July 01, 2018 - March 11, 2021 | March 12, 2021 - February 10, 2022 |
| *const* | 0.0000*** | -0.0000*** | -0.0000 |
| | (0.0000) | (0.0000) | (0.0000) |
| *BDM* | 0.0996*** | -0.0051 | 0.1395*** |
| | (0.0093) | (0.0157) | (0.0134) |
| *black_%* | -0.0602*** | -0.0052 | -0.0233* |
| | (0.0091) | (0.0169) | (0.0128) |
| *blue_%* | 0.0645*** | 0.0088 | 0.0824*** |
| | (0.0091) | (0.0179) | (0.0124) |
| *d_ann_messages_month* | | 0.0273 | |
| | | (0.0341) | |
| *d_ann_users_month* | -0.0275 | | 0.1406*** |
| | (0.0213) | | (0.0247) |
| *d_gen_users_month* | 0.3198*** | -0.2209*** | 0.5127*** |
| | (0.0205) | (0.0369) | (0.0228) |
| *discord_account* | -0.1011*** | 0.0494 | -0.3819*** |
| | (0.0143) | (0.0347) | (0.0193) |
| *floor_price_collection* | 0.3544*** | 0.5372*** | 0.4522*** |
| | (0.0101) | (0.0430) | (0.0129) |
| *following_count* | 0.1200*** | -0.0028 | 0.0597*** |
| | (0.0159) | (0.0389) | (0.0221) |
| *gray_%* | 0.0412*** | 0.0071 | 0.0762*** |
| | (0.0092) | (0.0169) | (0.0130) |
| *green_%* | 0.0664*** | 0.0112 | 0.0412*** |
| | (0.0085) | (0.0174) | (0.0115) |
| *like_count_month* | | -0.3151*** | |
| | | (0.0449) | |
| *listed_count* | -0.2856*** | 0.3272*** | -0.6503*** |
| | (0.0128) | (0.0344) | (0.0163) |
| *marketplace_collection* | -0.0835*** | 0.3896*** | 0.6688*** |
| | (0.0165) | (0.0423) | (0.0201) |
| *nft_type_erc721* | 0.1996*** | -0.1525*** | -0.0465*** |
| | (0.0141) | (0.0351) | (0.0108) |
| *norm_shannon_entropy* | -0.1539*** | -0.0121 | -0.2852*** |
| | (0.0107) | (0.0176) | (0.0161) |
| *num_of_colors* | 0.0108 | -0.0207 | 0.0686*** |
| | (0.0112) | (0.0189) | (0.0158) |
| *num_tweets* | -0.1694*** | | -0.5071*** |
| | (0.0203) | | (0.0281) |
| *orange_%* | -0.0145 | -0.0389** | 0.0264** |
| | (0.0098) | (0.0195) | (0.0133) |
| *purple_%* | -0.0080 | 0.0234 | -0.0100 |
| | (0.0085) | (0.0170) | (0.0116) |
| *quote_count_month* | -0.1008*** | | -0.0324 |
| | (0.0158) | | (0.0306) |
| *red_%* | 0.0887*** | 0.0272 | 0.1350*** |
| | (0.0088) | (0.0169) | (0.0122) |
| *reply_count_month* | | | 0.0517 |
| | | | (0.0324) |
| *twitter_account* | 0.1613*** | -0.0542** | 0.3055*** |
| | (0.0089) | (0.0213) | (0.0114) |
| *verified* | 0.1859*** | | 0.3863*** |
| | (0.0163) | | (0.0199) |
| *white_%* | 0.0213** | 0.0097 | -0.0154 |
| | (0.0086) | (0.0159) | (0.0117) |
| *year_collection_creation_2019* | 0.3228*** | 0.4629*** | -0.1387 |
| | (0.0157) | (0.0473) | (114.4002) |
| *year_collection_creation_2020* | -0.0390*** | 0.3153*** | -0.0086 |
| | (0.0122) | (0.0432) | (83.2368) |
| *year_collection_creation_2021* | -0.0748*** | 0.1644** | -5.7275 |
| | (0.0222) | (0.0645) | (151.5591) |
| *year_collection_creation_2022* | -0.0691*** | | -4.4126 |
| | (0.0120) | | (114.2373) |
| *yellow_%* | 0.0146 | 0.0212 | 0.0094 |
| | (0.0089) | (0.0190) | (0.0120) |
| BIC | -799744.31 | -189862.27 | -579142.38 |
| AIC | 103178.12 | 28229.21 | 58572.33 |
| Deviance | 103122.12 | 28179.20 | 58514.33 |
| Pearson $\chi^2$ | 82060.16 | 25665.88 | 56287.96 |

**Note:** Robust Standard Errors in parenthesis.
$***\rho < 0.01, **\rho < 0.05, *\rho < 0.1$

**Table 4. Negative binomial GLM regression results**. The following table presents the coefficients of the GLM regression relative to the number of sales (expected frequency) per NFT divided by the observed time series: from July 2018 to 10th February 2022, from July 2018 to 11th March 2021 and from 12th March 2021 to 10th February 2022. The number of observations is 246125, and the dependent variable is *number_of_trades* which counts the number of trades that each NFT encountered during its life. Variable selection was performed using the Variance Inflation Factor (VIF) method. At the bottom, four goodness of fit measures are reported: the Bayesian information criterion (BIC), the Akaike Information Criterion (AIC), the Deviance and the Pearson $\chi^2$. Standard errors are clustered at the firm level and reported in parentheses. ***, **, and * correspond to statistical significance at the 1%, 5%, and 10% level, respectively.

