Pricing Dynamics and Herding Behavior of NFTs

Gilbert Fridgen* Roman Kräussl‡†

Orestis Papageorgiou* Alessandro Tugnetti‡

August 12, 2023

Abstract

Non-fungible tokens (NFTs) took the digital art space by storm in 2021, generating massive amounts of volume and attracting a large number of users to a previously obscure part of blockchain technology. Still, very little is known about the determinants that contribute to NFT pricing and the market dynamics. This paper attempts to evaluate these factors analyzing 860,067 art NFTs that have been deployed on the Ethereum blockchain and have been involved in 317,950 sales. We introduce a first-of-its-kind ranking system for NFT artists, adding a novel dimension to our analysis. Our results suggest that market liquidity and intrinsic value strongly predict average price of NFTs. Contrarily, social media activity surprisingly shows a negative association with NFT prices. We also uncover the dominance of a small group of artists, demonstrating the pronounced 'superstar effect' and herding behavior within the NFT market. This market dynamic, however, shows signs of shifting towards independent decision-making during extreme market volatility.

Keywords— Non-fungible tokens (NFTs), Price Index, Herding

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, 2021). Inevitably, such popularity also attracted substantial criticism about their utility and whether they can revolutionize the digital asset sector. However, NFTs are not just a fleeting trend; they are part of a broader technological evolution in the blockchain space. The crash in 2022 might be viewed as a correction, a recalibration of the true underlying value of NFTs, in line with a more comprehensive perspective of market dynamics. The drop in value does not undermine the inherent properties that make NFTs unique; instead, it may pave the way for a more sustainable growth pattern that aligns with long-term financial models. These criticisms 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, 2023) 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 (Sotheby's, 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, Dowling (2022a) provides a pioneering exploration of Decentraland LAND NFTs, drawing a mental association between virtual and physical land, and identifying possible inefficiencies in pricing behavior akin to early cryptocurrency pricing. This notion of inefficiency is further examined in a broader perspective by Cheah and Fry (2015) and Urquhart (2016), underlining the nascent nature of NFT markets. Schaar and Kampakis (2022) take a quantitative approach to investigate the CryptoPunks collection, highlighting an impressive average monthly return of 34.19% over three years and the significant role that rarity plays in determining prices. This emphasis on rarity is also central to the work of Mekacher et al. (2022), who delve into 410 collections to demonstrate that rarer NFTs command higher prices and are less risky. Horky et al. (2022) employ a combination of econometric tools and machine learning in their study of the digital art market through SuperRare, finding that Hedonic pricing models furnish valuable insights into NFT prices, independent of cryptocurrencies.

Broadening the scope to encompass the entire market, Dowling (2022b) work stands out by interrogating the connection between NFT pricing and cryptocurrencies, uncovering limited volatility transmission effects but notable co-movement between the two. Ante (2022) adds nuance to this relationship by revealing how Bitcoin price shocks trigger an increase in NFT sales, while Ether price shocks have a converse effect. Borri et al. (2023) take an ambitious step by constructing a comprehensive dataset for the overall NFT market, unearthing the nuanced interplay between creator fees, rarity, visual characteristics, and prices. Kong and Lin (2021) enrich this perspective by emphasizing the role of well-connected investors in the pricing dynamics of NFTs. Mazur (2021) shifts the focus to the risk and return characteristics of NFT-based startups, illustrating a staggering investment multiple of 40 over the long term. In a groundbreaking study, Nadini et al. (2021) map the structure and evolution of the entire NFT market, identifying sale history and visual features as key price predictors. Hemenway Falk et al. (2022) introduce a novel dimension by exploring the value of creator royalties, uncovering their potential to reshape the NFT market through risk-sharing, dynamic pricing, and price discrimination.

While the existing literature on NFT pricing and determinants has provided a comprehensive view of the market, the exploration of herding behavior in the context of art NFTs presents a unique and uncharted territory. Herding behavior, as observed in traditional financial markets (Banerjee, 1992; Shiller, 1995; Cipriani and Guarino, 2008), refers to the inclination of investors to follow the majority's actions rather than making independent decisions based on intrinsic value. Art NFTs offer a perfect primer to study this phenomenon, primarily due to their parallel nature with the physical art market and the novel complexities they introduce.

In the traditional art market, the phenomenon of herding behavior has been well-documented. Azarmi and Menny (2013) study of the fine arts market explores the tendency of investors to gravitate towards well-established artists, leaving the quality of the work secondary to the historical auction performance of the artist. They found that a small fraction of artists dominates financial activity and that contemporary and less-established artists are subject to more herding behavior. This observation reflects a pattern where investors follow trends and the actions of others, often leading to inflation in the value of specific artists or styles. Art NFTs amplify these dynamics by adding layers of complexity. Unlike physical art, where authenticity, provenance, and physical condition play a role in valuation, art NFTs operate on a digital platform where these factors are replaced by digital scarcity, network effects, and integration with cryptocurrencies. The rapid rise of art NFTs, marked by astronomical prices and intense media coverage, has many resemblances to the herding behavior observed in the traditional art market. Investors and collectors rush towards certain NFTs, often driven more by hype and the actions of others than a deep understanding of the digital art's inherent value.

Moreover, the volatility and correlation with the broader cryptocurrency market add to the intrigue of art NFTs. The fluctuations in prices, often driven by external factors such as changes in Ethereum or Bitcoin values, create a turbulent environment where herding behavior can thrive. The allure of quick profits, media attention, and the novelty of owning a unique digital asset can lead to a cascade effect where one investment decision triggers another, often without a rational analysis of the asset's underlying value. The work of Bao et al. (2022) provides the first empirical evidence on herding in the NFT market. They identify three waves of herding in the NFT market, with daily market returns becoming more volatile during these periods. Their findings also reveal that herding is more likely to emerge as the proportion of newcomers increases and that media exposure drives investors' attention when herding arises. The dynamics of herding in the NFT market were further explored, showing a connection with the return on Ethereum but a diminishing effect with the return on Bitcoin. Yousaf and Yarovaya (2022) examine herding behavior across three cryptocurrency classes, including NFTs. Their time-varying analysis identified herding in conventional cryptocurrencies and DeFi assets during the most recent bubble of 2021 but failed to demonstrate evidence of herding in NFTs during various market conditions. This contrast presents an intricate picture of the NFT market, suggesting that herding behavior may manifest differently across various aspects of the broader cryptocurrency space. The speculative nature of NFTs and DeFi markets, often perceived as bubble behaviors, has been explored by Wang et al. (2022). They document that both NFT and DeFi markets exhibit speculative bubbles, with NFT bubbles being more recurrent and having higher explosive magnitudes. The bubbles in these markets are correlated with market hype and broader cryptocurrency market uncertainty, though they also recognize periods where bubbles are not detected, indicating intrinsic value in these markets.

NFT enthusiasts argue that the characteristics observed, such as provable digital scarcity, are traits of most industries with limited track records, and that NFTs have unique attributes with transformative potential. This nexus between herding behavior, speculative bubbles, and the inherent value of NFTs opens up avenues for understanding the art market's financial dynamics and the broader economic landscape. The integration of insights from traditional herding behavior theories (*e.g.*, Banerjee, 1992; Shiller, 1995) with contemporary studies on NFTs and art markets provides a nuanced understanding of market dynamics. This interdisciplinary approach is instrumental in drawing parallels between classical assets like securities, physical art, and collectibles, and emerging digital assets like NFTs.

What sets NFTs apart is the unprecedented access to a wealth of real-time data about transactions and financial characteristics of these digital assets. Unlike traditional markets where information might be fragmented, delayed, or obscured by various market frictions, the blockchain technology underpinning NFTs ensures that every transaction is transparent, timestamped, and publicly accessible. The ability to track and analyze these transactions in real-time opens up new avenues for research, allowing for a more nuanced understanding of market dynamics, pricing mechanisms, and investor behavior. As an example, researchers can study market frictions in a way that was previously unattainable. Market frictions, such as transaction costs, information asymmetry, and liquidity constraints, play a critical role in asset pricing and investment strategies. They have been studied extensively in traditional financial markets by authors like Amihud and Mendelson (1986), Stiglitz (1989) and Vayanos and Wang (2012). In the context of NFTs, these frictions might manifest differently, and their impact on market dynamics could be distinct from what is observed in traditional markets. By leveraging the granular data available through NFT transactions, this paper aims to explore how market frictions operate in a digital asset environment, and how they may affect pricing, liquidity, and investor behavior. Specifically, the goal of this paper is to navigate the complex interplay between the pricing of art NFTs and the herding behavior that characterizes their market dynamics, thus bridging the gap between the two streams of literature presented before. 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. Our sample comprises 860,067 art NFTs that have been deployed on the Ethereum blockchain and have been involved in 317,950 sales.

Our approach integrate hedonic and repeat sales methods, with corrections for sample selection bias, to dissect the multifarious factors contributing to the pricing structure of art NFTs. Simultaneously, we delve into herding behavior bias, analyzing it from both the artist's perspective and the transactional viewpoint. Our findings confirm the positive correlation between the number of trades and the floor price of NFTs with their average price, elucidating critical elements of the NFT market's pricing structure. Further, our data suggests an unexpected inverse relationship between social media activity and NFT prices, presenting a challenge to popular conceptions of the NFT market. Remarkably, our exploration of herding behavior revealed a small group of artists who command most of the market's activity, mirroring trends seen in traditional art markets, while also spotlighting the market's susceptibility to potential speculative bubbles.

These insights into the NFT art market's pricing dynamics and liquidity constraints represent a significant contribution to the existing financial literature, providing a rich context for understanding a new asset class within a blockchain-enabled environment. The complex interplay between intrinsic and extrinsic factors, coupled with the unique attributes of NFTs, adds new dimensions to traditional financial theories. This exploration opens up new avenues for academic inquiry, extending classic finance models to a digital frontier, and offers practical implications for investors, collectors, platforms, and regulators seeking to navigate the burgeoning NFT landscape (Glaser et al., 2014; Dyhrberg, 2016). The synthesized understanding from our study aligns with the evolving discourse in financial literature and marks a vital step toward a more nuanced comprehension of NFTs and their role within the broader financial ecosystem.

The remainder of this paper is structured as follows. Section 2 discusses the data extraction techniques. Section 3 describes the econometric models used to construct the price indexes and to study the herd bias. Section 4 discusses our findings. Section 5 concludes the paper.

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. Our study's methodology is particularly groundbreaking as it merges financial, descriptive, and social network features of art NFTs—a first-of-its-kind approach in the financial literature. By integrating these diverse aspects, we strive to provide a more comprehensive understanding of the art NFT market. Please, refer to Section A.1 in Appendix A for a detailed explanation of data extraction, cleaning and preparation processes.

Our sample comprises 860,067 art NFTs that have been deployed on the Ethereum blockchain and have been involved in 317,950 sales between 15^{th} July 2018 and 10^{th} February 2022. To facilitate a deeper understanding of this extensive dataset, we have organized the information into two distinct tables. Table 1 presents the descriptive statistics, summarizing key features such as median, standard deviation, and other relevant statistical measures that provide an overall picture of the art NFT market. These statistics offer insights into the general trends, distributions, and characteristics that define the market landscape. Table A1, on the other hand, provides a detailed breakdown of the variables used in our analysis.

Each variable is accompanied by a concise explanation, outlining its relevance and how it contributes to our understanding of the pricing and herding behavior in the NFT art space.

[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, 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.Furthermore, the application of the Jarque-Bera (JB) and KPSS tests to our data uncovers intricate statistical properties. Notable deviations from normality and the existence of unit roots in continuous variables such as *BGCI_return* and *num_trades* emphasize the non-linear characteristics of the art NFT market. The JB statistics reveal pronounced skewness and kurtosis, corroborating the heterogeneous quality of the data. This is in harmony with the fragmented and multifarious nature of the NFT market (Caporale et al., 2021).

The variables *floor_price* and *last_price* highlight the extensive range of pricing that characterizes the art NFT market. This dynamic pricing behavior correlates with observations in financial literature, and it frequently results from complex factors such as rarity and artist recognition. Rarity, in particular, acts as a recognized determinant of value in both conventional and digital art markets, creating a perception of exclusivity and uniqueness (Renneboog and Spaenjers, 2013; Schaar and Kampakis, 2022). In the realm of NFTs, rarity's definition further refines and authenticates via blockchain technology, thereby augmenting its allure. Conversely, artist recognition plays an instrumental role in the valuation of art pieces, and eminent artists demand elevated prices that reflect their consolidated reputation and brand (Mandel, 2009).

An examination of the discrete variables panel provides compelling insights into the structural composition of the NFT ecosystem. The prevalence of platforms such as OpenSea might be ascribed to its accessible interface, varied assortment of listed NFTs, and robust community engagement. The triumph of the platform is congruent with academic research on network effects, where platforms that captivate more users gain advantages in terms of augmented liquidity and information dissemination, thereby fortifying their market dominance (Parker and Van Alstyne, 2005). Likewise, the inclination for marketplace collections, as exemplified by Foundation (FND), over individual artists might be anchored in the trust and credibility typically associated with renowned collections. Foundation, to illustrate, curates pieces from distinguished artists and facilitates a more seamless acquisition process. This predilection aligns with scholarly research concerning consumer behavior, where branded collections often evoke greater trust and thus attract a larger number of buyers (Aaker, 1996).

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.

The observed illiquidity and fragmentation in the NFT art market, as depicted in our study, resonate with characteristics commonly associated with emerging markets. Illiquidity refers to the scarcity of buyers and sellers in the market, leading to wider bid-ask spreads, higher transaction costs, and difficulty in executing large trades without significant price impact (Amihud and Mendelson, 1986; Bekaert et al., 2007). Fragmentation, on the other hand, implies a market divided into smaller disconnected segments, often due to lack of standardization or regulatory disparities (Madhavan, 2000). These features are consistent with a market in its infancy, where high volatility and uncertainty prevail (Schwert, 1989), as seen in the early stages of many financial markets.

