NBER WORKING PAPER SERIES

SELECTION-NEGLECT IN THE NFT BUBBLE

Dong Huang
William N. Goetzmann

Working Paper 31498
http://www.nber.org/papers/w31498

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2023

We thank Nick Barberis, Milad Nozari, Christophe Spaenjers and Amy Whitaker for helpful conversations. We thank SuperRare for providing data. We also thank participants at the 22nd ACEI International Conference on Cultural Economics and the 12th Portuguese Financial Network Conference for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w31498

Selection-Neglect in the NFT Bubble
Dong Huang and William N. Goetzmann
NBER Working Paper No. 31498
July 2023
JEL No. G1,G12,G14,G4,G40,G41

ABSTRACT

Using transaction data from a large non-fungible token (NFT) trading platform, this paper examines how the behavioral bias of selection-neglect interacts with extrapolative beliefs, accelerating the boom and delaying the crash in the recent NFT bubble. We show that the price-volume relationship is consistent with extrapolative beliefs about increasing prices which were plausibly triggered by a macroeconomic shock. We test the hypothesis that agents prone to selection-neglect formed even more optimistic beliefs and traded more aggressively than their counterparts during the boom. When liquidity for NFTs declined, observed NFT prices were subject to severe selection bias due in part to seller loss aversion delaying the onset of the crash. Finally, we show that market participants with sophisticated bidding behavior were less subject to selection bias and performed better.

Dong Huang
Yale School of Management
165 Whitney Ave.
New Haven, CT 06511
USA
dong.huang@yale.edu

William N. Goetzmann
School of Management
Yale University
Box 208200
New Haven, CT 06520-8200
and NBER
william.goetzmann@yale.edu

1. Introduction

By some measures, observed prices in the NFT art marketplace rose by a factor of more than 10,000 over a period from the summer of 2019 to the autumn of 2021, only to decline by summer 2022 in equally spectacular fashion, leaving market participants and observers puzzled by what market forces and psychological forces could have led to this event. The parallels to Dutch Tulipmania are unmistakable. NFT artworks – like rare tulips -- generate no economic dividends. They reflect what collectors are willing to pay for ownership and aesthetic enjoyment – and the beliefs and hopes by speculators about future demand. The NFT bubble is a rare opportunity for economists to study a speculative bubble in great detail. Blockchain records of NFT bids, offers, and transactions provide rich material to test theories about unusual market dynamics. In this paper, we use the COVID-era NFT bubble to test psychological theories about expectations and beliefs in asset markets. Our focus is on whether market participants rationally inferred prices and future demand from transactions, or instead whether conditioning upon a censored sample of realized returns was an important source of selection bias which may have supported the asset bubble.

Selection bias is a particular concern in the study of markets in which transactions are relatively rare events that depend on circumstances with a variety of conditioning variables.[2] This conditioning, if it is systematic, make observed prices biased measures of mean population value, and price trends biased measures of unrealized capital appreciation. Nevertheless, market transaction data are often taken as representative of unrealized values and returns.[3] In this paper we consider the case of economic agents who observe market prices and returns in an asset market and infer unobserved asset values and returns, despite a significant sample selection bias in the data generating process. We find evidence in one prominent NFT marketplace that participants conditioned their expectations and transactions on trends estimated from censored data, and that this may have played an important role in bubble formation.

Economic experiments have found considerable evidence that decision makers can be prone to selection-neglect and are misled by the observed information.[4] Theory shows that this can lead to market dislocation. For example, Hirshleifer and Plotkin (2021) find that a combination of salience and selection-neglect can explain investment booms. In contrast, non-experimental empirical evidence about selection-neglect is limited. Records from the NFT marketplace SuperRare over the period 2018-2022 allow us to

[2] For example, selection biases may arise in housing market (Goetzmann & Peng, 2006), mutual fund advertising (Koehler and Mercer, 2009), venture capital evaluation (Korteweg and Sorensen, 2010), saving and consumption decisions (Han et al., 2019), art investment (Korteweg et al., 2016), and media censorship (Enke, 2020).

[3] For example, housing market appraisals typically rely on recent comparable sales without considering the effects of seller reserves on the probability of observing a comparable sale. For extrapolative expectations in housing markets See Case and Shiller (1988 & 2003) and Goetzmann, Peng and Yen (2012).

[4] cf. Esponda and Vespa (2018), Barron et al. (2019), Enke (2020), and López-Pérez et al. (2022).

observe or infer beliefs of individual agents during a period of high market activity, and to estimate the magnitude of selection bias as well as the conditioning thresholds that helped to determine transaction decisions.

In this paper, we first test whether the market for NFTs in the 2021-2022 period was indeed a bubble – defined modestly as a sharp increase by orders of magnitude in value followed by a dramatic decline. To do this, we estimate an econometric model that adjusts for sample selection bias. The model addresses the problem of selection neglect that a simple averaging of prices – or more sophisticated methods of index construction such as repeat-sale-regression – does not. Our model allows us to estimate the censoring process that generated observed transaction prices. Next, we test whether market participants' neglect of the bias interacted with extrapolative beliefs and may have contributed to the NFT bubble.