| | Dependent Variable: *number_of_trades* | | |
|---|---|---|---|
| Variable | July 01, 2018 - February 10, 2022 | July 01, 2018 - March 11, 2021 | March 12, 2021 - February 10, 2022 |
| *const* | 0.0000*** | -0.0000*** | 0.0000 |
| | (0.0000) | (0.0000) | (0.0000) |
| *BDM* | -0.0030 | -0.0085 | 0.0013 |
| | (0.0082) | (0.0288) | (0.0085) |
| *BGCI* | 0.0229*** | -0.0020 | 0.0297*** |
| | (0.0077) | (0.0372) | (0.0072) |
| *black_%* | -0.0245*** | 0.0114 | -0.0279*** |
| | (0.0075) | (0.0292) | (0.0078) |
| *blue_%* | 0.0403*** | -0.0023 | 0.0249*** |
| | (0.0077) | (0.0315) | (0.0080) |
| *d_ann_messages_month* | 0.0728*** | | 0.0887*** |
| | (0.0175) | | (0.0174) |
| *d_gen_messages_month* | -0.0147 | 0.0058 | -0.0556*** |
| | (0.0174) | (0.0346) | (0.0172) |
| *discord_account* | 0.0564*** | 0.0574 | 0.0485*** |
| | (0.0125) | (0.0593) | (0.0129) |
| *floor_price* | -0.5378*** | -0.6336*** | -0.5413*** |
| | (0.0106) | (0.0496) | (0.0105) |
| *floor_price_collection* | -0.0289*** | 0.0124 | -0.0207** |
| | (0.0092) | (0.0374) | (0.0097) |
| *following_count* | -0.2897*** | -0.0221 | -0.2374*** |
| | (0.0124) | (0.0568) | (0.0116) |
| *gray_%* | 0.0458*** | 0.0298 | 0.0272*** |
| | (0.0077) | (0.0296) | (0.0080) |
| *green_%* | 0.0807*** | -0.0377 | 0.0863*** |
| | (0.0071) | (0.0293) | (0.0073) |
| *last_price* | 0.4719*** | 0.5923*** | 0.4807*** |
| | (0.0120) | (0.0571) | (0.0121) |
| *listed_count* | -0.0908*** | | -0.1007*** |
| | (0.0114) | | (0.0111) |
| *marketplace_collection* | -0.0037 | -0.0464 | 0.0132 |
| | (0.0202) | (0.0684) | (0.0213) |
| *nft_type_erc721* | -0.3452*** | -0.1696*** | -0.3102*** |
| | (0.0039) | (0.0330) | (0.0035) |
| *norm_shannon_entropy* | -0.0950*** | -0.0044 | -0.1058*** |
| | (0.0091) | (0.0330) | (0.0097) |
| *num_of_colors* | -0.0135 | 0.0134 | -0.0006 |
| | (0.0094) | (0.0330) | (0.0099) |
| *num_tweets* | 0.3163*** | 0.0437 | 0.2336*** |
| | (0.0141) | (0.0420) | (0.0120) |
| *orange_%* | 0.0816*** | -0.0016 | 0.0765*** |
| | (0.0078) | (0.0311) | (0.0082) |

## Table 4 - Cont.

| Variable | Dependent Variable: *number_of_trades*<br>July 01, 2018 - February 10, 2022 | July 01, 2018 - March 11, 2021 | March 12, 2021 - February 10, 2022 |
|---|---|---|---|
| *platform_last_sale_LooksRare* | -0.0184***<br>(0.0059) | | -0.0101*<br>(0.0059) |
| *platform_last_sale_OpenSea* | | -0.0840**<br>(0.0411) | |
| *platform_last_sale_Rarible* | -0.0903***<br>(0.0057) | | -0.0605***<br>(0.0047) |
| *platform_last_sale_SuperRare* | 0.0991***<br>(0.0172) | | 0.0600***<br>(0.0138) |
| *purple_%* | 0.0219***<br>(0.0070) | -0.0132<br>(0.0295) | 0.0018<br>(0.0073) |
| *quote_count_month* | | -0.0566<br>(0.0508) | -0.0471***<br>(0.0109) |
| *red_%* | -0.0090<br>(0.0074) | 0.0167<br>(0.0295) | -0.0077<br>(0.0077) |
| *reply_count_month* | -0.1004***<br>(0.0129) | | |
| *timediff* | 0.1735***<br>(0.0116) | 0.0380<br>(0.0608) | 0.1166***<br>(0.0121) |
| *twitter_account* | 0.0300***<br>(0.0068) | 0.0062<br>(0.0270) | 0.0639***<br>(0.0075) |
| *verified* | -0.1479***<br>(0.0179) | | -0.1635***<br>(0.0169) |
| *white_%* | 0.0731***<br>(0.0071) | 0.0427<br>(0.0263) | 0.0644***<br>(0.0074) |
| *year_collection_creation_2019* | 0.0080<br>(0.0169) | | -0.0311**<br>(0.0141) |
| *year_collection_creation_2020* | -0.0412***<br>(0.0081) | 0.0568<br>(0.0548) | -0.0817***<br>(0.0070) |
| *year_collection_creation_2021* | 0.0600***<br>(0.0191) | 0.0686<br>(0.0875) | -0.0434**<br>(0.0173) |
| *year_collection_creation_2022* | -0.0248*<br>(0.0127) | | -0.0855***<br>(0.0133) |
| *year_last_sale_2019* | -0.1106***<br>(0.0089) | | |
| *year_last_sale_2020* | | 0.0484<br>(0.0624) | |
| *year_last_sale_2021* | 0.0397**<br>(0.0178) | 0.0254<br>(0.0816) | -0.1773***<br>(0.0070) |
| *year_last_sale_2022* | 0.2015***<br>(0.0170) | | |
| *yellow_%* | -0.0418***<br>(0.0074) | 0.0173<br>(0.0302) | -0.0543***<br>(0.0077) |
| BIC | -213164 | -11686.11 | -196585.19 |
| AIC | 63480 | 3797.82 | 58548.26 |
| Deviance | 13895.96 | 256.98 | 12437.65 |
| Pearson $\chi^2$ | 47300 | 353 | 37900 |

**Note:** Robust Standard Errors in parenthesis.
$***\rho < 0.01$, $**\rho < 0.05$, $*\rho < 0.1$

**Table 5. Log-normal GLM regression results**. This table presents the coefficients of the GLM regression relative to the logarithm of the average price (expected severity) per NFT divided by the observed time series: from July 2018 to 10th February 2022, from July 2018 to 11th March 2021 and from 12th March 2021 to 10th February 2022. The number of observations is 246125 and the dependent variable is *avgprice* which measure the average price per transfer per NFT. Variable selection performed using the Variance Inflation Factor (VIF) method. At the bottom, four goodness of fit measures are reported: the Bayesian information criterion (BIC), the Akaike Information criterion (AIC), the Deviance and the Pearson $\chi^2$. Standard errors are clustered at the firm level and reported in parentheses. ***, **, and * correspond to statistical significance at the 1%, 5%, and 10% level, respectively.