In the context of multi-sold NFTs, our analysis reveals intriguing patterns, such as each successive sale taking less time on average and the profit margin diminishing with each sale. These trends suggest a dynamic market shaped by a blend of intrinsic and extrinsic factors. In particular, extrinsic factors, such as market hype and overall crypto market trends, add further complexity to pricing dynamics. Similar to phenomena observed during technology bubbles (Ofek and Richardson, 2003) or the behavior of meme stocks (Costola et al., 2021), the influence of social media, influencers, and broader market sentiment can create rapid price swings and speculative behavior in the NFT market.

3 Approach

The methodology section of our analysis is split into two distinct parts, each addressing a unique aspect of the art NFT market. In the first part, we delve into the price dynamics of art NFTs, adopting a two-sided approach that employs both hedonic and repeat-sales regression models. The hedonic pricing model is a revealed preference method used to estimate the influence of various features on the price of a good, and it is particularly suited for studying markets where the goods being transacted are highly differentiated, like the art NFT market. The repeat-sales regression model, on the other hand, focuses on items that have been sold more than once, thus allowing us to control for the intrinsic qualities of each item and isolate the pure price effect over time.

The second section of the methodology focuses on the study of herding bias within the art NFT market. This involves examining the tendency of investors to follow the actions or strategies of a larger group, potentially leading to collective behavior that may not necessarily align with individual information or preferences. The investigation into herding bias sheds light on the social and psychological factors driving investment decisions in the burgeoning NFT space, offering a nuanced understanding of market dynamics and investor behavior.

3.1 Pricing Dynamics

The hedonic regression approach applied to the study of art NFTs price dynamics is rooted in the consumer theory of utility maximization, where NFTs are considered bundles of characteristics, and the price reflects the consumer's willingness to pay for these characteristics (Lancaster, 1966; Rosen, 1974). This approach has been widely used to study price determinants in various markets, such as real estate (Muehlenbachs et al., 2015), traditional art (Chanel et al., 1996; Renneboog and Spaenjers, 2013; Adams et al., 2021), and more recently, in the analysis of digital assets like NFTs (Borri et al., 2023; Horky et al., 2022).

For the analysis of art NFTs in this study, we commenced by employing a hedonic regression model, described by the following equation:

$$\ln(y+1) = \beta_0 + \boldsymbol{\beta}\mathbf{X} + \gamma D + \delta T \ + \epsilon \tag{1}$$

where y is the average price of the NFT, $\mathbf{X}$ represents a vector of continuous variables, D denotes control variables, T stands for time coefficients, and ϵ is the error term. Here, the continuous variables include features such as color proportions, Shannon's entropy, and the total number of colors present, reflecting both aesthetic and complexity aspects of the NFTs. The control variables represents a set of dummy variables accounting for intrinsic characteristics of the digital assets and presence on social medias. Time coefficients are incorporated to construct a price index, accounting for temporal variations and market trends.

To correct the potential selection bias in observed price data, a Heckman two-stage regression was performed. Selection bias arises when the sample selection for observation is not random, and there is a correlation between the observable selection and the unobservable factors affecting the dependent variable. In the context of art NFTs, this bias can occur when only the prices of traded NFTs are observed, while untraded NFTs remain unobserved. This non-random selection can create a bias in estimating the relationship between price and characteristics of the NFTs, as the unobserved factors affecting the decision to trade might also influence the price.

The inverse Mills ratio (IMR) is derived from the Probit estimation and encapsulates the likelihood of the NFT being traded. This IMR is incorporated into the hedonic regression as a new variable, λ, ensuring unbiased estimates:

$$\lambda = \frac{\phi(\gamma \mathbf{X})}{\Phi(\gamma \mathbf{X})} \tag{2}$$

The second stage involved the incorporation of the IMR into the hedonic regression. Indeed, by performing this, the model corrects for the selection bias by capturing the unobserved factors affecting both the selection process and the dependent variable:

$$\ln(y+1) = \beta_0 + \boldsymbol{\beta}\mathbf{X} + \gamma D + \delta T + \theta\lambda + \epsilon \tag{3}$$

The two-stage approach ensures that the relationship between the price and characteristics of NFTs is estimated without the bias induced by non-random selection.

Multicollinearity was also addressed using the Variance Inflation Factor (VIF) method, ensuring that the explanatory variables were not highly correlated (O'brien, 2007).

In the context of hedonic modeling for art NFT prices, the time-varying and invariant variables are crucial in determining the underlying price index. However, it is vital to ensure that the hedonic regression parameters are stable over time. If not, this instability can lead to biased estimates of time dummy coefficients and consequent price indexes (Kuminoff et al., 2010).

To mitigate this issue, we employ the chained Fisher index, derived from sectional hedonic regressions, following the methodology used in classical financial literature (Diewert, 1976; Triplett, 2004). Specifically, we construct the Fisher index by first calculating the Laspeyres and Paasche indexes and then obtaining the geometric mean of these two indexes.

Given the estimated β coefficients from the hedonic regression and the normalized weights q for each characteristic, the Laspeyres index for a given period t is calculated as:

$$\text{Laspeyres}_t = \frac{\sum_{j=1}^{J} \beta_{j,t+1} q_{j,t}}{\sum_{j=1}^{J} \beta_{j,t} q_{j,t}} \tag{4}$$

Similarly, the Paasche index for the same period is determined as:

$$\text{Paasche}_t = \frac{\sum_{j=1}^{J} \beta_{j,t+1} q_{j,t+1}}{\sum_{j=1}^{J} \beta_{j,t} q_{j,t+1}} \tag{5}$$

The Fisher index is then the geometric mean of these two indexes:

$$F_{t+1} = \sqrt{\text{Laspeyres}_t \times \text{Paasche}_t} \tag{6}$$

The Fisher index can be chained across time to generate a consistent price index for art NFTs. By calculating the Fisher index in this manner, the analysis ensures an unbiased understanding of pricing mechanisms in the evolving field of art NFTs, accommodating potential time instability in the hedonic regression parameters and offering a distinct perspective on market behavior.

As a final step, we employ repeat sales regression (RSR) index. Repeat sales regression focuses specifically on the items that have been sold multiple times, thereby controlling for the unobservable heterogeneity of those assets (Bailey et al., 1963). This approach offers a complementary perspective to hedonic regression by capturing price changes for identical NFTs, thus eliminating the need to control for detailed characteristics. Initially, the regression model is specified as:

$$\ln(y+1) = \alpha + \sum_{t=1}^{T} \beta_t D_t + \epsilon_t \tag{7}$$

where β_t represents the coefficients for the time dummies D_t, and ϵ_t is the error term.

However, a significant concern in repeat sales regression is, again, the potential selection bias and the heteroskedasticity of the error term. Indeed, as specified before, in repeat sales regression, the sample only consists of properties that have been sold multiple times. This specific selection can introduce bias since properties that are sold more than once may not be a random subset of all properties. They might have unique characteristics that make them more or less likely to be resold, leading to a systematic deviation from the overall population.

To correct for this bias and account for the varying holding periods between repeat sales, we utilize the Case-Shiller three-stage regression method (Case and Shiller, 1989).

We begin by conducting the repeat sales regression as previously described, extracting the residuals. This initial step lays the foundation for modeling the heteroskedasticity in the error term. Heteroskedasticity refers to the situation where the variability of the error term is not constant across all levels of the explanatory variables. In the context of repeat sales regression, this non-constant variability could be related to factors like the holding period, which might have varying effects on the error term across different observations. For this reason, we regress the squared residuals against a constant and the specific holding periods of art NFTs:

$$u_t^2 = \gamma_0 + \gamma_1 \times holding\_period + \zeta_t \tag{8}$$

By modeling the squared residuals in this manner, we capture the pattern of heteroskedasticity and its relationship with the holding period. Finally, we fit the original repeat sales regression using Generalized Linear Models (GLMs) with weights derived from the square root of the fitted values from the second stage. These weights are instrumental in accounting for the heteroskedasticity in the residuals, resulting in more efficient and unbiased estimates.

Compared to the hedonic approach, it provides a more direct insight into price changes without the need to model detailed characteristics. Furthermore, the combination of hedonic and repeat sales regression provides a comprehensive understanding of the art NFT market, leveraging the strengths of both methodologies to capture the dynamics of price determination (Goetzmann, 1992; Mei and Moses, 2002; Whitaker and Kräussl, 2020).

3.2 Herd Bias

The notion of herding behavior, as posited by the cascade theory, suggests that individuals in a market may disregard their private information and follow the decisions of predecessors. In the context of the NFT art market, this can manifest in the form of preferences aligning with recognized or leading artists, thereby creating self-reinforcing patterns of demand. Such mechanisms can lead to a superstar phenomenon where a small number of artists dominate the market. This phenomenon has been empirically investigated in traditional art markets (Rosen, 1981), but the exploration in the digital art space is novel and adds a layer of complexity due to the uniqueness of NFTs.

Our analytical framework encompasses three primary dimensions: artistic quality as reflected by market metrics, herding behavior rooted in classical financial theories, and the intricate relationship between historical performance, popularity, and influence factors like social media engagement. This last dimension, particularly, provides insights into how external elements, such as an artist's Discord account presence or deployer status, interact with traditional market dynamics to shape an artist's ranking and demand. By leveraging Azarmi and Menny (2013) multinomial logit model, we collectively analyze consumer choices based on the individual market shares of artists rather than focusing on separate purchasing decisions.

Our analytical representation takes the following form:

$$\log\left(\Pi_{it}\right) = \alpha_i + \alpha_t + \sum_{j=1}^{J} \beta_j \left(X_{ijt} - \bar{X}_{jt}\right) + \epsilon_{it} \tag{9}$$

Here, Π_{it} symbolizes the selection made by art enthusiasts and is rooted in the relative demand attributed to artist i at time instance t. We define this relative demand as:

$$\Pi_{it} = \frac{S_{it}}{\widetilde{S}_t} = \frac{\frac{V_{it}}{V_t}}{\sqrt[n]{\prod_{j=1}^{n} \frac{V_{jt}}{V_t}}} \tag{10}$$

In this mathematical construct, the relative demand Π_{it} is derived by dividing an artist's market share S_{it} by the geometric mean of the individual market shares across all artists, represented by $\widetilde{S}_t$. The term S_t is ascertained by dividing the monthly market capitalization V_{it} specific to an artist by the cumulative market capitalization V_t for all n artists at the time t.

Within the context of our model, we consider each sample as an isolated market, premised on the assumption that consumers do not harbor predetermined preferences for a specific artist or art form within a given sample category. For example, an individual evaluating works from a particular NFT artist would exhibit equal interest in diverse artistic expressions, irrespective of the creator. The intrinsic value of art by artist i is symbolized by α_i, the k-th explanatory variable is denoted by X_{ikt}, and ϵ_t refers to the error term of the model. The corresponding arithmetic means of these underlying variables are represented by $\bar{\alpha}$, $\bar{X}_{kt}$, and $\bar{\varepsilon}_t$.

As explanatory variable, the cumulative score yearly ranking lagged by one period captures the inertia in an artist's popularity, reflecting how past performance continues to affect current standings. The yearly rank of the artist and the previous monthly rank provide insights into the temporal dynamics of an artist's market position. The interaction terms in the regression model serve specific purposes in understanding the dynamics of the NFT art market. The complexity of the NFT art market is further dissected through specific interaction terms in the model. The term involving yearly rank and Discord account (*yr*discord*) captures the influence of social media engagement on an artist's rank, reflecting how an active presence on platforms like Discord can amplify the artist's visibility and market appeal. On the other hand, the interaction between yearly rank and deployer artist status (*yr*deployer_artist*) distinguishes between artists who are independent creators and those associated with companies. This distinction is pivotal in understanding how the nature of the deployer, whether an individual artist or a company, interacts with the artist's rank to affect demand. In the NFT art market, the interaction effects help explain how information cascades might form. For instance, a high yearly rank, coupled with active engagement on Discord, might create a positive feedback loop where popularity leads to more visibility, further enhancing demand. Conversely, the deployer's nature might affect how information about the artist disseminates, with individual artists potentially benefiting from a more personalized connection with collectors.

We complement and enrich the study of herding behavior and artist stardom by examining Cross-Sectional Absolute Deviation (CSAD). This method serves as an analytical tool for discerning how market participants adapt to fluctuations within the market. By evaluating the cumulative sum of deviations from a certain trend or mean, we identify instances of herding behavior, where investors are inclined to follow prevailing trends rather than base their decisions on fundamental analysis. Such herding behavior forms a common thread between the stardom effect and cascade theory, as well as the study of CSAD. These interconnected areas pertain to the manner in which investors or market participants might succumb to trends, overlook private information, or make determinations grounded in the conduct of others, rather than a rational evaluation of value. In the context of NFTs, this exploration gains particular relevance. The NFT market, being relatively nascent and heavily swayed by social dynamics, celebrity endorsements, and market hype, often leads to behaviors where individuals align with the crowd instead of pursuing independent analysis. We recognize this alignment as a phenomenon where a few artists dominate the market, a pattern that we meticulously trace through the application of CSAD.