An NFT is a digital asset that is "tokenized" - i.e., registered and transferable -- on a blockchain. Virtually unknown outside of a small group of creators and collectors before 2019, NFTs attracted attention in March, 2021 with the multi-million dollar auction of a work by Mike Winkleman, a.k.a. Beeple. This suddenly opened a new digital medium for artists to create and sell art and also introduced a new market for speculation in a novel collectibles market.[5] The transaction-level data we use are provided by SuperRare, a popular, curated online marketplace for high quality digital artworks. The sample covers all transactions, including purchases, sales, bids, and listings of artworks, from April 2018 to June 2022. This period contains the complete boom and at least a substantial part of the bust of the NFT bubble, and thus the data is potentially informative about investor trading behavior during the bubble episode.[6]

To start, our paper documents several features of the market's price and volume dynamics, which serve as evidence of a speculative bubble. As in DeFusco et al. (2022) and many others, we show that there is a significant lead-lag relationship between trading volume and market prices, which we estimate by a repeat-sale regression. More specifically, prices and volume comove and changes in volume generally lead changes in prices by one year. In addition, a rise in price dispersion is present in the market. The cross-sectional volatility of transaction prices jumps in November 2019 and continues increasing until the peak in early 2022, suggesting greater disagreement among market participants during the growth phase of the bubble. We also report the dynamics of bids and asks, as they are potentially informative about buyers' and owners' expectations about current and future prices.

As in real estate markets and art markets (cf. Goetzmann and Peng, 2006; Korteweg et al., 2016), sellers in the NFT market may set publicly observable reservation prices or have implicit ones in mind. In

[5] For a recent account of the history of the NFT market, see Amy Whitaker & Nora Burnett Abrams,The Story of NFTs : Artists, Technology, and Democracy (Rizzoli Electa, 2023).

[6] According to the data website NonFungible.com, the monthly sales of NFTs reached a peak of nearly $4.8 billion USD in September 2021. But by the end of our sample period, this number is only about $1.9 billion, more than 60% less than its maximum. The monthly dollar volume drops even further to $390 million in September 2022.

the presence of seller reserves, transaction prices may provide biased information about the market. Specifically, due to rational factors like liquidity constraints or irrational factors like loss aversion, seller reserves of NFTs can differ from market consensus value. Transactions will not happen, and prices remain unobserved until one buyer's bid meets or exceeds the reservation price. Therefore, transaction prices deliver inadequate knowledge about true market conditions.

We provide two pieces of evidence supporting the existence of selection bias in transaction prices. First, we proxy the potential transaction price of each untraded NFT by the mid-point of its latest bid and ask each month. If an untraded NFT has a price from only one side of the market, that price is used as the proxy. We then estimate the repeat-sale regression augmenting the actual transaction prices and the mid-point prices. The combined-price index is much lower than the RSR index using only transaction prices, suggesting significant positive selection bias in the baseline RSR.

Second, we conjecture that seller loss aversion is an important mechanism that generates selection bias. Loss aversion and hence a disposition effect suggests that investors are more likely to realize potential gains than losses, conditioned on a reference point. The reference point is typically taken as the purchase price of the asset, however it need not be. We observe that seller asking prices are often multiples of the purchase price. This may indicate an aversion to selling at a profit which does not meet expectations of an even higher gain. Using the listing data and following the analytical framework of Genesove and Mayer (2001) and Beggs and Graddy (2009), we show that sellers with potential losses tend to set 25% higher asking prices than those with potential gains. Since asking prices can be regarded as a proxy for seller's reservation price, this pattern indicates the reluctance to realize potential losses.

To estimate a selection-corrected index, we adopt the data-augmented Markov Chain Monte-Carlo (MCMC) algorithm implemented by Korteweg et al. (2016). We specify the selection equation as a function of returns since purchase and estimate the model parameters by recursively simulating latent observations and updating the ex-post parameter distributions. The selection-corrected index is lower than the baseline-RSR index and similar in magnitude to the bid-ask augmented RSR. Importantly, the corrected and uncorrected indexes identify the timing of the boom and bust in NFT prices. The corrected index declines much earlier than the uncorrected index.[7]

A natural question is whether investors are aware of the selection bias and adjust their beliefs accordingly, or instead they neglect it and act as if the observed prices are representative of true market conditions. This selection-neglect bias can be particularly influential if NFT investors are extrapolators whose beliefs about future prices rely on past observed prices. We hypothesize that if extrapolators fail to

[7] Using subset five major NFT collections, Borri et al (2022) find the typical correlation between selection-corrected and uncorrected indexes to range from .70 to .91. They find significant differences in the mean weekly returns for three of five majorNFT collections, selected ex post, consistent with selection bias, although they characterize it as economically insignificant (1% to 3% weekly).

account for selection bias, they will form even more over-optimistic price expectations due to their observing upward-biased prices. When true market conditions begin to deteriorate, the declining trend may be masked by selection bias. Investors who are unaware of the bias may then continue purchasing NFTs and delay the realization of a downturn in observed market prices.