| | Dependent Variable: *avgprice* | | |
|---|---|---|---|
| Variable | July 01, 2018 - February 10, 2022 | July 01, 2018 - March 11, 2021 | March 12, 2021 - February 10, 2022 |
| *const* | -0.0000<br>(0.0000) | -0.0000<br>(0.0000) | -0.0000<br>(0.0000) |
| *BDM* | 0.0030<br>(0.0243) | 0.0013<br>(0.0714) | 0.0024<br>(0.0256) |
| *BGCI* | 0.0033<br>(0.0241) | 0.0022<br>(0.0963) | 0.0049<br>(0.0224) |
| *black_%* | -0.0015<br>(0.0229) | -0.0025<br>(0.0732) | -0.0017<br>(0.0241) |
| *blue_%* | 0.0016<br>(0.0235) | 0.0010<br>(0.0780) | 0.0016<br>(0.0248) |
| *d_ann_messages_month* | 0.0058<br>(0.0616) | | 0.0167<br>(0.0614) |
| *d_gen_messages_month* | -0.0057<br>(0.0587) | 0.0049<br>(0.0857) | -0.0230<br>(0.0590) |
| *discord_account* | -0.0040<br>(0.0410) | 0.0087<br>(0.1528) | -0.0003<br>(0.0428) |
| *floor_price* | 0.2196***<br>(0.0673) | 0.1922<br>(0.2736) | 0.2085***<br>(0.0687) |
| *floor_price_collection* | 0.0099<br>(0.0289) | 0.0105<br>(0.0967) | 0.0180<br>(0.0307) |
| *following_count* | -0.0058<br>(0.0443) | 0.0093<br>(0.1630) | -0.0088<br>(0.0429) |
| *gray_%* | -0.0003<br>(0.0236) | 0.0009<br>(0.0733) | -0.0004<br>(0.0250) |
| *green_%* | -0.0033<br>(0.0228) | -0.0052<br>(0.0734) | -0.0027<br>(0.0240) |
| *last_price* | 0.4536***<br>(0.0695) | 0.7241**<br>(0.2827) | 0.4201***<br>(0.0708) |
| *listed_count* | -0.0021<br>(0.0378) | | -0.0007<br>(0.0373) |
| *marketplace_collection* | -0.0100<br>(0.0680) | -0.0258<br>(0.1725) | -0.0132<br>(0.0718) |
| *nft_type_erc721* | 0.0022 | 0.0053 | 0.0008 |
| *norm_shannon_entropy* | 0.0017<br>(0.0279) | -0.0051<br>(0.0842) | 0.0002<br>(0.0297) |
| *num_of_colors* | 0.0005<br>(0.0289) | 0.0028<br>(0.0830) | -0.0003<br>(0.0306) |
| *num_tweets* | -0.0002<br>(0.0525) | 0.0030<br>(0.1305) | -0.0000<br>(0.0492) |
| *number_of_trades* | 0.0392<br>(0.0242) | -0.0134<br>(0.0978) | 0.0401<br>(0.0253) |
| *orange_%* | -0.0017<br>(0.0248) | 0.0003<br>(0.0789) | -0.0009<br>(0.0262) |

## Table 5 - Cont.

| Variable | Dependent Variable: *avgprice* | | |
|---|---|---|---|
| | July 01, 2018 - February 10, 2022 | July 01, 2018 - March 11, 2021 | March 12, 2021 - February 10, 2022 |
| *platform_last_sale_LooksRare* | 0.0007<br>(0.0194) | | 0.0009<br>(0.0204) |
| *platform_last_sale_OpenSea* | | -0.0293<br>(0.1101) | |
| *platform_last_sale_Rarible* | -0.0007<br>(0.0232) | | 0.0005<br>(0.0214) |
| *platform_last_sale_SuperRare* | 0.0147<br>(0.0543) | | 0.0133<br>(0.0464) |
| *purple_%* | -0.0020<br>(0.0220) | -0.0020<br>(0.0726) | -0.0028<br>(0.0230) |
| *quote_count_month* | | -0.0004<br>(0.1326) | -0.0032<br>(0.0409) |
| *red_%* | -0.0004<br>(0.0228) | 0.0006<br>(0.0743) | 0.0009<br>(0.0240) |
| *reply_count_month* | -0.0019<br>(0.0411) | | |
| *timediff* | 0.0230<br>(0.0370) | -0.0087<br>(0.1551) | 0.0285<br>(0.0383) |
| *twitter_account* | 0.0022<br>(0.0227) | 0.0028<br>(0.0686) | 0.0036<br>(0.0237) |
| *verified* | -0.0014<br>(0.0596) | | -0.0016<br>(0.0621) |
| *white_%* | -0.0012<br>(0.0225) | -0.0004<br>(0.0676) | -0.0009<br>(0.0238) |
| *year_collection_creation_2019* | 0.0076<br>(0.0560) | | 0.0103<br>(0.0493) |
| *year_collection_creation_2020* | 0.0038<br>(0.0312) | -0.0167<br>(0.1410) | 0.0052<br>(0.0281) |
| *year_collection_creation_2021* | 0.0225<br>(0.0635) | 0.0134<br>(0.2248) | 0.0225<br>(0.0590) |
| *year_collection_creation_2022* | 0.0124<br>(0.0419) | | 0.0144<br>(0.0449) |
| *year_last_sale_2019* | -0.0014<br>(0.0220) | | |
| *year_last_sale_2020* | | 0.0205<br>(0.1502) | |
| *year_last_sale_2021* | 0.0183<br>(0.0544) | 0.0160<br>(0.2014) | 0.0033<br>(0.0256) |
| *year_last_sale_2022* | 0.0172<br>(0.0538) | | |
| *yellow_%* | -0.0012<br>(0.0230) | 0.0023<br>(0.0773) | -0.0029<br>(0.0241) |
| BIC | -802409.78 | -805932 | -806321 |
| AIC | 100512.65 | 7661.12 | 104979.26 |
| Deviance | 189909.67 | 9813.85 | 180456.04 |
| Pearson $\chi^2$ | 189886.95 | 9813.86 | 180456.04 |

**Note:** Robust Standard Errors in parenthesis.
$***\rho < 0.01$, $**\rho < 0.05$, $*\rho < 0.1$

**Table 6. Performance of different NFT investment strategies.** This table shows the returns and the volatility of various trading scenarios for single-sold and multi-sold NFTs for the years 2021 and 2022.

**Performance in 2021**

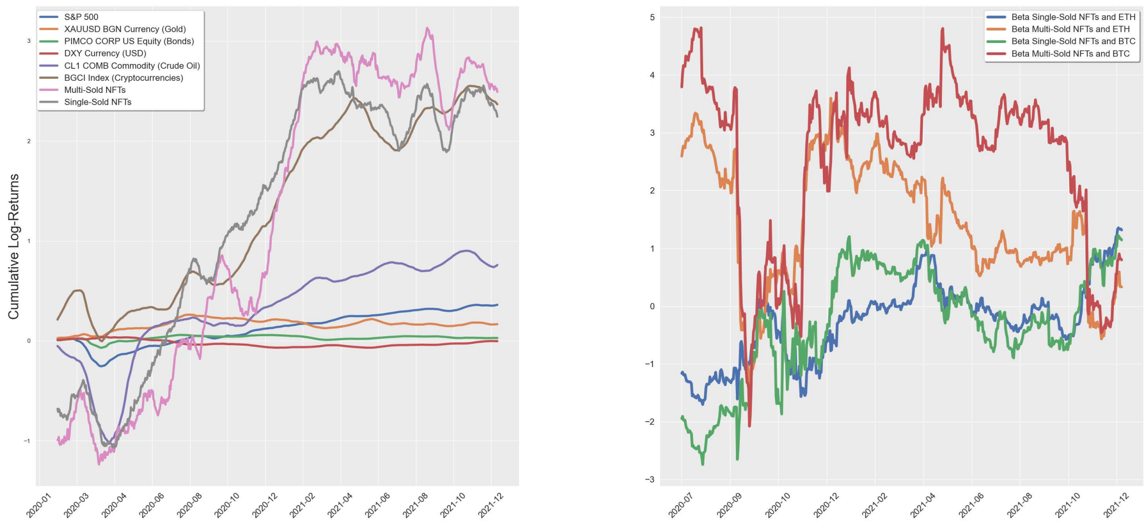
| | Single-Sold High Volatility Scenario | | Single-Sold Low Volatility Scenario | |
|---|---|---|---|---|
| | Median of Cumulative Log Returns | STD of Cumulative Log Returns | Median of Cumulative Log Returns | STD of Cumulative Log Returns |
| BGCI | 0.84874 | 0.14223 | 0.84874 | 0.14223 |
| Top-5 | 0.80461 | 0.41901 | 0.65597 | 1.28491 |
| Top-10 | 1.33326 | 0.52441 | 1.01000 | 0.81072 |
| Top-20 | 1.15193 | 0.30680 | 1.04637 | 0.23695 |