Initially, we adapt the CSAD measure, as formulated by Chang et al. (2000):

$$\mathrm{CSAD}_t = \frac{1}{N} \sum_{i=1}^{N} \left| R_{i,t} - R_{m,t} \right|, \tag{11}$$

Further, to enhance the robustness of the findings and capture the dynamic interactions in the market, we extend the methodology by introducing a lagged dependent variable, as advised by Fu and Wu (2021):

$$\mathrm{CSAD}_t = \alpha + \gamma_1 \left|R_{m,t}\right| + \gamma_2 R_{m,t}^2 + \gamma_3 \mathrm{CSAD}_{t-1} + \varepsilon_t. \tag{12}$$

The convergence between individual and market returns in this model reflects the manifestation of artist stardom, where market preferences tend to converge towards specific artists.

To delve deeper into the dynamics of herding, we employ the fixed transition probabilities Markov-switching model (FTP-MS) (Diebold and Rudebusch, 1999), integrating the methodology developed by Bao et al. (2022). This approach enables the distinction between different market regimes, providing insights into how herding behavior fluctuates over time:

$$\mathrm{CSAD}_t = \begin{cases} \alpha + \gamma_{1,1} \left|R_{m,t}\right| + \gamma_{2,1} R_{m,t}^2 + \gamma_3 \mathrm{CSAD}_{t-1} + \varepsilon_{1,t}, & \text{if } s_t = 1, \\ \alpha + \gamma_{1,2} \left|R_{m,t}\right| + \gamma_{2,2} R_{m,t}^2 + \gamma_3 \mathrm{CSAD}_{t-1} + \varepsilon_{2,t}, & \text{if } s_t = 2, \end{cases} \tag{13}$$

with transition probabilities expressed as:

$$P\left(s_{t+1} = i \mid s_t = j\right) = p_{ij}, \quad i, j \in \{1, 2\}$$

The transition matrix for this first-order Markov-switching process takes the form:

$$P = \begin{pmatrix} P_{1,1} & 1 - P_{2,2} \\ 1 - P_{1,1} & P_{2,2} \end{pmatrix}$$

where $0 < P_{i,i} < 1$ and $0 < P_{j,j} < 1$. In this context, regime 1 represents the state with herding, characterized by a negative $\gamma_{2,1}$, while regime 2 denotes the state without herding, indicated by positive or insignificant $\gamma_{2,2}$.

The constant term α serves as a non-switching parameter, symbolizing the long-term average of CSAD_t, while the transition probabilities, $P\left(s_{t+1} = i \mid s_t = j\right)$, govern the progression between these two states. The probability $1 - P_{2,2}$ specifically elucidates the likelihood of the emergence of herding behavior.

By employing this Markov-switching framework, we are able to estimate the time-varying probability of herding existence. This approach complete the understanding of how herding mechanisms operate within the NFT market, reflecting the complex interplay between individual preferences and market-wide trends. It adds depth to our analysis of artist stardom, illustrating how certain artists might become market leaders, thereby shaping the dynamics of preference alignment and demand cascades in the digital art space.

4 Discussion of Results

4.1 Price Indeces Design

Table 2 illustrates our hedonic regression results. We identify several significant pricing determinants for art NFTs. A pivotal outcome, observed consistently across all our models, is the positive and significant coefficient on the variables log($number\_of\_trades$) and $log(floor\_price)$, with values around 1.08 and 0.92 respectively. This suggests that NFTs with a higher number of trades and a higher floor price are likely to command a higher average price. This finding aligns with existing literature on asset pricing, such as Campbell and Shiller (1988) and Fama and French (1993), highlighting the relevance of these factors in financial markets. In particular, the variable $log(floor\_price)$ may be capturing the price floor effect, suggesting that the minimum acceptable price for the NFT could be a significant determinant of its average price. As Kahneman and Tversky (1979) illustrate, this minimum price often sets a psychological benchmark or reference point for transactions, influencing final prices. In a similar vein, the role of a base price or minimum price as a determinant of the final price of an asset has been emphasized by Campbell and Shiller (1988). In the context of our study, a higher floor price could convey a higher perceived value of the NFT, resulting in a higher average price. Further, the floor price might encapsulate the price floor effect. In auction theory, Milgrom and Weber (1982) underscore the importance of the reserve price (the lowest price at which the seller is willing to sell the item) in determining the final price. This concept could apply to NFTs as well, where the floor price acts as a reserve price and impacts the final price of the NFT.

On the other hand, the negative coefficients on the variables $\log(following\_count)$, $\log(quote\_count\_month)$, and $\log(post\_count\_month)$ are indeed unexpected, given that one would typically expect higher levels of social media activity to positively influence the price of NFTs. However, our regression results suggest that this is not the case for the NFT market, at least within the context of the observed data. One possible explanation for these counterintuitive results is the oversaturation hypothesis. Essentially, when an NFT or a collection of NFTs is excessively quoted or posted about on social media, it could lead to oversaturation of information. This oversaturation could reduce the perceived rarity or exclusivity of the NFTs, thereby driving down its price. The concept of rarity is particularly important in the NFT market, as the unique and non-reproducible nature of NFTs is a key part of their value proposition. This idea aligns with the findings of Dellarocas (2003), which discusses the potential downsides of excessive online exposure. Oversaturation of information can overwhelm potential buyers, dilute the perceived value, and even lead to skepticism or mistrust if the promotional activity appears too aggressive or inauthentic. It is also worth noting that the negative relationship between social media activity and NFT prices could be specific to the type of activity being measured. For instance, while the number of followers, quote counts, and post counts might be negatively associated with NFT prices, other types of social media activity or engagement, such as likes or shares, could potentially have a positive impact on prices.

The application of Heckman's two-step model, as exhibited in models $1'$, $2'$, and $3'$, reveals compelling insights about the potential selection bias in the study of art NFT pricing dynamics. The Inverse Mills ratio (λ), derived from the Probit model in the first step of the Heckman process, is significantly negative, confirming the presence of selection bias in the sample of traded NFTs. The negative sign of the Mills ratio suggests that the unobserved factors influencing the decision to trade NFTs are inversely correlated with the error term in the average price equation. This implies that NFTs less likely to be traded (due to unobserved factors) tend to have higher prices than what could be predicted based solely on their observable characteristics. In conclusion, higher-priced NFTs, which are less likely to be traded, might contain superior artistic or unique features not captured by the observable characteristics.

In terms of the robustness of the results, the consistency of most variable coefficients across both the hedonic and Heckman models indicates that our findings are resilient to adjustments for selection bias. However, some variables, such as $BGCI\_return$, demonstrate significant changes in their coefficients when transitioning from the hedonic models to the Heckman models. The coefficient substantial increase suggests that the relationship between these variables and NFT prices is subject to the influence of selection bias. As such, the return on the Galaxy Crypto Index could be capturing some of the unobserved factors affecting the decision to trade NFTs. Previous studies have found a significant impact of selection bias on the relationship between asset prices and market-wide indicators (Ang et al., 2020; Bianchi and Babiak, 2022). Moreover, its pronounced positive coefficient in our model suggests a correlation between the performance of cryptocurrencies, as measured by the Bloomberg Galaxy Crypto Index, and the pricing of NFTs. This correlation reinforces the idea that NFTs, as a subset of the broader crypto asset market, are likely influenced by the market dynamics and investor sentiment in the cryptocurrency space.

[Table 2 about here.]

The application of the repeat sales regression (RSR) model drastically reduced the number of observations from 190,201 to 53,027, a reduction of approximately 72%. This is expected as the RSR model only considers assets that have been traded more than once. This is a crucial characteristic of the RSR model, which leverages repeated sales data to control for the unobserved time-invariant characteristics of the assets. Furthermore, Figure 1 draws the art NFT price indices. We observe that all four indices - the Hedonic, Hedonic-Heckman, Repeat Sales, and Repeat Sales-Case Shiller - display a similar trend of price growth over the period from February 2020 to February 2022. There is a particularly noticeable spike in price growth around February 2021. This surge in price indices coincides with several significant events in the NFT market. For instance, in March 2021, a digital artwork by the artist Beeple was sold as an NFT at Christie's auction for a staggering $69 million. This sale not only marked a historic moment for digital art but also significantly boosted the perceived value of NFTs. Furthermore, in the same month, the first-ever tweet by Twitter CEO Jack Dorsey was sold as an NFT for $2.9 million. These high-profile sales events likely contributed to the increased attention and perceived value of NFTs, which is reflected in the sharp increase in the price indices around this time. In the subsequent period, from April to June 2021, the indices continue to rise, albeit at a slower pace. This could be linked to several notable developments in the NFT market. In April 2021, Playboy announced its entry into the NFT market, and Edward Snowden's NFT sold for $5.5 million at an auction for charity. These events could have sustained the attention and interest in NFTs, contributing to the continued growth in their prices. The sharp spike in the indices around December 2021 coincides with significant investments in the NFT space. Dragonfly Capital, a cryptocurrency venture capital firm, announced a substantial investment of $225 million in DeFi and NFTs. This massive investment likely boosted market confidence and contributed to the surge in NFT prices reflected in the price indices.

Interestingly, after correcting for selection bias, both the hedonic and repeat sales indices become slightly more conservative. This suggests that selection bias have led to an overestimation of the price indices in the original hedonic and repeat sales models. Moreover, the Repeat Sales - Case Shiller model, which corrects for both heteroskedasticity and selection bias, shows slightly more volatility, particularly around December 2021. This could be attributed to the increased sensitivity of this model to changes in repeat sales data. This unique sensitivity stems from the model's design, which leverages repeat sales data to capture pure price changes, while simultaneously accounting for potential heteroskedasticity and selection bias. As a result, any substantial variations in repeat sales data are likely to be more pronounced in this model, leading to increased volatility in the derived price index.

Despite the different methodologies and underlying assumptions, it is noteworthy that the four indices exhibit a high degree of correlation, indicating they capture similar underlying pricing dynamics in the art investment market. This correlation may initially seem counter-intuitive due to the distinct focus of the hedonic and repeat sales models. The hedonic model evaluates the value of individual characteristics of a piece of art (such as artist, age, size, and material), while the repeat sales model concentrates on price changes for identical pieces of art over time. However, these findings are in line with existing studies on art market indices. For instance, Renneboog and Spaenjers (2013) found that both hedonic and repeat sales indices for art investment exhibit similar trends over time, despite differences in their construction. This could be due to the fact that both models, even with their different perspectives, are ultimately attempting to capture the same underlying market factors driving price changes in the art market.

[Figure 1 about here.]

The hedonic and repeat sales regression offer robust analytical frameworks for understanding the pricing dynamics of our digital assets. However, they are not exempt from potential issues. A significant concern in these models is the time instability of the parameters, which can lead to biased estimates of the price indices (Kuminoff et al., 2010). This instability can stem from changing market dynamics, evolving consumer preferences, or shifts in the mix of characteristics present in the traded NFTs over time.

The Laspeyres and Paasche indices, as shown in Figure 2a, provide a clear picture of the monthly price dynamics of the NFT market. we observe the Laspeyres and Paasche indices which represent the monthly price dynamics of the NFT market. Despite observable fluctuations in the market, these indices remain relatively stable and mirror each other closely throughout the studied period. This stability suggests that the cost of purchasing a specific bundle of NFT characteristics has remained consistent over time, a finding in line with the assumptions of the hedonic model (Rosen, 1974). However, a significant boom in the NFT market was observed starting from September 2021. Despite this surge, the Laspeyres and Paasche indices remained stable, prompting us to conduct a more detailed investigation into the underlying causes. To this end, we plot the changes in the β coefficients and quantities[1] over time in Figure 2b. The β coefficients, indicative of the implicit prices of NFT characteristics, show substantial variability throughout the period under consideration. This variability reflects the evolving nature of the NFT market, capturing changes in consumer preferences and market dynamics. Conversely, the quantities of the NFT characteristics, denoted by q, follow a smoother trajectory, suggesting that the mix of characteristics in the sold NFTs has remained relatively constant over time. This stability in quantities, coupled with the variability in β coefficients cause the steady nature of the Laspeyres and Paasche indices during the observed market boom. Essentially, the stability of the Fisher index, which is a geometric mean of the Laspeyres and Paasche indices, implies that changes in prices (β) are generally balanced by changes in quantities (q) (Diewert, 1976). In other words, the increase in quantities of different characteristics sold during the market boom in September 2021 could have counterbalanced the price increases.

This analysis allows us to understand the Fisher index as a relative measure, comparing prices and quantities in one period to those in another. As such, the Fisher index compares each month's index to the previous month's. If changes in prices and quantities from one month to the next are relatively small, the Fisher index will remain relatively stable.

[Figure 2 about here.]

The graphical representation of the number of trades and floor price trends in Figure 3 further underscores the dynamic nature of the NFT market. We observe a significant increase in the number of trades, or market activity, towards the end of the period, which

[1] Intended as the median of characteristics in our dataset

coincides with a rise in the floor price. This pattern suggests a potential correlation between market activity, prices, and the quantities of characteristics traded. In other words, as the volume of trade and average prices increased (evidenced by the rise in the floor price), the quantities of NFT characteristics sold also increased, effectively stabilizing the price indices.

[Figure 3 about here.]

4.2 Herding Behaviour

Following classical financial theories, we believe that consumers tend to follow the crowd, aligning their preferences and demands with popular artists, thereby creating a cascade effect. This herding behavior is manifested in the NFT market, leading to a superstar phenomenon where a handful of artists dominate the market, as proposed by Rosen (1981).

In our research, we employ a comprehensive ranking system for artists. This system works by considering a diverse range of market metrics and social factors which encompass various aspects of an artist's presence in the NFT market, allowing us to evaluate the artists' market positions and popularity. The ranking process is performed monthly to reflect the dynamic nature of the NFT market.