| | Multi-Sold High Volatility Scenario | | Multi-Sold Low Volatility Scenario | |
|---|---|---|---|---|
| | Median of Cumulative Log Returns | STD of Cumulative Log Returns | Median of Cumulative Log Returns | STD of Cumulative Log Returns |
| BGCI | 0.84874 | 0.14223 | 0.84874 | 0.14223 |
| Top-5 | -0.56452 | 0.38207 | 5.49503 | 0.68806 |
| Top-10 | 0.36603 | 0.23467 | 4.63010 | 0.92329 |
| Top-20 | 0.35326 | 0.17991 | 0.62306 | 0.68816 |

**Performance in 2022**

| | Single-Sold High Volatility Scenario | | Single-Sold Low Volatility Scenario | |
|---|---|---|---|---|
| | Median of Cumulative Log Returns | STD of Cumulative Log Returns | Median of Cumulative Log Returns | STD of Cumulative Log Returns |
| BGCI | -0.86930 | 0.30537 | -0.86930 | 0.30537 |
| Top-5 | -0.60019 | 0.89431 | -2.31249 | 1.26698 |
| Top-10 | 0.15386 | 0.54128 | -2.40762 | 0.85753 |
| Top-20 | -2.02811 | 0.54457 | -1.15983 | 0.83629 |

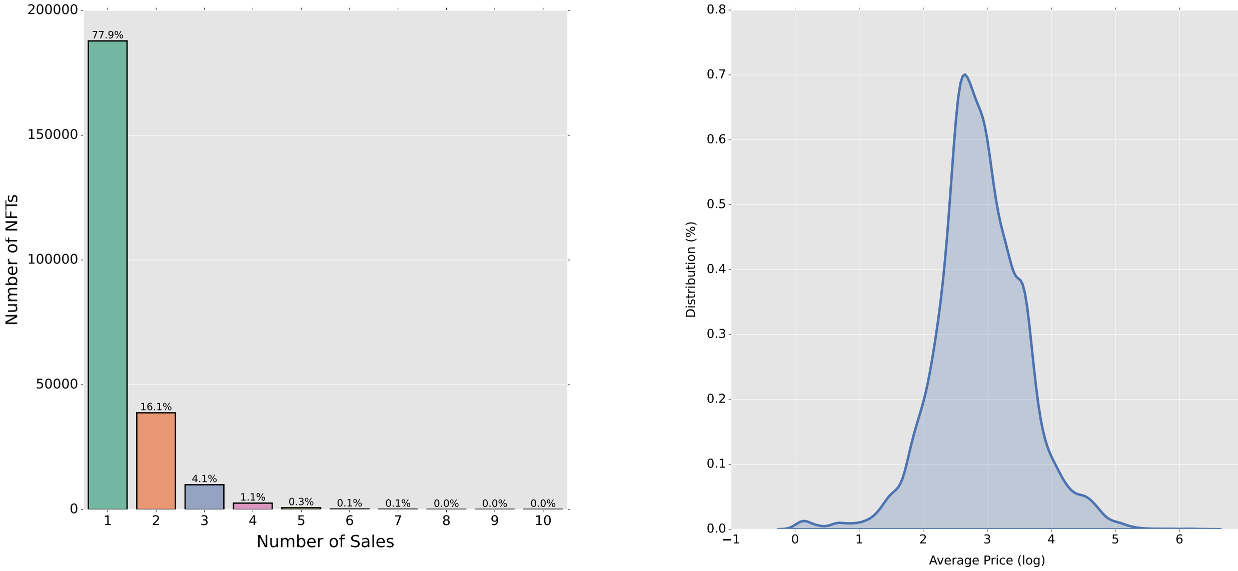
| | Multi-Sold High Volatility Scenario | | Multi-Sold Low Volatility Scenario | |
|---|---|---|---|---|
| | Median of Cumulative Log Returns | STD of Cumulative Log Returns | Median of Cumulative Log Returns | STD of Cumulative Log Returns |
| BGCI | -0.86930 | 0.30537 | -0.86930 | 0.30537 |
| Top-5 | -1.58835 | 0.56728 | -2.74530 | 0.67600 |
| Top-10 | -1.16717 | 0.72243 | -3.09742 | 0.78026 |
| Top-20 | -1.24324 | 0.76291 | -1.27920 | 1.19775 |




**(a)** 30-day rolling mean cumulative log-returns of NFTs and indexes.

**(b)** Betas of multi-sold and single-sold NFTs vs. ETH & BTC on a 180-day rolling window.

**Figure 1. NFT volatility and correlation with the cryptocurrency market**. Figure 1a showcases that the art NFT market had higher returns than the cryptocurrency market for the most part of 2021 and that multi-sold NFTs perform better than single-sold. Figure 1b demonstrates that the returns of multi-sold NFTs are more volatile and have a stronger correlation with ETH and BTC.




**(a)** Sold NFT sales distribution

**(b)** Sold NFT (log) average price distribution

**Figure 2. Dependent variables distribution**. This figure illustrates the distribution of the dependent variables modelled for the expected frequency and severity of trading for the art NFT market between July 2018 and February 2022. Figure 2a shows the spread of the expected numbers of sales in the form of a discrete variable. Figure 2b expresses the continuous distribution of the logarithmic transformation of the average prices for each NFT considered.