We factor in the ratio of an artist's sold to unsold artworks, which gives us an understanding of market demand for their work. The higher the ratio, the greater the demand for the artist's work. Similarly, we account for the average price of sold NFTs. A higher average price typically indicates a higher perceived value or desirability of an artist's work. The market capitalization for each artist, calculated on a monthly basis, is also considered. This measure, defined as the total value of their sold artworks within a given month, provides an insight into the artist's overall market presence and financial impact. We also take into account the average number of trades for each sold NFT. This measure provides an indication of the liquidity and secondary market activity of an artist's work. More trades generally suggest greater interest and engagement from collectors and investors. Additionally, we look at the number of unique buyer-seller pairs, which gives us an idea of the diversity of interactions associated with an artist's works. A greater number of unique pairs usually indicates a broader collector base and more varied market activity. Lastly, we consider the selling time of each sold NFT. This measure indicates the average time it takes for an artist's work to sell. Shorter selling times often suggest higher demand and faster market turnover.

By considering all these factors, we compute a composite score for each artist. This score, where each factor has an equal weight, serves as the basis for our monthly ranking of artists.

In order to contextualize the variables employed in our study, we start with an examination of the descriptive statistics illustrated in Table 3. The substantial variation in *Volume*—with a median value of 2811.84—exposes the disparity between artists, highlighting an uneven playing field where a select few artists garner significant market activity, while the majority struggle to gain traction. As this variable reflects the market activity for individual artists in a given month, this could potentially signal illiquidity issues, where the market is concentrated around a small number of artists, akin to the "superstar" phenomenon. Moreover, *Share*, which represents the proportion of the total monthly market capitalization attributable to an individual artist, further accentuates this fragmentation. The median value of 0 indicates that the market is disproportionately dominated by a few artists, suggesting a significant skew in market power. Lastly, the *Demand* variable, representing an artist's popularity relative to other artists, underscores the fluctuating and potentially illiquid nature of the NFT art market. With a median value of 1.04 and a wide range of values, 'Demand' demonstrates how rapidly consumer preferences can shift in the NFT art market. This suggests a market prone to sudden swings and potentially fragile liquidity, as demand can quickly pivot from one artist to another. This pattern underscores the complex interplay of individual artists' attributes and broader market dynamics, illustrating the intrinsic challenges of navigating the fragmented and highly volatile NFT art market.

[Table 3 about here.]

Table 4 presents the findings from our regression models, dividing the artists into three distinct categories based on their monthly rankings. The categories are as follows: the highest ranked artists (Model 1), the mid-tier artists (Model 2), and the lower-tier artists (Model 3). One of the significant findings of our analysis is that the variable $cum\_score\_year\_rank_{t-1}$ only impacts the choice of art for the highest ranked artists in our sample (Model 1). This suggests that for the highest ranked artists, a higher score in the previous year does not necessarily lead to higher demand in the current year, which supports Rosen (1981) explanation that art enthusiasts are driven by novelty and seek fresh artistic expressions.

For emerging and contemporary artists, such as those represented in Models 2 and 3, the quality of art does not significantly impact the choice of art, indicating that Rosen's explanation of stardom may not apply to all art market segments. This implies that superstar phenomena in these two market segments are not likely to be due talent differences but more likely due to other factors. Indeed, we find that $year\_rank_t$ is negative and statistically significant across all models, suggesting that artists with higher yearly rankings have lower demand, potentially indicating art enthusiasts' preference for niche artists or their inclination to discover new talent rather than following the crowd.

The variable $month\_rank_{t-1}$ further underscores the complexity of the art choice dynamics. It is negative for the highest and lowest ranked artists (Model 1 and 3), suggesting that a higher rank in the previous month decreases the current demand. However, the relationship is positive for mid-tier artists (Model 2), which could be attributed to a "momentum effect" where mid-ranking artists who perform well in one month continue to attract attention in the subsequent month. This result highlights the importance of the current attention given to an artist in shaping consumer choice, consistent with the findings of the empirical study.

The interaction term $year\_t * discord\_account$ is negative for the highest ranked artists (Model 1), which suggests that active engagement on Discord does not necessarily lead to increased demand for these artists' works. This finding may appear counter-intuitive, particularly considering the widely held perception that social media engagement enhances artists' visibility and amplifies their market appeal, as underscored by Bao et al., 2022. Yet, it's plausible that for top-tier artists who already enjoy significant visibility and recognition, additional exposure via Discord offers diminishing returns. These artists may have already reached a saturation point in their exposure where additional visibility on a platform like Discord doesn't noticeably bolster demand. Instead, their reputation, past performance, and the intrinsic appeal of their art might be the primary drivers of demand.

Turning our attention to the interaction term $year\_t * deployer\_artist$, we find that it's positive for the highest ranked artists (Model 1). Artists who independently deploy their work, as opposed to those associated with or acting on behalf of a company, are met with higher demand. This finding could be indicative of art enthusiasts' preference for the authenticity, creativity, and individuality that independent artists often embody. However, the dynamics shift when we look at mid-tier and lower-tier artists (Model 2 and 3). For these artists, the term is negative, implying that independent deployment might not be as beneficial. This could be due to the additional challenges and responsibilities that come with operating independently. Lesser-known artists might lack the resources, network, or reputation that companies provide, thereby facing hurdles in establishing themselves in the competitive NFT art market. Furthermore, art enthusiasts may perceive company-associated artists as more reliable or credible, especially in the volatile, rapidly-evolving NFT space.

Significantly, our findings strongly support the hypothesis of herding behavior. The herding variable, which is represented by the yearly cumulative score, shows a clear and consistent picture: the historical auction performance of artists (as represented by their rank) has a profound influence on consumer choice. This is particularly pronounced for the contemporary artist sample (Model 3), where herding has a stronger impact on consumer behavior, and the effect of past performance is even more apparent. These results substantiate the hypothesis that herding behavior is caused by informational cascades, wherein the signal conveyed through the actions of other consumers dominates private information on the quality of an artist.

[Table 4 about here.]

The clear result obtained from our regression is the conspicuous inequalities in the art NFT market: the users art choices are highly skewed towards the highest-ranked artists. This result implies a heavy concentration of market influence amongst a small number of artists.

Indeed, the presence of stardom in art markets is characterized by the fact that a select group of artists disproportionately command both attention and revenue. These 'star' artists dominate the marketplace, often overshadowing their less well-known peers. Their works are sought after by collectors and command high prices, contributing to a skewed distribution of wealth within the artist community. To visually illustrate this concentration, we turn to the 2021 Market Cap Lorenz Curve in Figure 4a. The Lorenz curve's substantial deviation from the equality line—indicative of a scenario where every artist shares an equal proportion of the market cap—provides a clear visualization of the pronounced inequality in the distribution of market capitalization among artists. The curve's steep incline towards the end of the distribution highlights that a small fraction of artists command a disproportionately large share of the market cap, reinforcing our regression findings.

Next, we use the Gini Index, a single measure derived from the Lorenz curve, to quantify the degree of inequality over time. Figure 4b shows the Gini Index for the three different samples across a series of months. The relatively high and fluctuating Gini Index values across all three samples underscore a persistent and varying level of inequality in the market cap distribution among artists. The most pronounced inequality is observed in the Top-Tier artists sample, which aligns with our regression analysis indicating the dominance of the highest-ranked artists.

[Figure 4 about here.]

Building on our prior analysis, we aim to identify if and when herding behavior is likely to occur in the entire art NFT market, especially during periods of significant market fluctuations. Understanding this dynamic is essential as it could reveal the market's resilience or susceptibility to volatility and provide insights into how investors might respond to various market scenarios.

the OLS results in Table 5 suggests that market volatility, represented by the absolute market return, has a positive influence on the Cross-Sectional Absolute Deviation (CSAD). The significant positive coefficient indicates that increased market volatility is associated with increased herding behavior. In other words, when the market experiences significant ups and downs, investors are more likely to follow the crowd, potentially leading to inflated or deflated asset prices. The lagged CSAD also has a significant positive coefficient, suggesting a persistence of herding behavior over time. This implies that if herding behavior was present in the past, it is likely to continue into the future, potentially creating self-reinforcing trends in the NFT market. Most interestingly, however, is the negative coefficient of the squared market return, which implies a non-linear relationship between market returns and herding behavior. This suggests that extreme market conditions might actually deter herding behavior. One explanation could be that even casual investors, when faced with extreme market conditions, may exhibit risk-averse behavior. The prospect of a potential market bubble or crash during periods of high returns might lead these investors to exercise more caution. Rather than following the crowd, they might rely more on their own judgment or private information, thus leading to less herding behavior. Additionally, periods of high returns could instigate profit-taking behavior amongst investors. It's possible that during such times, investors, rather than continuing to follow the upward trend, decide to sell their holdings to realize profits. This could disrupt the momentum of the trend and consequently reduce the prevalence of herding behavior (Drehmann et al., 2005).

Furthermore, extreme market conditions might exacerbate the issue of information asymmetry. Casual investors, recognizing their informational disadvantage compared to more informed market participants, may choose to refrain from following the crowd. They might prefer to adopt a wait-and-see approach or make decisions based on their own research or intuition, leading to a decrease in herding behavior (Park and Sabourian, 2011). Finally, as the NFT market matures and evolves, it's possible that investors, including casual ones, become more discerning and sophisticated. They may rely less on the behavior of others and more on their own analysis or assessment of the intrinsic value of assets. This shift in behavior could be more pronounced during extreme market conditions, resulting in less herding behavior (Economou et al., 2011).

In the Markov Switching Model (MSM), the coefficients under Regime 0 (where herding is present) and Regime 1 (where herding is absent) provide more specific insights into the dynamic nature of herding behavior in the NFT market. In Regime 0, the squared market return has a larger negative coefficient compared to the OLS model, reinforcing the idea that extreme market conditions, particularly during periods of high returns, could deter herding. In Regime 1, where herding is absent, the squared market return displays a positive, albeit insignificant, coefficient. This suggests that extreme market conditions do not necessarily deter the formation of herding behavior when it is not already present. The transition probabilities, show that herding behavior, once initiated, tends to persist. However, the market can also rapidly transition from a non-herding to a herding state.

These factors contribute to the nonlinearity in herding behavior as observed in our analysis and proposed by Fu and Wu (2021). They underscore the complexity of investor behavior, suggesting that even casual investors in the NFT market may not always follow the crowd, especially during periods of extreme market volatility.

[Table 5 about here.]

Figure 5 effectively demonstrates the temporal oscillation between regimes of herding behavior and independent decision-making within the NFT market, as postulated by our model. Specifically, the shaded areas, correlating with heightened market returns, suggest periods of pronounced herding behavior. These instances are closely aligned with the OLS regression and Markov Switching Model (MSM) findings, which reveal a positive relationship between absolute market returns and CSAD, hence indicating increased herding behavior during market volatility.

Initially, it may appear that the graphical representation and the regression model contradict each other. At first glance, the shaded areas seem to suggest a positive correlation between rising market returns and herding behavior, despite the negative coefficient for the squared market return in the regression model. Specifically, the negative coefficient for the squared market return in the regression model implies a non-linear relationship between market returns and herding behavior. In this context, herding behavior might decrease under conditions of extreme market returns (either extremely high or low). However, this doesn't mean that any increase in market returns will automatically lead to a decrease in herding. Instead, it indicates that beyond certain thresholds of market returns, investors might start showing more independent behavior, possibly due to heightened risk perception during periods of extreme volatility or the desire to cash in their gains during periods of high returns.

The shaded areas in the graph represent periods of herding, which seem to correlate with instances of significant market returns. This suggests that while extreme market returns might deter herding, periods of high, but not extreme, market returns can still induce herding behavior. The shaded periods are likely times when investors, driven by the market momentum, decide to follow the crowd, leading to an increase in herding behavior. Moreover, the shaded regions align with notable events in the NFT space that likely incited increased market activity and investor interest. For example March 2021 coincides with Beeple $69 million auction. This watershed event attracted significant media attention and public interest, leading to increased trading activity and a surge in market returns. Similarly, the shaded region in September 2021 coincides with the period when the "Loot (for Adventurers)" project gained significant traction. This project, and the numerous derivatives it inspired, led to a surge in trading activity, manifesting as increased market returns and herding behavior.

The peaks in herding behavior in January and February 2022 coincide with a significant downturn in the NFT market, largely spurred by the decrease in value of major cryptocurrencies. This period of market distress likely amplified uncertainty and risk perception among investors. In response to this heightened uncertainty, many investors may have chosen to follow the behavior of the majority, leading to an increase in herding behavior.

[Figure 5 about here.]

In essence, the dynamics of investor behavior in the NFT market are not solely dictated by market returns. While these returns undeniably play a pivotal role, they do not operate in isolation. It is a symphony of factors such as significant market events, social dynamics, and the level of hype surrounding NFTs that collectively shape investor behavior. Depending on the context, these factors can either reinforce or dampen herding tendencies among investors. Given these intricate dynamics, we employ a smoothing coefficient GAM regression to dissect these complex interactions and patterns. We select two important variables, the percentage of newcomers, intended as the proportion of investors who have traded only once, and the BCGI return. Figure 6a illustrates the former relation. In the initial stages, from zero to around 0.44 newcomer proportion, β_2 remains relatively stable and close to zero. This could suggest that the entrance of new investors into the NFT market doesn't immediately disturb the market's return volatility. It's plausible that the blend of newcomers' behaviors at this stage, encompassing both herd-like actions and independent decision-making, contributes to this stability. This phase is reminiscent of the Efficient Market Hypothesis in which markets rapidly absorb new information into asset prices, hindering any investor group's ability to consistently outperform.