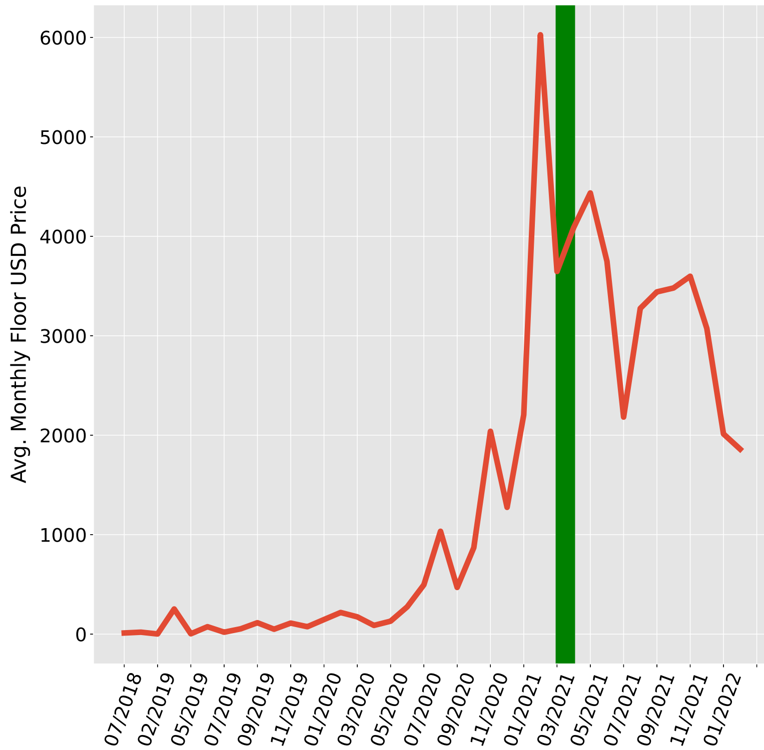
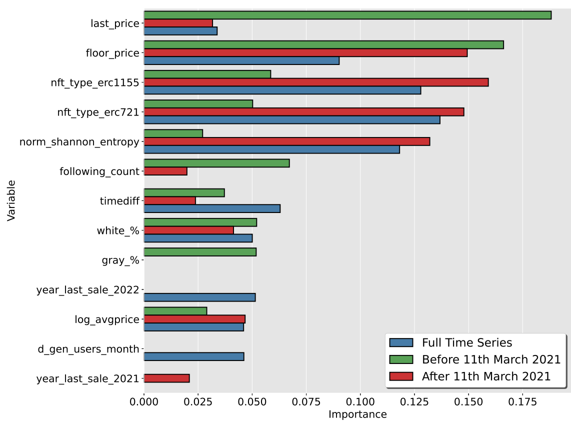
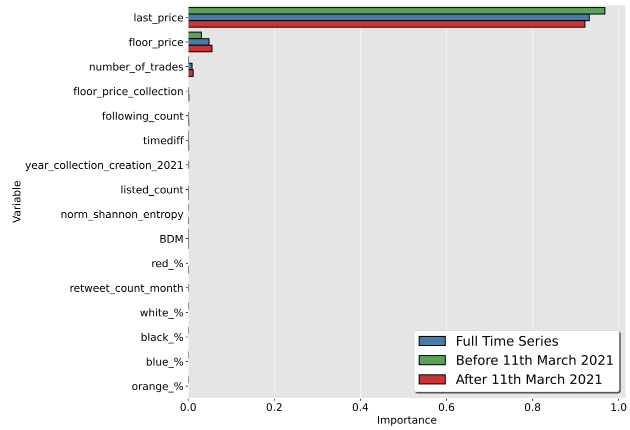



**Figure 3. NFTs Floor Price Dynamics.** This figure represents the monthly average floor price trend of each NFT under analysis in the period July 2018 - February 2022. The solid green area represents a structural break.

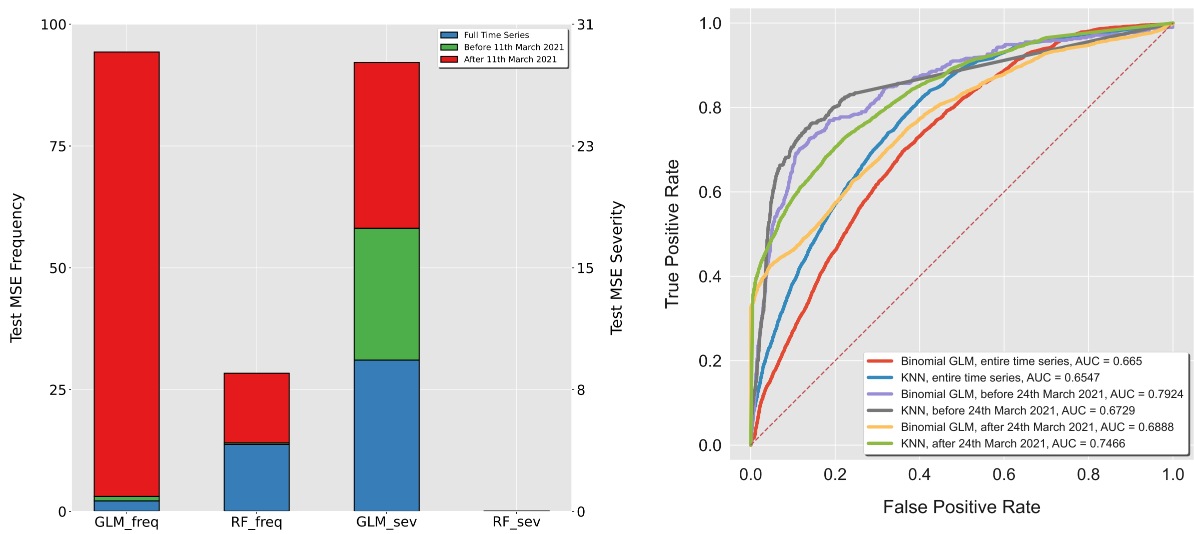



**(a)** RF Expected Frequency Var. Importance




**(b)** RF Expected Severity Var. Importance

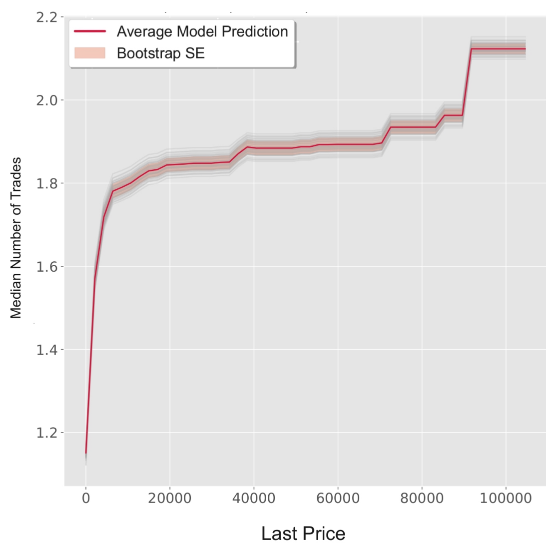
**Figure 4. Random Forest (RF) variables**. Figure 4a shows the important variables for the RF relative to expected frequencies, while Figure 4b shows the relevant variables for the expected severity.



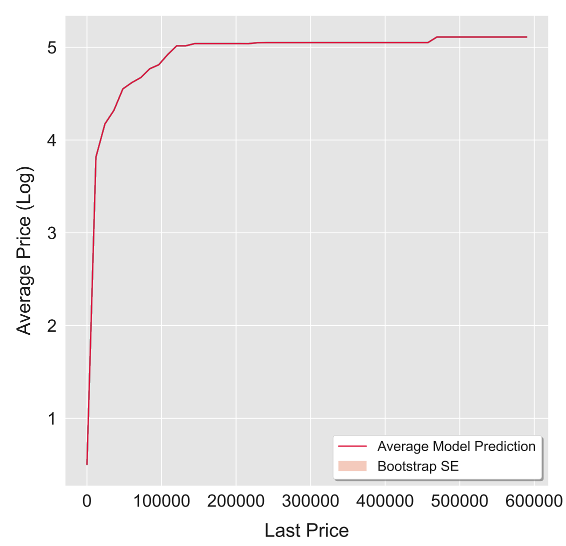

**(a)** Test MSE Frequency-Severity Modeling

**(b)** ROC Curve, Binomial GLM vs KNN

**Figure 5. Test MSE and ROC curve.** Figure 5a shows the comparison among the GLM and RF algorithms' predictive performances by calculating the test mean squared error (MSE) on each different model. The test MSE is defined as the average squared difference between the estimated values and the actual one calculated on the test set, as a measure of out-of-sample performance. Figure 5b shows the ROC curves belonging to binomial GLM and KNN for each of the tested datasets. Generally, the curve for a binary classification problem plots the true positive rate as a function of the false positive rate. The points of the curve are obtained by sweeping the classification threshold from the most positive classification value to the most negative.




**(a)** PDP and ICE *last_price*,
RF Frequency




**(b)** PDP and ICE *last_price*,
RF Severity

**Figure 6. RF Models *last_price* PDP and ICE.** This figure illustrates the partial dependence plots (PDP) and individual conditional expectations (ICE) functions for the variable *last_price* relative to the RF models used in the study. PDPs and ICE plots are tools used to visually analyze the impact of different features on the output. PDP plots display the marginal contribution of one or two features, while ICE plots show the prediction changes for each individual data point, given by bootstrapping the original data set (Bootstrap Standard Errors (SE) are plotted, too). Both types of plots utilize similar methods to model the prediction changes, but PDP plots focus on the overall trends of one or two features, while ICE plots provide a more detailed view of the variations. Figure 6a and Figure 6b illustrate the effect of the last price for each NFT on the expected frequency and expected severity, respectively.

# Appendix A

**Table A1. Variable Descriptions.** The following table list all the variables used in the analysis alongside their short description

| Variable | Description |
|---|---|
| *avgprice* | Average Non-Fungible Token (NFT) price |
| *BDM* | Block Decomposition Method per NFT image |
| *BGCI* | daily Bloomberg Galaxy Crypto Index last price in USD |
| *black_%* | Percentage of black color in the NFT image |
| *blue_%* | Percentage of blue color in the NFT image |
| *d_ann_messages_month* | Average number of messages in Discord ”Announcement” channel per NFT collection per month |
| *d_ann_users_month* | Average number of unique active users in Discord ”Announcement” channel per NFT collection per month |
| *d_gen_messages_month* | Average number of messages in Discord ”General” channel per NFT collection per month |
| *d_gen_users_month* | Average number of unique active users in Discord ”General” channel per NFT collection per month |
| *day_last_sale* | Day of last sale per single NFT |
| *discord_account* | Presence of a Discord account representing the artist or the NFT collection |
| *floor_price* | NFT floor price |
| *floor_price_collection* | NFT collection floor price |
| *followers_count* | Number of followers per Twitter account |
| *following_count* | Number of following per Twitter account |
| *gray_%* | Percentage of grey color in the NFT image |
| *green_%* | Percentage of green color in the NFT image |
| *last_price* | NFT last selling price |
| *like_count_month* | Average number of ”like” reactions per Twitter account per month |
| *listed_count* | Average number of public lists memberships per Twitter account |
| *marketplace_collection* | NFT belonging to one of the following collections: Foundation (FND), Editional, KnownOrigin, SuperRare |
| *month_collection_creation* | Month of creation per NFT collection |
| *month_last_sale* | Month of last sale per single NFT |
| *nft_type* | Belonging to the ERC721 or ERC1155 protocol per single NFT |
| *norm_shannon_entropy* | Normalized Shannon Entropy per NFT image |
| *num_colors* | Number of present colors in the NFT image |
| *num_tweets* | Average number of ”tweets” per Twitter account |
| *num_trades* | Number of trades per NFT |
| *orange_%* | Percentage of orange color in the NFT image |
| *platform_last_sale* | Blockchain platform of last sale per NFT |
| *purple_%* | Percentage of purple color in the NFT image |
| *quote_count_month* | Average number of ”Retweets” with comments per Twitter account per month |
| *red_%* | Percentage of red color in the NFT image |
| *reply_count_month* | Average number of reply to ”tweets” per Twitter account per month |
| *retweet_count_month* | Average number of ”retweets” per Twitter account per month |
| *sale_occurrence* | Sale occurrence per NFT |
| *timediff* | Time difference (in days) between the collection creation and the selling dates per NFT |
| *twitter_account* | Presence of a Twitter account representing the artist or the NFT collection |
| *verified* | Presence of ”verified” badge per Twitter account |
| *white_%* | Percentage of white color in the NFT image |
| *year_collection_creation* | Year of creation per NFT collection |
| *year_last_sale* | Year of last sale per single NFT |
| *yellow_%* | Percentage of yellow color in the NFT image |

**Table A2. Binomial GLM Robustness Table**. The following table presents the coefficients change for the model with the complete time series as each of the variables is added into the regression