As we move past the 0.44 mark, a striking transition occurs—β_2 takes a sharp downward turn. Market return volatility decreases as the NFT market continues to be populated by newcomers. This unexpected development could be attributed to a *learning effect*, where newcomers, through accumulating experience and gaining a better understanding of the market, are less inclined towards herd behavior, thus contributing to reduced market return volatility (Parker and Van Alstyne, 2005). Yet, the plot reveals another intriguing twist beyond the newcomer proportion of 0.7. Here, β_2 starts climbing again, albeit more gently. This upward shift might be a manifestation of a *saturation effect* where a market dominated by newcomers leads to collective uncertainty, reigniting herding behavior and subsequently increasing market return volatility (Cipriani and Guarino, 2008).

The relationship between β_2 and BGCI return in Figure 6b presents a linear path. The increasing relationship observed between β_2 and the Bloomberg Galaxy Crypto Index (BGCI) suggests that the herding behavior in the NFT market becomes more pronounced as the returns on the BGCI increase. However, our previous analysis found no necessarly correlation between market volatility and herding behavior in the NFT market. This is somehow surprising. The NFT market, being a part of the broader cryptocurrency market, is known for its high volatility. Yet, instead of triggering herding, this high volatility doesn't seem to impact the herding behavior in the NFT market. One possible interpretation is that NFT investors, perhaps accustomed to the inherent volatility of the crypto space, don't resort to herding behavior during volatile market conditions. Instead, they might be focusing more on the potential high returns that individual NFTs can offer. This could be attributed to the unique and idiosyncratic value of NFTs which is tied to specific digital assets, making each NFT investment a distinct proposition. This interpretation aligns with the findings of Corbet et al. (2021), who have emphasized the unique characteristics and behaviors in the cryptocurrency and NFT markets compared to traditional financial markets.

Moving on, instead of market volatility, positive market sentiment appears to be a stronger driver of herding in the NFT market. This suggests that when the overall crypto market (as represented by the BGCI) is doing well, it creates a positive sentiment that encourages investors to follow the crowd, thus increasing herding behavior. This is similar to the findings of Demirer and Kutan (2006) in the context of emerging markets, where positive market sentiment, rather than volatility, led to increased herding. In the context of the NFT market, this positive sentiment could be influenced by a variety of factors. For example, the profitable NFT auctions of 2021 could have created a positive buzz, or massive blockchain adoption by the industry could enhance overall market sentiment, driving investors towards herding behavior.

[Figure 6 about here.]

5 Conclusions

In our study, we analyze the determinants of value and the herding behavior in the art NFT market. We use, to our knowledge, the most extensive data set ever gathered in this field. A significant component of our research design involves forming the first-ever ranking system for NFT artists, which presents a milestone in the NFT research domain.

Our key findings contribute significantly to the current understanding of NFT pricing dynamics. We identified a positive correlation between the number of trades and the floor price of NFTs with their average price, providing a crucial insight into the pricing factors within the NFT market. Additionally, we have challenged the general belief about the impact of social media, demonstrating a negative correlation between social media activity and NFT prices, due to possible information oversaturation. Furthermore, we have revealed a strong herding behavior within the art NFT market. This behavior is reflected in the unequal distribution of market activities, with a small group of top-tier artists dominating the landscape. Despite observed herding behaviors, our findings also suggest that during periods of extreme market volatility, investors may revert to independent decision-making, highlighting the intricate interplay between market conditions and investor behavior. We have further contributed to the existing literature by uncovering the intricate oscillation between collective and individual decision-making within the NFT market. This novel finding suggests that significant market events and social trends have the power to either reinforce or undermine investor herding behavior, thereby enriching our understanding of NFT market dynamics. Lastly, we have identified that the positive sentiment in the broader cryptocurrency market, as reflected by rising crypto index returns, intensifies the herding behavior in the NFT market.

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

Aaker, D. A. (1996). Measuring Brand Equity Across Products and Markets. *California Management Review*, *38*(3), 102–120.

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

Amihud, Y., & Mendelson, H. (1986). Asset Pricing and the Bid-Ask Spread. *Journal of Financial Economics*, *17*(2), 223–249.

Ang, A., Liu, J., & Schwarz, K. (2020). Using Stocks or Portfolios in Tests of Factor Models. *Journal of Financial and Quantitative Analysis*, *55*(3), 709–750.

Ante, L. (2022). The Non-Fungible Token (NFT) Market and Its Relationship with Bitcoin and Ethereum. *FinTech*, *1*(3), 216–224.

Azarmi, T. F., & Menny, P. R. P. (2013). Star Artists and Herding in Fine Arts' Market: Theory and Empirical Evidence. *SSRN Journal, ID 2232307.*

Bailey, M. J., Muth, R. F., & Nourse, H. O. (1963). A Regression Method for Real Estate Price Index Construction. *Journal of the American Statistical Association*, *58*(304), 933–942.

Banerjee, A. V. (1992). A Simple Model of Herd Behavior. *The Quarterly Journal of Economics*, *107*(3), 797–817.

Bao, T., Ma, M., & Wen, Y. (2022). Herding in the Non-fungible Token (NFT) Market. *Journal of Behavioral and Experimental Finance, forthcoming.*

Basel, A. (2022). Collectors Are Spending More, Not Less, on NFT Art, According to the Art Market Report.

Bekaert, G., Harvey, C. R., & Lundblad, C. (2007). Liquidity and Expected Returns: Lessons from Emerging Markets. *The Review of Financial Studies*, *20*(6), 1783–1831.

Bianchi, D., & Babiak, M. (2022). On the Performance of Cryptocurrency Funds. *Journal of Banking & Finance*, *138*, 106467.

Borri, N., Liu, Y., & Tsyvinski, A. (2023). The Economics of Non-Fungible Tokens. *SSRN Journal ID 4052045*.

Campbell, J. Y., & Shiller, R. J. (1988). The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. *The Review of Financial Studies*, *1*(3), 195–228.

Caporale, G. M., Kang, W.-Y., Spagnolo, F., & Spagnolo, N. (2021). Cyber-Attacks, Spillovers and Contagion in the Cryptocurrency Markets. *Journal of International Financial Markets, Institutions and Money*, *74*, 101298.

Case, K. E., & Shiller, R. J. (1989). The Efficiency of the Market for Single-Family Homes. *The American Economic Review*, *79*(1), 125–137.

Chainalysis. (2022). The Chainalysis 2021 NFT Market Report.

Chanel, O., Chanel, O., Gérard-Varet, L.-A., & Ginsburgh, V. (1996). The Relevance of Hedonic Price Indices: The Case of Paintings. *Journal of Cultural Economics*, *20*(1), 1–24.

Chang, E. C., Cheng, J. W., & Khorana, A. (2000). An Examination of Herd Behavior in Equity Markets: An International Perspective. *Journal of Banking & Finance*, *24*(10), 1651–1679.

Cheah, E.-T., & Fry, J. (2015). Speculative Bubbles in Bitcoin Markets? An Empirical Investigation in the Fundamental Value of Bitcoin. *Economics Letters*, *130*, 32–36.

Cipriani, M., & Guarino, A. (2008). Herd Behavior and Contagion in Financial Markets. *The B.E. Journal of Theoretical Economics*, *8*(1), 1–56.

Collins. (2021). Get Your Crypto at the Ready: NFTs Are Big in 2021. *Collins Dictionary Language Blog*.

Costola, M., Iacopini, M., & Santagiustina, C. R. M. A. (2021). On the "Mementum" of Meme Stocks. *Economics Letters*, *207*, 110021.

Dellarocas, C. (2003). The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. *Management Science*, *49*(10), 1407–1424.

Diebold, F. X., & Rudebusch, G. D. (1999). *Business Cycles: Durations, Dynamics, and Forecasting*.

Diewert, W. E. (1976). Exact and Superlative Index Numbers. *Journal of Econometrics*, *4*(2), 115–145.

Dowling, M. (2022a). Fertile LAND: Pricing Non-Fungible Tokens. *Finance Research Letters*, *44*, 102096.

Dowling, M. (2022b). Is Non-Fungible Token Pricing Driven by Cryptocurrencies? *Finance Research Letters*, *44*, 102097.

Drehmann, M., Oechssler, J., & Roider, A. (2005). Herding and Contrarian Behavior in Financial Markets: An Internet Experiment. *The American Economic Review*, *95*(5), 1403–1426.

Dyhrberg, A. H. (2016). Bitcoin, Gold and the Dollar – A GARCH Volatility Analysis. *Finance Research Letters*, *16*, 85–92.

Economou, F., Kostakis, A., & Philippas, N. (2011). Cross-Country Effects in Herding Behaviour: Evidence from Four South European Markets. *Journal of International Financial Markets, Institutions and Money*, *21*(3), 443–460.

Fama, E. F., & French, K. R. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, *33*(1), 3–56.

Financial Times. (2021a). How Digital Art Dominated 2021.

Financial Times. (2021b). How NFTs Became a $40bn Market in 2021.

Fu, J., & Wu, L. (2021). Regime-Switching Herd Behavior: Novel Evidence from the Chinese A-share Market. *Finance Research Letters*, *39*, 101652.

Glaser, F., Zimmermann, K., Haferkorn, M., Weber, M. C., & Siering, M. (2014). Bitcoin - Asset or Currency? Revealing Users' Hidden Intentions. *SSRN Journal, ID 2425247*.

Goetzmann, W. (1992). The Accuracy of Real Estate Indices: Repeat Sale Estimators. *The Journal of Real Estate Finance and Economics*, *5*(1), 5–53.

Hemenway Falk, B., Tsoukalas, G., & Zhang, N. (2022). Economics of NFTs: The Value of Creator Royalties. *SSRN Journal, ID 4284776*.

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*, *48*, 103007.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–291.

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*.

Kuminoff, N., Parmeter, C., & Pope, J. (2010). Which Hedonic Models Can We Trust to Recover the Marginal Willingness to Pay for Environmental Amenities? *Journal of Environmental Economics and Management*, *60*(3), 145–160.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, *74*(2), 132–157.

Li, S., & Chen, Y. (2023). How Non-Fungible Tokens Empower Business Model Innovation. *Business Horizons*, *66*(4), 543–544.

Madhavan, A. (2000). Market Microstructure: A Survey. *Journal of Financial Markets*, *3*(3), 205–258.

Mandel, B. R. (2009). Art as an Investment and Conspicuous Consumption Good. *American Economic Review*, *99*(4), 1653–1663.

Mazur, M. (2021). Non-Fungible Tokens (NFT). The Analysis of Risk and Return. *SSRN Journal, ID 3953535*.

Mei, J., & Moses, M. (2002). Art as an Investment and the Underperformance of Masterpieces. *The American Economic Review*, *92*(5), 1656–1668.

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.

Milgrom, P. R., & Weber, R. J. (1982). A Theory of Auctions and Competitive Bidding. *Econometrica*, *50*(5), 1089–1122.

Muehlenbachs, L., Spiller, E., & Timmins, C. (2015). The Housing Market Impacts of Shale Gas Development. *American Economic Review*, *105*(12), 3633–3659.

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.

O'brien, R. M. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity*, *41*(5), 673–690.

Ofek, E., & Richardson, M. (2003). DotCom Mania: The Rise and Fall of Internet Stock Prices. *The Journal of Finance*, *58*(3), 1113–1137.

Park, A., & Sabourian, H. (2011). Herding and Contrarian Behavior in Financial Markets. *Econometrica*, *79*(4), 973–1026.

Parker, G. G., & Van Alstyne, M. W. (2005). Two-Sided Network Effects: A Theory of Information Product Design. *Management Science*, *51*(10), 1494–1504.

Renneboog, L., & Spaenjers, C. (2013). Buying Beauty: On Prices and Returns in the Art Market. *Management Science*, *59*(1), 36–53.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, *82*(1), 34–55.

Rosen, S. (1981). The Economics of Superstars. *The American Economic Review*, *71*(5), 845–858.

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

Schwert, G. W. (1989). Why Does Stock Market Volatility Change Over Time? *The Journal of Finance*, *44*(5), 1115–1153.

Shiller, R. (1995). Conversation, Information, and Herd Behavior. *American Economic Review*, *85*(2), 181–85.

Sotheby's. (2021). NFTs: Redefining Digital Ownership and Scarcity.

Stiglitz, J. E. (1989). Markets, Market Failures, and Development. *The American Economic Review*, *79*(2), 197–203.

Triplett, J. (2004). Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products. *OECD Working Papers Series*, *9*.

Urquhart, A. (2016). The Inefficiency of Bitcoin. *Economics Letters*, *148*, 80–82.

Vayanos, D., & Wang, J. (2012). Liquidity and Asset Returns Under Asymmetric Information and Imperfect Competition. *The Review of Financial Studies*, *25*(5), 1339–1365.

Wang, Y., Horky, F., Baals, L. J., Lucey, B. M., & Vigne, S. A. (2022). Bubbles All the Way Down? Detecting and Date-Stamping Bubble Behaviours in NFT and DeFi Markets. *Journal of Chinese Economic and Business Studies*, *20*(4), 415–436.

Whitaker, A., & Kräussl, R. (2020). Fractional Equity, Blockchain, and the Future of Creative Work. *Management Science*, *66*(10), 4594–4611.

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.

Yousaf, I., & Yarovaya, L. (2022). Herding Behavior in Conventional Cryptocurrency Market, Non-Fungible Tokens, and DeFi Assets. *Finance Research Letters*, *50*, 103299.

Table 1. Descriptive Statistics Art NFTs: The table offers a detailed statistical overview of art NFTs, specifically focusing on their continuous and discrete variables. **Panel A** focuses on the continuous variables, providing a comprehensive statistical summary of the data extracted. This includes the median, standard deviation (denoted by σ), minimum, maximum, Jarque-Bera (JB) Statistic for normality testing, and the KPSS Test for stationarity testing across both the Aggregate and Transaction datasets. The Aggregate dataset encompasses augmented data per NFT, including data from Discord and Twitter, as well as color and image complexity data. In contrast, the Transaction dataset contains detailed information pertaining to every sale of the extracted art NFTs. The JB Statistic and KPSS Test provide insights into the distribution characteristics, with statistical significance indicated at various levels. **Panel B**, on the other hand, details the discrete variables in the both datasets, listing the unique values and the top frequent value with its corresponding frequency. The information compiled in this table stems from a rigorous data extraction process, leveraging on-chain and off-chain sources as described in Section A.1, providing a rich and diversified analysis of the NFTs, encompassing both their visual characteristics and market behavior.

Panel A: Continuous Variables

Variable	Median	σ	Min.	Max.	JB Stat.	KPSS Test
			Aggregate DataSet			
BGCI_return	0	0.03	-0.26	0.22	8.10e+04***	0.94***
black	0.04	0.27	0	1	1.15e+05***	5.01***
blue	0.04	0.23	0	1	2.24e+05***	14.28***
num_colors	7	2.43	1	9	2.58e+04***	3.15***
norm_shannon_entropy	0.29	0.12	0	0.53	1.69e+04***	5.38***
floor_price	613.22	2.38e+04	0.01	6.25e+06	1.14e+13***	2.47***
following_count	780	1268.19	0	1.76e+04	1.13e+06***	31.17***
gray	0.01	0.2	0	1	6.23e+05***	10.74***
green	0	0.17	0	1	1.44e+06***	7.63***
last_price	725.02	2.82e+04	0.01	6.25e+06	3.61e+12***	4.29***
listed_count	27	1322.58	0	8.76e+04	5.86e+09***	21.68***
quote_count_month	64	394.54	0	4.85e+04	4.62e+11***	11.07***
orange	0.03	0.2	0	1	4.22e+05***	17.01***
purple	0	0.13	0	1	6.34e+06***	10.13***
red	0.05	0.21	0	1	3.96e+05***	4.16***
timediff	48.97	242.55	0.01	1307.19	5.50e+05***	16.89***
num_trades	1	9.12	1	2372	9.03e+12***	7.08***
white	0.01	0.25	0	1	2.10e+05***	2.66***
yellow	0	0.1	0	1	1.23e+07***	1.64***
			Transaction DataSet			
average_price	748.48	2.21e+04	0.01	6.25e+06	1.44e+13***	0.72***
floor_price	459.44	1.98e+04	0	6.25e+06	3.29e+13***	0.62***
last_price	744.23	2.82e+04	0	6.25e+06	3.16e+12***	0.75***
max_price	976.74	3.15e+04	0.01	6.25e+06	1.31e+12***	1.13***
num_trades	1	285.24	1	2372	2.30e+07***	0.61***
usd_amount	679.07	2.39e+04	0.01	6.25e+06	9.16e+12***	0.71***

Panel B: Discrete Variables

Variable	Unique Values	Top Frequent Value	Frequency of Top Value
	Aggregate DataSet		
collection_name	512	Foundation (FND)	21.91
deployer_creator_generalities	8	Company	39.96
deployer_creator_name	291	Pak	3.49
discord_account	2	1	61.86
discord_handle	237	SuperRare	7.65
marketplace_collection	2	0	70.09
nft_type	2	erc721	97.76
platform_of_last_sale	6	OpenSea	69.54
twitter_account	2	1	97.58
twitter_handle	419	foundation	21.91
verified	2	0	64.17
	Transaction DataSet		
buyer	95399	0x8888888888e9997e...	0.62
collection_name	513	Foundation (FND)	15.29
nft	240614	0x73da73ef3a6982109c4.../8	0.67
platform	6	OpenSea	74.93
seller	73207	0x8c9f364bf7a56ed058...	1.08

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2. Hedonic and Heckman Two-Stage Model Estimations. This table presents the estimation results of hedonic models (Models 1, 2, and 3) and Heckman two-stage models (Models 1', 2', and 3') applied to the average price of Art NFTs collected from June 1, 2018, to February 10, 2022. The dependent variable is the logarithm of the average price plus one. The models progressively include additional controls: Model 1 and 1' do not include social network controls or market fixed effects, Model 2 and 2' introduce social network controls, and Model 3 and 3' further include market fixed effects. Month fixed effects are included in all models. Each entry in the table is the coefficient estimate from the respective model, with the standard error reported in parentheses. Asterisks denote significance levels: * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. The number of observations for all models is 191,201. The adjusted R^2, residual standard error, and F-statistics are reported at the bottom of the table. The *mills_ratio* row represents the Inverse Mills Ratio, which is included only in the Heckman models (1', 2', and 3') as a result of the first stage selection equation.

			Dependent Variable: ln(*average_price* + 1)			
	(1)	(1')	(2)	(2')	(3)	(3')
const	-0.513*** (0.007)	-0.491*** (0.007)	-0.461*** (0.009)	-0.454*** (0.009)	-0.665*** (0.013)	-0.661*** (0.013)
black	-0.002 (0.004)	0.004 (0.004)	-0.005 (0.004)	-0.002 (0.004)	-0.007* (0.004)	-0.006 (0.004)
white	0.013*** (0.004)	0.015*** (0.004)	0.009** (0.004)	0.010** (0.004)	0.021*** (0.004)	0.022*** (0.004)
norm_shannon_entropy	0.005 (0.011)	0.015 (0.011)	0.036*** (0.011)	0.038*** (0.011)	0.068*** (0.011)	0.068*** (0.011)
num_of_colors	0.001 (0.000)	0.000 (0.000)	0.001 (0.000)	0.000 (0.000)	-0.001** (0.000)	-0.001** (0.000)
BGCI_return	0.112*** (0.027)	0.696*** (0.059)	0.100*** (0.027)	0.357*** (0.060)	0.099*** (0.027)	0.200*** (0.061)
log(*number_of_trades*)	1.076*** (0.004)	1.077*** (0.004)	1.080*** (0.004)	1.081*** (0.004)	1.108*** (0.004)	1.108*** (0.004)
log(*floor_price*)	0.929*** (0.001)	0.928*** (0.001)	0.924*** (0.001)	0.924*** (0.001)	0.918*** (0.001)	0.918*** (0.001)
log(*following_count*)	-0.005*** (0.001)	-0.005*** (0.001)	-0.010*** (0.001)	-0.009*** (0.001)	-0.006*** (0.001)	-0.006*** (0.001)
log(*listed_count*)	0.024*** (0.000)	0.024*** (0.000)	0.026*** (0.000)	0.026*** (0.000)	0.019*** (0.001)	0.019*** (0.001)
log(*quote_count_month*)	0.002** (0.001)	0.002** (0.001)	-0.016*** (0.001)	-0.015*** (0.001)	-0.011*** (0.001)	-0.011*** (0.001)
log(*retweet_count_month*)	0.007*** (0.001)	0.007*** (0.001)	0.015*** (0.001)	0.014*** (0.001)	0.012*** (0.001)	0.012*** (0.001)
log(*like_count_month*)	0.002*** (0.000)	0.002*** (0.000)	0.014*** (0.000)	0.014*** (0.000)	0.008*** (0.000)	0.008*** (0.000)
log(*reply_count_month*)	0.004*** (0.001)	0.004*** (0.001)	0.015*** (0.001)	0.014*** (0.001)	0.011*** (0.001)	0.011*** (0.001)
log(*post_count_month*)	-0.017*** (0.000)	-0.017*** (0.000)	-0.035*** (0.000)	-0.035*** (0.000)	-0.021*** (0.000)	-0.021*** (0.000)
log(*number_of_pictures*)	0.001*** (0.000)	0.001*** (0.000)	0.004*** (0.000)	0.004*** (0.000)	0.002*** (0.000)	0.002*** (0.000)
mills_ratio		-6.146*** (0.547)		-2.708*** (0.563)		-1.065*** (0.570)
Social Network Controls	✗		✓		✓	
Market FE	✗		✗		✓	
Month FE	✓		✓		✓	
Observations			191,201			
Adjusted R^2	0.692	0.731	0.849	0.861	0.930	0.932
Residual Std. Error	0.606	0.406	0.575	0.405	0.402	0.402
F Statistic	69435.176*** (df = 36)	67606.269*** (df = 37)	64423.102*** (df = 39)	62820.379*** (df = 40)	48058.466*** (df = 53)	47169.173*** (df = 54)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3. Herding Behaviour Descriptive Statistics. This table provides descriptive statistics for the continuous and discrete variables used in the study of herd bias. The continuous variables are split into two datasets: the Artists dataset and the Transactions dataset. Each variable's median value is displayed, along with its standard deviation (σ), minimum and maximum values. The Jarque-Bera statistic (JB Stat.), which tests for normality in the data distribution, is also provided. A triple asterisk denotes a significance level of $p < 0.01$. The KPSS test results for stationarity are also shown, with a single asterisk denoting $p < 0.1$. For instance, in the Artist DataSet, *sold_ratio*, which shows that the average artist sells only 2% of their work, reveals a great disparity as some artists manage to sell all their work. The wide-ranging $floor\_price$ highlights the diversity in the pricing strategy across artists, with the average price being 380.06, while some NFTs are priced as high as 1.23e+06. Trade activity, represented by the $number\_of\_trades$, is generally limited but can peak dramatically in certain cases, reaching up to 1799 trades. The $buyer\_seller\_pair$ variable confirms this trend, as despite the median indicating just 7 unique pairs, some artists attract thousands of unique pairs. NFT selling times, captured by $timediff$, typically do not exceed three months, but some artworks may require over three years to find a buyer. The $Volume$ variable, a measure of an artist's monthly market activity, further supports this imbalance: while the median market activity is 2811.84, the most popular artists can stir activity levels up to 1.63e+07. *Share* suggests that the NFT market can be dominated by a few artists, with some accounting for up to 82% of the total monthly market capitalization, despite the median share being zero. Finally, *Demand* variable summarize the shifting nature of consumer preferences in the NFT market, with a wide range from 0 to 5461.62, despite the average artist having a demand score close to 1. In the Transactions dataset, discrete variables such as *discord_account* and *deployer_artist* provide additional insights. Most artists appear to have a Discord account, while independent artists constitute the majority.

Panel A: Continuous Variables

Variable	Median	σ	Min.	Max.	JB Stat.	KPSS Test
			Artists DataSet			
sold_ratio	0.02	0.16	0	1	2.50e+04***	0.16*
floor_price	380.06	3.36e+04	1.02	1.23e+06	7.39e+07***	0.15*
number_of_trades	2.41	46.01	2	1799	9.68e+07***	0.28*
buyer_seller_pair	7	345.81	0	7256	3.96e+06***	0.08*
timediff	88.69	186.02	1.04	1231.58	8150.97***	0.31*
score	776.53	1.50e+05	62.28	2.72e+06	2.16e+06***	0.28*
cum_score	1.47e+07	1.59e+07	484.87	5.93e+07	254.1***	7.94***
cum_score_year_rank	447	471.36	1	1589	230.69***	8***
Volume	2811.84	8.90e+05	1.02	1.63e+07	2.25e+06***	0.28*
Share	0	0.05	0	0.82	1.19e+06***	0.11*
Demand	1.04	171.49	0	5461.62	2.24e+07***	0.41***
			Transactions DataSet			
market_return	0	0	-0.01	0.06	5.10e+04***	0.42***
CSAD	0.01	0.01	0	0.11	4.08e+04***	0.52***
newcomer_proportion	0.33	0.12	0	1	65.73***	1.11***

Panel B: Discrete Variables

Variable	Unique Values	Top Frequent Value	Frequency of Top Value
		Transactions DataSet	
discord_account	2	Yes	55.52
deployer_artist	2	Individual	87.02

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4. Multinomial Logit Model Capturing Herd Bias. The table presents the findings from a comprehensive regression analysis based on a multinomial logit model, investigating the determinants of demand for artworks in the NFT market. We divide artists into three categories according to their monthly rankings: the highest ranked artists (Model 1), mid-tier artists (Model 2), and lower-tier artists (Model 3). The dependent variable in each model is the logarithm of demand for artworks by each artist category. A diverse array of explanatory variables is included, reflecting historical performance metrics, artist rankings, and social media engagement. For instance, the cumulative score yearly ranking lagged by one period ($cum\_score\_year\_rank_{t-1}$) has a negative coefficient in Model 1, suggesting that a higher score in the previous year does not necessarily translate into higher demand in the current year for top-tier artists. The interaction term between the year and the presence of a Discord account ($year\_t * discord\_account$) also has a negative coefficient in Model 1, suggesting that active engagement on Discord does not necessarily lead to increased demand for these top-tier artists' works. On the other hand, the interaction term between the year and the deployer artist status ($year\_t * deployer\_artist$) shows a positive coefficient for Model 1, implying that top-tier artists who independently deploy their work enjoy higher demand. Each entry in the table delineates the coefficient estimates, with standard errors provided in parentheses. Artist and month fixed effects are included to control for unobserved time-invariant artist-specific characteristics and common temporal shocks. The total number of observations, the adjusted R^2, and the F-statistic are reported for each artist category. Significance levels are denoted as follows: $*p < 0.1, **p < 0.05, ***p < 0.01$.