| | Dependent Variable: *sale_occurrence* | | | |
|---|---|---|---|---|
| Variable | (1) | (2) | (3) | (4) |
| *const* | 0.000*** (0.000) | 0.000*** (0.000) | -0.000*** (0.000) | -0.000*** (0.000) |
| *black_%* | -0.051*** (0.008) | -0.110*** (0.008) | -0.075*** (0.009) | -0.068*** (0.009) |
| *white_%* | 0.020*** (0.008) | -0.004 (0.008) | -0.015* (0.008) | -0.002 (0.008) |
| *gray_%* | 0.046*** (0.008) | 0.024*** (0.009) | 0.055*** (0.009) | 0.055*** (0.009) |
| *green_%* | 0.067*** (0.008) | 0.072*** (0.008) | 0.058*** (0.008) | 0.072*** (0.008) |
| *blue_%* | 0.063*** (0.008) | 0.027*** (0.008) | 0.043*** (0.009) | 0.070*** (0.009) |
| *yellow_%* | -0.000 (0.008) | -0.002 (0.009) | 0.023** (0.009) | 0.018** (0.009) |
| *purple_%* | -0.009 (0.008) | -0.006 (0.008) | 0.014 (0.008) | 0.003 (0.009) |
| *orange_%* | -0.011 (0.009) | -0.042*** (0.009) | -0.038*** (0.010) | -0.018* (0.010) |
| *red_%* | 0.071*** (0.008) | 0.037*** (0.008) | 0.068*** (0.009) | 0.092*** (0.009) |
| *norm_shannon_entropy* | -0.119*** (0.010) | -0.062*** (0.010) | -0.118*** (0.011) | -0.142*** (0.011) |
| *BDM* | 0.066*** (0.008) | 0.061*** (0.009) | 0.096*** (0.009) | 0.092*** (0.009) |
| *year_collection_creation_2019* | | 0.294*** (0.013) | 0.337*** (0.015) | 0.394*** (0.016) |
| *year_collection_creation_2020* | | -0.001 (0.011) | -0.078*** (0.012) | -0.021 (0.013) |
| *year_collection_creation_2021* | | -0.209*** (0.020) | -0.520*** (0.023) | -0.456*** (0.025) |
| *year_collection_creation_2022* | | -0.045*** (0.011) | -0.191*** (0.012) | -0.183*** (0.013) |
| *floor_price_collection* | | 0.233*** (0.008) | 0.376*** (0.010) | 0.356*** (0.010) |
| *nft_type_erc721* | | 0.165*** (0.014) | 0.107*** (0.014) | 0.114*** (0.015) |
| *marketplace_collection* | | 0.186*** (0.010) | -0.013 (0.016) | 0.133*** (0.017) |
| *BGCI* | | 0.475*** (0.013) | 0.700*** (0.015) | 0.732*** (0.015) |
| *twitter_account* | | | 0.168*** (0.009) | 0.172*** (0.009) |
| *following_count* | | | 0.138*** (0.015) | 0.036** (0.016) |
| *num_tweets* | | | -0.188*** (0.020) | -0.091*** (0.021) |
| *listed_count* | | | -0.496*** (0.013) | -0.394*** (0.013) |
| *verified* | | | 0.167*** (0.014) | 0.214*** (0.016) |
| *quote_count_month* | | | -0.003 (0.015) | -0.097*** (0.016) |
| *discord_account* | | | | -0.212*** (0.015) |
| *d_ann_users_month* | | | | -0.063*** (0.022) |
| *d_gen_users_month* | | | | 0.400*** (0.021) |
| BIC | -792623.65 | -798225.58 | -801597.76 | -802413.29 |
| AIC | 110456.72 | 104780.46 | 101352.55 | 100499.86 |
| Deviance | 11043.72 | 10474.46 | 10130.55 | 10044 |
| Residual Std. Error | 1.000(df = 79989) | 1.000(df = 79981) | 1.000(df = 79975) | 1.000(df = 79972) |
| F Statistic | (df = 10; 79989) | (df = 18; 79981) | (df = 24; 79975) | (df = 27; 79972) |

**Note:** Robust Standard Errors in parenthesis.
$***\rho < 0.01$, $**\rho < 0.05$, $*\rho < 0.1$

**Table A3. Negative Binomial GLM Robustness Table**. The following table presents the coefficients change for the model with the complete time series as each of the variables is added into the regression

| | Dependent Variable: *number_of_trades* | | | |
|---|---|---|---|---|
| Variable | (1) | (2) | (3) | (4) |
| *const* | 0.000*** (0.000) | -0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) |
| *black_%* | -0.014* (0.008) | -0.042*** (0.007) | -0.028*** (0.007) | -0.025*** (0.008) |
| *white_%* | 0.144*** (0.007) | 0.098*** (0.007) | 0.078*** (0.007) | 0.073*** (0.007) |
| *gray_%* | -0.013 (0.008) | 0.045*** (0.008) | 0.040*** (0.008) | 0.046*** (0.008) |
| *green_%* | 0.114*** (0.007) | 0.079*** (0.007) | 0.085*** (0.007) | 0.081*** (0.007) |
| *blue_%* | 0.008 (0.008) | 0.037*** (0.008) | 0.043*** (0.008) | 0.040*** (0.008) |
| *yellow_%* | -0.066*** (0.008) | -0.042*** (0.007) | -0.041*** (0.007) | -0.042*** (0.007) |
| *purple_%* | 0.069*** (0.007) | 0.024*** (0.007) | 0.024*** (0.007) | 0.022*** (0.007) |
| *orange_%* | 0.089*** (0.008) | 0.090*** (0.008) | 0.080*** (0.008) | 0.082*** (0.008) |
| *red_%* | -0.041*** (0.008) | -0.016** (0.007) | -0.013* (0.007) | -0.009 (0.007) |
| *norm_shannon_entropy* | 0.069*** (0.009) | -0.100*** (0.009) | -0.108*** (0.009) | -0.095*** (0.009) |
| *BDM* | -0.074*** (0.008) | 0.003 (0.008) | 0.007 (0.008) | -0.003 (0.008) |
| *num_of_colors* | -0.001 (0.010) | 0.002 (0.009) | -0.017* (0.009) | -0.013 (0.009) |
| *year_collection_creation_2019* | | 0.028* (0.016) | 0.005 (0.017) | 0.008 (0.017) |
| *year_collection_creation_2020* | | -0.007 (0.008) | -0.048*** (0.008) | -0.041*** (0.008) |
| *year_collection_creation_2021* | | -0.028* (0.017) | 0.058*** (0.019) | 0.060*** (0.019) |
| *year_collection_creation_2022* | | -0.050*** (0.011) | -0.021* (0.013) | -0.025** (0.013) |
| *year_last_sale_2019* | | -0.107*** (0.009) | -0.116*** (0.009) | -0.111*** (0.009) |
| *year_last_sale_2021* | | 0.054*** (0.018) | 0.038** (0.018) | 0.040** (0.018) |
| *year_last_sale_2022* | | 0.245*** (0.017) | 0.200*** (0.017) | 0.202*** (0.017) |
| *floor_price* | | -0.582*** (0.010) | -0.536*** (0.010) | -0.538*** (0.011) |
| *last_price* | | 0.500*** (0.012) | 0.475*** (0.012) | 0.472*** (0.012) |
| *floor_price_collection* | | -0.058*** (0.008) | -0.026*** (0.009) | -0.029*** (0.009) |
| *timediff* | | 0.070*** (0.009) | 0.169*** (0.011) | 0.173*** (0.011) |
| *nft_type_erc721* | | -0.318*** (0.003) | -0.336*** (0.004) | -0.345*** (0.004) |
| *platform_last_sale_LooksRare* | | -0.014** (0.006) | -0.019*** (0.006) | -0.018*** (0.006) |
| *platform_last_sale_Rarible* | | -0.090*** (0.006) | -0.099*** (0.006) | -0.090*** (0.006) |
| *platform_last_sale_SuperRare* | | -0.040*** (0.013) | 0.105*** (0.015) | 0.098*** (0.016) |
| *BGCI* | | 0.067*** (0.007) | 0.033*** (0.008) | 0.023*** (0.008) |

## Table A3 - Cont.

| Variable | Dependent Variable: *number_of_trades* (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| *twitter_account* | | | 0.024*** | 0.030*** |
| | | | (0.007) | (0.007) |
| *following_count* | | | -0.270*** | -0.290*** |
| | | | (0.012) | (0.012) |
| *num_tweets* | | | 0.319*** | 0.316*** |
| | | | (0.014) | (0.014) |
| *listed_count* | | | -0.074*** | -0.090*** |
| | | | (0.010) | (0.010) |
| *verified* | | | -0.212*** | -0.150*** |
| | | | (0.013) | (0.015) |
| *reply_count_month* | | | -0.106*** | -0.101*** |
| | | | (0.013) | (0.013) |
| *discord_account* | | | | 0.057*** |
| | | | | (0.012) |
| *d_ann_messages_month* | | | | 0.073*** |
| | | | | (0.017) |
| *d_gen_messages_month* | | | | -0.015 |
| | | | | (0.017) |
| BIC | -792623.65 | -798225.58 | -801597.76 | -802413.29 |
| AIC | 110456.72 | 104780.46 | 101352.55 | 100499.86 |
| Deviance | 11043.72 | 10474.46 | 10130.55 | 10044 |
| Residual Std. Error | 1.000(df = 79989) | 1.000(df = 79981) | 1.000(df = 79975) | 1.000(df = 79972) |
| F Statistic | (df = 10; 79989) | (df = 18; 79981) | (df = 24; 79975) | (df = 27; 79972) |

**Note:** Robust Standard Errors in parenthesis.
$***\rho < 0.01$, $**\rho < 0.05$, $*\rho < 0.1$

**Table A4. Log-Normal GLM Robustness Table**. The following table presents the coefficients change for the model with the complete time series as each of the variables is added into the regression

| | Dependent Variable: *avgprice* | | | |
|---|---|---|---|---|
| Variable | (1) | (2) | (3) | (4) |
| *const* | $0.000^{***}$ | 0.000 | 0.000 | 0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| *black_%* | 0.021 | -0.002 | -0.002 | -0.002 |
| | (0.023) | (0.023) | (0.023) | (0.023) |
| *white_%* | 0.006 | 0.002 | -0.000 | -0.000 |
| | (0.022) | (0.022) | (0.022) | (0.022) |
| *gray_%* | $0.039^{*}$ | 0.000 | -0.000 | -0.000 |
| | (0.023) | (0.023) | (0.023) | (0.023) |
| *green_%* | -0.036 | -0.001 | -0.003 | -0.003 |
| | (0.023) | (0.023) | (0.023) | (0.023) |
| *blue_%* | 0.011 | 0.001 | 0.002 | 0.002 |
| | (0.024) | (0.023) | (0.023) | (0.023) |
| *yellow_%* | $-0.053^{**}$ | -0.001 | -0.002 | -0.002 |
| | (0.023) | (0.023) | (0.023) | (0.023) |
| *purple_%* | -0.028 | -0.002 | -0.002 | -0.003 |
| | (0.022) | (0.022) | (0.022) | (0.022) |
| *orange_%* | -0.004 | -0.001 | -0.001 | -0.001 |
| | (0.025) | (0.025) | (0.025) | (0.025) |
| *norm_shannon_entropy* | $0.080^{***}$ | 0.000 | 0.000 | 0.001 |
| | (0.027) | (0.028) | (0.028) | (0.028) |
| *BDM* | $0.074^{***}$ | 0.004 | 0.004 | 0.004 |
| | (0.023) | (0.024) | (0.024) | (0.024) |
| *num_of_colors* | $0.062^{**}$ | -0.001 | 0.001 | 0.001 |
| | (0.029) | (0.029) | (0.029) | (0.029) |
| *year_collection_creation_2019* | | 0.005 | 0.003 | 0.004 |
| | | (0.054) | (0.055) | (0.056) |
| *year_collection_creation_2021* | | 0.011 | 0.023 | 0.023 |
| | | (0.054) | (0.063) | (0.063) |
| *year_collection_creation_2022* | | 0.005 | 0.012 | 0.012 |
| | | (0.037) | (0.042) | (0.042) |
| *year_last_sale_2019* | | -0.003 | -0.003 | -0.003 |
| | | (0.022) | (0.022) | (0.022) |
| *year_last_sale_2021* | | 0.019 | 0.018 | 0.018 |
| | | (0.054) | (0.054) | (0.054) |
| *year_last_sale_2022* | | 0.023 | 0.018 | 0.019 |
| | | (0.053) | (0.053) | (0.053) |
| *floor_price* | | $0.159^{***}$ | $0.162^{***}$ | $0.162^{***}$ |
| | | (0.057) | (0.057) | (0.057) |
| *last_price* | | $0.513^{***}$ | $0.511^{***}$ | $0.511^{***}$ |
| | | (0.060) | (0.060) | (0.060) |
| *floor_price_collection* | | 0.010 | 0.011 | 0.011 |
| | | (0.026) | (0.028) | (0.029) |
| *timediff* | | 0.014 | 0.027 | 0.027 |
| | | (0.027) | (0.034) | (0.034) |
| *nft_type_erc721* | | -0.000 | -0.000 | -0.001 |
| | | (0.021) | (0.022) | (0.022) |
| *platform_last_sale_LooksRare* | | 0.001 | 0.001 | 0.001 |
| | | (0.019) | (0.019) | (0.019) |
| *platform_last_sale_Rarible* | | -0.001 | -0.000 | -0.000 |
| | | (0.023) | (0.023) | (0.023) |
| *platform_last_sale_SuperRare* | | 0.003 | 0.012 | 0.014 |
| | | (0.041) | (0.049) | (0.051) |
| *BGCI* | | 0.008 | 0.004 | 0.004 |
| | | (0.023) | (0.024) | (0.024) |
| *twitter_account* | | | 0.002 | 0.002 |
| | | | (0.022) | (0.023) |
| *following_count* | | | -0.011 | -0.009 |
| | | | (0.042) | (0.044) |
| *num_tweets* | | | 0.002 | 0.000 |
| | | | (0.052) | (0.052) |
| *listed_count* | | | 0.003 | 0.002 |
| | | | (0.032) | (0.033) |
| *verified* | | | -0.011 | -0.011 |
| | | | (0.043) | (0.047) |
| *reply_count_month* | | | -0.006 | -0.006 |
| | | | (0.041) | (0.041) |
| *discord_account* | | | | -0.002 |
| | | | | (0.041) |
| *d_ann_messages_month* | | | | 0.012 |
| | | | | (0.060) |
| *d_gen_messages_month* | | | | -0.009 |
| | | | | (0.057) |
| BIC | -792623.65 | -798225.58 | -801597.76 | -802409.78 |
| AIC | 110456.72 | 104780.46 | 101352.55 | 100512.65 |
| Deviance | 199542.03 | 189914.68 | 189910.04 | 189909.67 |
| Residual Std. Error | 2.967(df = 22666) | 2.896(df = 22650) | 2.896(df = 22644) | 2.896(df = 22641) |
| F Statistic | (df = 11; 22666) | (df = 27; 22650) | (df = 33; 22644) | (df = 36; 22641) |

**Note:** Robust Standard Errors in parenthesis.
$***\rho < 0.01$, $**\rho < 0.05$, $*\rho < 0.1$