	Dependent Variable: log($demand_{it}$)		
	(1)	(2)	(3)
$const$	5.120*** (0.092)	3.908*** (0.113)	4.491*** (0.379)
$cum\_score\_year\_rank_{t-1}$	-0.002* (0.001)	-0.04 (0.001)	0.001** (0.002)
$year\_rank_t$	-0.080*** (0.006)	-0.064*** (0.002)	-0.075*** (0.005)
$month\_rank_{t-1}$	-0.017*** (0.011)	0.003* (0.006)	-0.016** (0.012)
$year_t$ *$discord\_account$	-0.003** (0.003)	0.01 (0.001)	0.004* (0.002)
$year_t$ *$deployer\_artist$	0.007*** (0.005)	-0.002* (0.002)	-0.005* (0.003)
Artist FE	✓	✓	✓
Month FE	✓	✓	✓
Observations	786	793	802
Adjusted R^2	0.791	0.713	0.505
F Statistic (robust)	259.22*** (df = 523)	348.57*** (df = 459)	104.69*** (df = 513)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 5. OLS and Markov Switching Models for Herd Bias. The table presents the coefficients of the Ordinary Least Squares (OLS) regression and the Markov Switching Model (MSM) to investigate the presence and dynamics of herding behavior in the art NFT market. The dependent variable is the Cross-Sectional Absolute Deviation (CSAD), a measure of herding. The first column shows the OLS regression results, illustrating the overall dynamics of herding behavior in the market. The absolute market return ($|R_{m,t}|$) and lagged CSAD ($CSAD_{t-1}$) both have positive coefficients, indicating that market volatility and past herding behavior increase the tendency for herding. Interestingly, the squared market return ($R^2_{m,t}$) has a negative coefficient, suggesting that extreme market conditions may deter herding behavior. The next two columns display the coefficients under two different regimes in the MSM: Regime 0, which signifies the presence of herding, and Regime 1, which signifies the absence of herding. In Regime 0, the negative coefficient of the squared market return is larger than in the OLS model, reinforcing the nonlinear relationship between market returns and herding. In Regime 1, the squared market return displays a positive, though insignificant, coefficient, suggesting that extreme market conditions do not necessarily deter the formation of herding behavior when it is not already present. The last column shows the transition probabilities ($p[0 \rightarrow 0]$ and $p[1 \rightarrow 0]$) in the MSM, providing insights into the persistence and transition dynamics of herding behavior. The table contains 4040 observations. The model's goodness of fit is represented by the log-likelihood, Bayesian Information Criterion (BIC), and Akaike Information Criterion (AIC). Significance levels are denoted as follows: $***p < 0.01, **p < 0.05, *p < 0.1$

	Dependent Variable: $CSAD_t$			
	OLS	Markov Switching Model		Non-Switching
		Regime 0	Regime 1	Parameters
const	0.0018*** (0.000)	0.0015*** (0.000)	0.0110*** (0.001)	
$\lvert R_{m,t}\rvert$	1.9494*** (0.094)	2.0903*** (0.092)	3.7242*** (0.732)	
$R^2_{m,t}$	**-1.3950*** ** (0.626)	**-5.7345**** (4.151)	**2.3418** (3.114)	
$CSAD_{t-1}$	0.1683*** (0.017)	0.1215*** (0.021)	0.1111*** (0.027)	
σ^2		0.067*** (0.002)	0.004** (0.010)	
p[0→0]				0.9631*** (0.012)
p[1→0]				0.7139*** (0.133)
Observations	4040			
Log Likelihood	1786.3	1839.752		
BIC	-3543	-3601.485		
AIC	-3563	-3653.504		
Note:				*p<0.1; **p<0.05; ***p<0.01

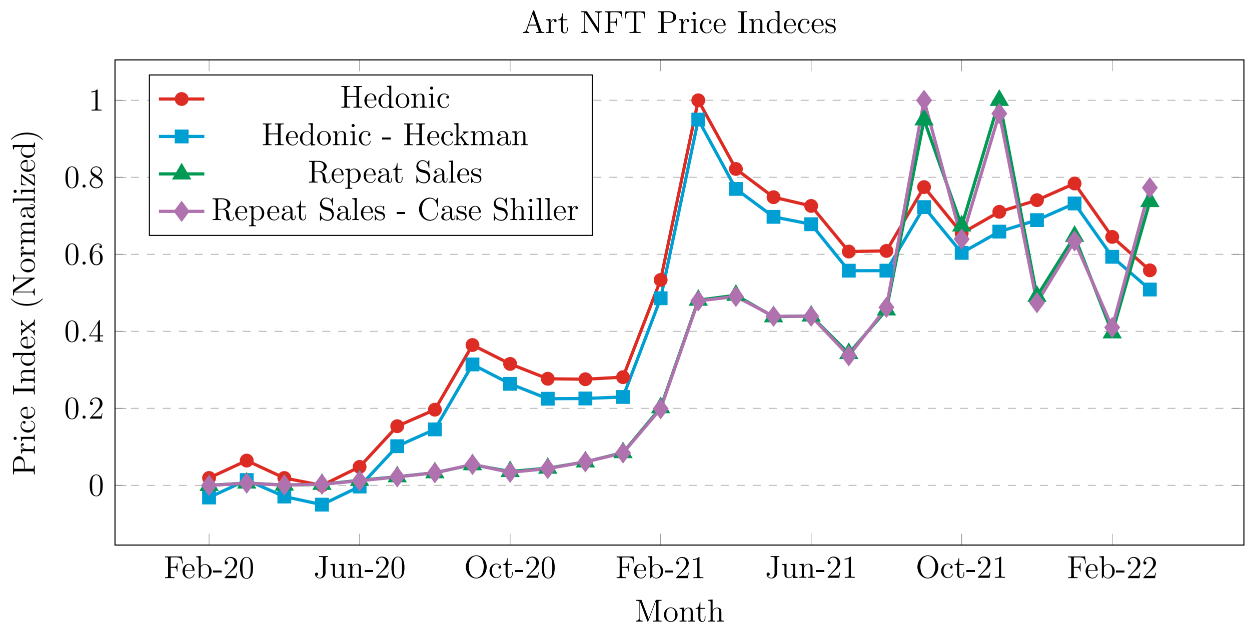
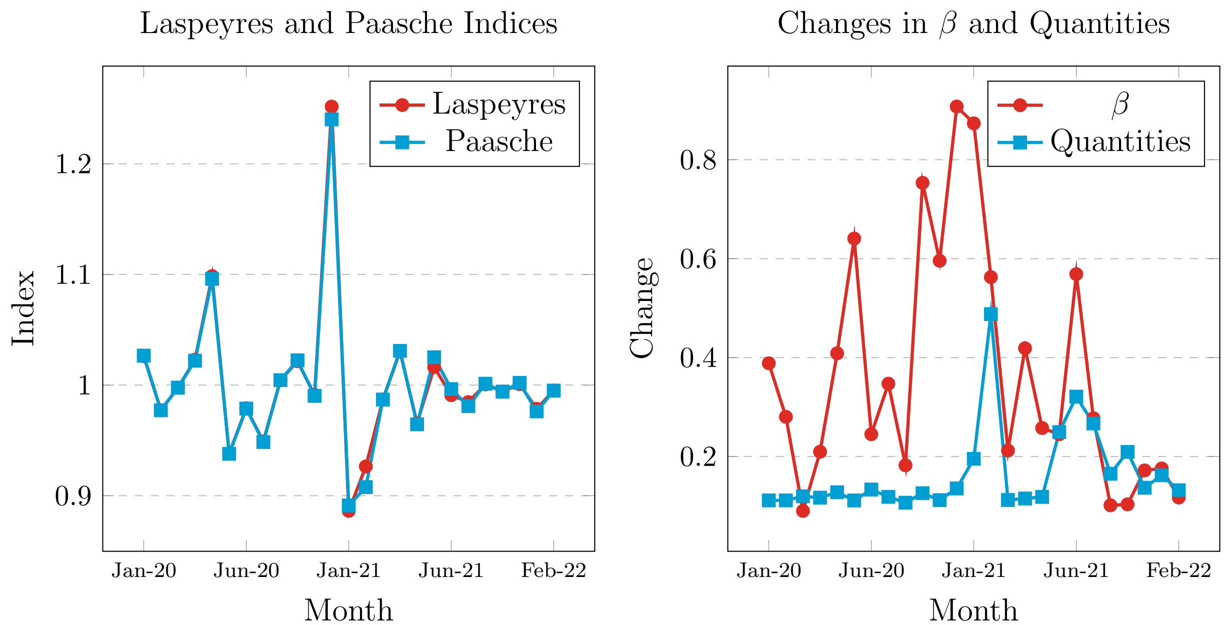


Figure 1. Art NFT Price Indeces. The figure presents the four different art NFT price indices - the Hedonic, Hedonic-Heckman, Repeat Sales, and Repeat Sales-Case Shiller. These indices, derived from distinctive methodologies and underlying assumptions, trace the evolution of art NFT prices from February 2020 to February 2022. Notably, all indices display a correlated pattern of price growth over this period, despite the nuanced differences in their construction. The Hedonic model, for instance, evaluates the prices of art NFTs based on their observable characteristics, while the Repeat Sales model focuses on price changes for identical art NFTs traded over time. The Hedonic-Heckman model augments the Hedonic model by correcting for potential selection bias. The Repeat Sales-Case Shiller model, on the other hand, extends the Repeat Sales model by accounting for both heteroskedasticity and selection bias. A significant spike in price growth is evident around February 2021, which aligns with several high-profile events in the NFT market, such as the sale of Beeple's digital artwork for $69 million at Christie's auction and the sale of Twitter CEO Jack Dorsey's first-ever tweet as an NFT for $2.9 million. These events significantly bolstered the perceived value and attention toward NFTs, leading to a steep rise in price indices. In contrast, after correcting for selection bias, the Hedonic-Heckman and Repeat Sales-Case Shiller indices become slightly more conservative, suggesting that selection bias might have led to an overestimation of the price indices in the original hedonic and repeat sales models. This highlights the importance of accounting for selection bias to ensure accurate estimation of asset prices. Furthermore, the Repeat Sales-Case Shiller model exhibits slightly more volatility, particularly around December 2021, which could be attributed to its increased sensitivity to changes in repeat sales data.



(a) Laspeyres and Paasche Indices **(b) Changes in β and Quantities**

Figure 2. Fisher Index. Figure **(a)** presents the Laspeyres and Paasche indices (components of the overall Fisher price index) over the examined period from January 2020 to February 2022. These indices reflect the monthly price dynamics of the NFT market. Despite observable fluctuations in the market, particularly a significant surge in September 2021, these indices remain relatively stable. This stability suggests a consistent cost for a specific bundle of NFT characteristics, underpinning the assumptions of the hedonic model. Figure **(b)** illustrates the changes in the β coefficients and quantities over the same time period. The β coefficients, which represent the implicit prices of NFT characteristics, show substantial variability, capturing the evolving nature of the NFT market and changes in consumer preferences. Conversely, the quantities, denoted by q, remain relatively constant, indicating a stable mix of NFT characteristics sold over time. This contrast between the variability of β coefficients and the stability of q suggests that the changes in prices are generally balanced by changes in quantities. This balance further highlights the stability of the Laspeyres and Paasche indices, and consequently the Fisher index, even during the market boom in September 2021.

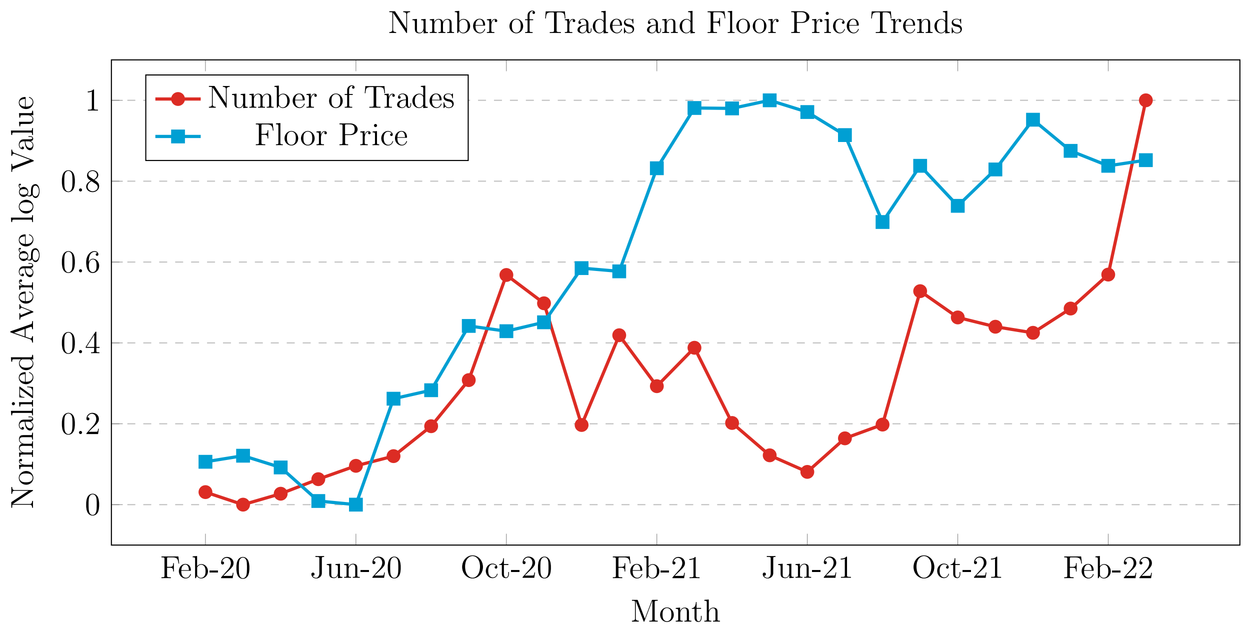
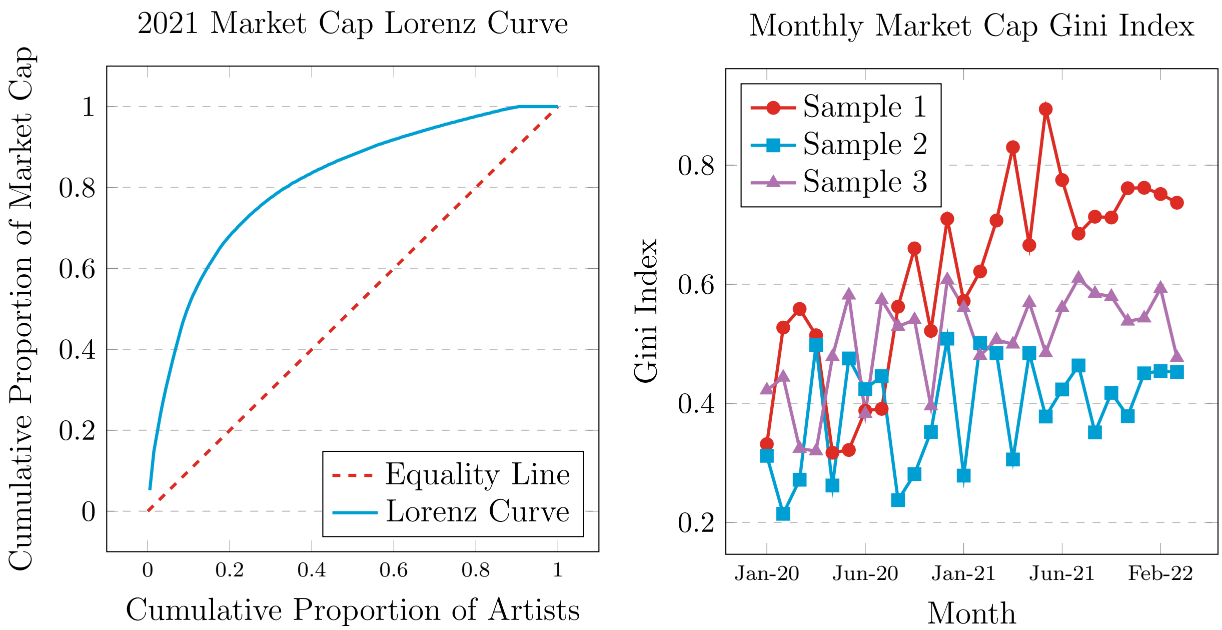


Figure 3. Number of Trades and Floor Price Trends. The graph presents the evolution of the number of trades and floor price trends in the art NFT market from February 2020 to February 2022. The number of trades (depicted in red) and floor prices (depicted in blue) are represented on a normalized logarithmic scale. An observable upswing in both metrics is evident from around October 2021, suggesting a significant market boom during this period. This upward trend likely reflects evolving consumer preferences, technological advancements, increased public awareness, and broader market acceptance of NFTs as digital assets. Importantly, as trading volume and average prices increased, so did the quantities of NFT characteristics sold, effectively balancing the price indices.



(a) Lorenz Curve **(b) Gini Index**

Figure 4. Lorenz Curve and Gini Index for NFT Artists. In the analysis of the 2021 Market Cap Lorenz Curve, a striking deviation from the line of equality is evident. This line of equality serves as a baseline, symbolizing an idealistic scenario where each artist enjoys an equal proportion of market capitalization—an equitable distribution of wealth and influence. The Lorenz curve, however, tells a different story, one marked by stark disparities and profound inequality. The Lorenz curve's x-axis measures the cumulative proportion of artists, a metric that ranges from the least to the most successful. The y-axis, on the other hand, quantifies the cumulative proportion of the market cap, thereby reflecting the accumulated wealth and influence within the artist community. The substantial bowing of the Lorenz curve away from the line of equality brings to light the pronounced inequality in the distribution of market capitalization among artists. The curve takes a steep ascent towards the end of the distribution, suggesting that a minor fraction of artists wield a disproportionately large share of the market cap. This phenomenon, termed market concentration, is brought into stark relief by the graphical representation, corroborating the findings from our comprehensive regression analysis. We turn next to the Monthly Market Cap Gini Index to further investigate the degree and temporal dynamics of this inequality. The Gini Index, a derivative measure from the Lorenz curve, encapsulates the inequality within a single, powerful metric. Plotted over time, it offers a chronological snapshot of inequality, tracing its ebb and flow through different periods. Three different samples were analyzed across a series of months (Top, Mid and Low Tier Artists), each revealing its own tale of inequality. However, the common thread across all three is the high Gini Index values that fluctuate over time, underscoring a persistent and variable level of inequality in the market cap distribution among artists. Among these, Sample 1 exhibits the most pronounced inequality, reflecting the dominance of the highest-ranked artists—a finding that echoes the conclusion drawn from our regression analysis.

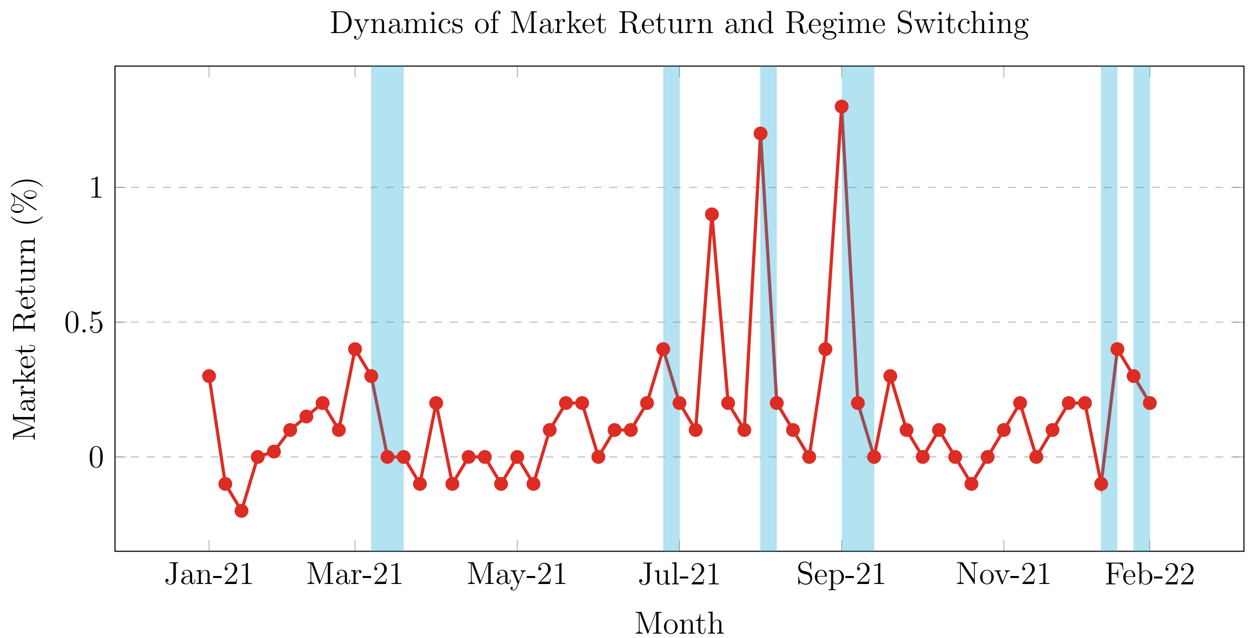
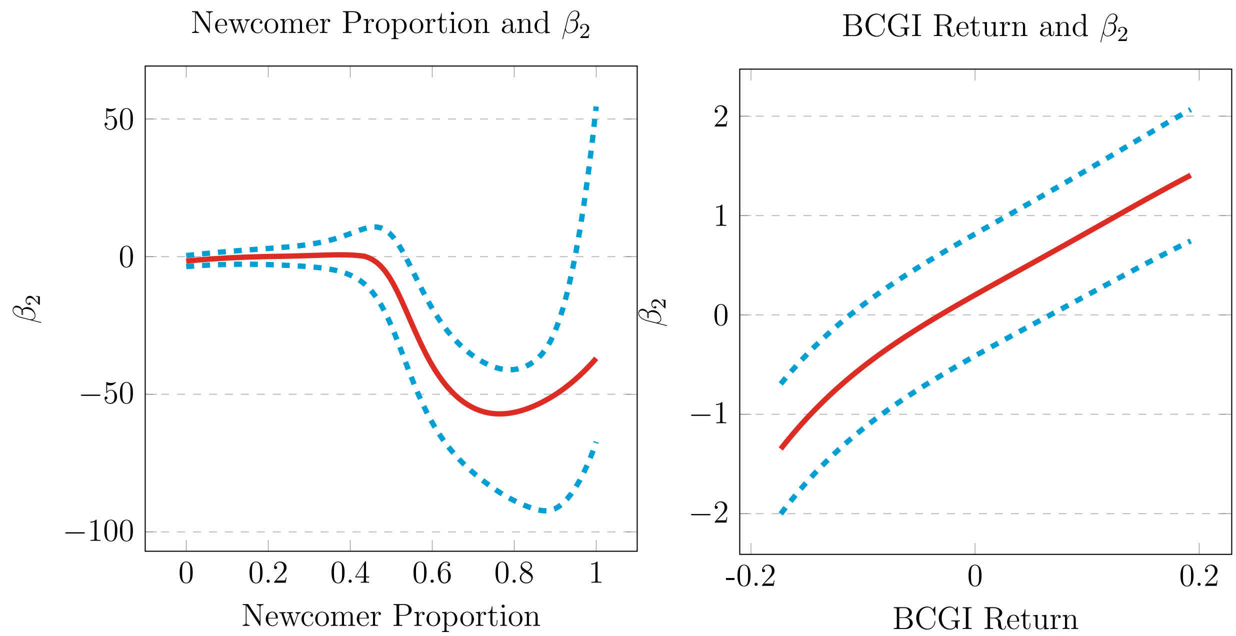


Figure 5. Dynamics of Market Return and Regime Switching. This graph provides a temporal illustration of the NFT market returns and the corresponding prevalence of herding behavior. The shaded areas represent periods of herding behavior and correlate with significant market returns. This alignment which indicate a positive relationship between absolute market returns and CSAD, offers a nuanced visual representation of herding behavior during periods of market volatility. The graph showcases oscillation between herding and independent decision-making, reflecting the complex, non-linear relationship between market returns and herding behavior. For instance, high but not extreme market returns still induce herding behavior, as indicated by the shaded areas. These shaded periods, driven by market momentum, correspond to key events in the NFT space that amplified market activity. In March 2021, the $69 million auction of Beeple's artwork led to increased trading activity. Similarly, the "Loot (for Adventurers)" project in September 2021 ignited a surge in market returns and herding behavior. The peaks in herding behavior in January and February 2022 coincide with a significant downturn in the NFT market, largely spurred by the decrease in the value of major cryptocurrencies. This period of market distress likely amplified uncertainty and risk perception among investors, leading to an increase in herding behavior. The graph, thus, encapsulates the interplay of market returns, investor behavior, social dynamics, and significant market events, all of which contribute to the dynamics of herding in the NFT market.



(a) Newcomer Proportion and β_2

(b) BCGI Return and β_2

Figure 6. β_2 dynamics, newcomers and BCGI returns. This graph presents the relationships between the β_2 coefficient, the percentage of newcomer investors in the NFT market, and the Bloomberg Galaxy Crypto Index (BGCI) returns. The β_2 coefficient represents the sensitivity of squared market returns and the cross-sectional absolute deviation, a proxy for herding behavior. The graph shows three distinct phases for the relation between β_2 and the newcomer proportion. Initially, from zero to around 0.44 newcomer proportion, β_2 remains relatively stable, indicating that the entrance of new investors doesn't immediately disturb the market's return volatility. Beyond the 0.44 mark, β_2 sharply decreases, suggesting a possible learning effect where newcomers, gaining experience and understanding of the market, reduce their herding behavior and consequently lower market return volatility. However, once the newcomer proportion exceeds 0.7, β_2 begins to rise again, suggesting a possible saturation effect where a market overpopulated by newcomers may reintroduce collective uncertainty, triggering increased herding behavior and thus higher market return volatility. In contrast, the relationship between β_2 and BGCI return reveals a linear trend, suggesting that herding behavior becomes more pronounced with increasing BGCI returns. This trend indicates that positive market sentiment, rather than market volatility, may be a key driver of herding behavior in the NFT market. It suggests that when the overall crypto market is performing well, investors in the NFT market are more likely to engage in herding behavior. This dynamic underscores the uniqueness of the NFT market, where the idiosyncratic value of individual NFTs and the overall market sentiment appear to shape investor behavior more than market volatility.

Appendix A

A.1 Data Extraction, Cleaning and Preparation

For the on-chain part of our analysis, we use Dune, Etherscan and the APIs of Alchemy and OpenSea, while for the off-chain portion, we employ the price oracle of Chainlink and the APIs of Discord and Twitter.

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[2]. 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

[2]The data extraction stopped in February 2022 due to the Russo-Ukrainian war, which introduced substantial levels of systematic volatility, making it impractical to examine the volatility of an individual asset class.

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.[3] 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[4]. 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. In the case of GIF files, we average the values of each frame.

[3] 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.

[4] 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.

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 is. Wash trading can greatly affect the price, the traded volume and the selling frequency. 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 (*Transaction DataSet* in Table 1), and the other comprising aggregated data per NFT, which was augmented with data from Discord and Twitter, and colour and image complexity data (*Aggregate DataSet* in Table 1). 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.

A.2 Feature Explanations

Table A1. Variable Descriptions. The following table list all the variables extracted from on and off blockchain sources 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 month
d_ann_users_month	Average number of unique active users in Discord "Announcement" channel per month
d_gen_messages_month	Average number of messages in Discord "General" channel per month
d_gen_users_month	Average number of unique active users in Discord "General" channel per month
day_last_sale	Day of last sale per single NFT
deployer_creator_generalities	Deployer type: Single Artist (Male, Female), Artists Collab., Company
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 between: 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