We first test whether NFT investors extrapolated past returns at the beginning of the bubble. We find that, in aggregate, the past returns of both the baseline RSR and the selection-adjusted index are positively correlated to market trading volume. We then test whether selection bias has additional explanatory power to predict trading volume. Specifically, we proxy selection bias in each month as the difference between the rolling six-month returns of the standard repeat-sale index and the selection-adjusted index. This selection-bias spread is then included as an explanatory variable in a regression of trading volume on past adjusted returns. To test whether this relationship is asymmetric in returns, we include interaction terms between the selection bias spread and the positive and negative portions of past returns. Regression results show that a one-standard deviation increase in selection bias amplifies volume sensitivity about positive past true returns by almost double and weakens the relationship when past returns are negative. These findings support the hypothesis that selection bias influenced NFT investors' trade decisions.

Next, we test for differing abilities among market participants to perceive and correct for selection bias. We use two metrics of sophisticated bidding behavior as an indicator of potential awareness of true market conditions. The first is a measure of attentiveness to the broad cross-section of NFT market values. We hypothesize that investors bidding of NFTs across a range of price tiers are more likely to be attentive to general trends rather than focused on a particular market segment. . For each investor in each month, we measure bid dispersion as the range of her logarithmic bid prices scaled by the median in the previous three months. We use this as a proxy for her attention to the overall market price distribution. If investors' bids vary across a wider range, this suggests that they may better understand the actual price distribution at high and low extremes and are thus less subject to selection bias in the transaction prices compared to their counterparts. As such, the trading intensity of high-bid dispersion investors should be more sensitive to actual market price declines than that of low-bid dispersion investors. In the data, we measure trading intensity as market aggregate purchase volume by high vs. low bid dispersion groups and also by the number of individual investor purchases. We find results consistent with the hypothesis that there is heterogeneity in the propensity for selection-neglect.

The second measure of investor awareness to market conditions is the fraction of NFTs purchased below the asking price. In the data we observe that about 40% of transactions are settled at a price below seller asking price. We hypothesize that buyers in these transactions tend to be better in assessing the values of NFTs and correcting for overly optimistic beliefs. Similar to the results with bid dispersion, we find that investors with more NFTs purchased below asking prices are more responsive to price declines than the others.

Finally, by adding investors' characteristics as additional explanatory variables in the standard repeat-sale regressions, we examine whether investors with greater bid dispersion or a higher fraction of NFTs bought below asking prices have better investment performance. For realized returns, we use returns of all completed repeat-sale pairs as the dependent variable. To account for unrealized returns of NFTs that remain unsold in investors' inventory by the end of our sample period, we proxy them with MCMC selection-corrected index returns in the same holding periods. We find that astute investors identified by either measure outperformed their counterparts, even after including unrealized returns.

Our research contributes in several ways to the literature about behavior during a bubble. First, we use the NFT market to test the interaction of extrapolative beliefs with selection-neglect in the ascending and declining stages of a bubble. Our findings with market data supplement experimental evidence showing suboptimal decisions induced by individuals prone to selection-neglect (cf. Koehler and Mercer, 2009). The results are also closely related to theoretical papers such as Jehiel (2018), who proves that overoptimism can arise in equilibrium in the presence of selection-neglect, and Frick et al. (2022), who suggests that the neglect of social assortativity can lead to misperception of income inequality and political attitude polarization.

Second, we provide supportive evidence for behavioral biases, including loss aversion and extrapolative beliefs, in a new speculative market. These two behavioral biases have been documented in different markets around the world, including equity (Odean, 1998), real estate (Genesove and Mayer, 2001; Glaeser and Nathanson, 2017), derivatives (Pearson et al., 2021), and art markets (Pénasse and Renneboog, 2021). They have been proved to be critical in explaining bubbles in those markets.

Current work on NFTs mainly focuses on market dynamics and market structure (cf. Kräussl and Tugenetti, 2022, Borri et al., 2023, and Lommers, Kim, and Baioumy, 2023). Borri et al. (2023), examine factors that may affect NFT prices, including quality, rarity, and visual characteristics. A number of papers examine investor behavior in the NFT market. For instance, Oh, Rosen, and Zhang (2022) document the intensive participation of experienced investors in primary market sales of NFT collectables. Barbon and Ranaldo (2023) highlight the role of experienced investors in the NFT markets. Borri et al. (2023) test the impact of money illusion when the value of cryptocurrencies that denominate NFTs changes. However, the role of behavioral heuristics and biases and their potential contribution to the NFT bubble have not been directly explored. Our paper addresses this issue by examining the interactions between loss aversion, extrapolative beliefs, and selection-neglect.

The remainder of this paper is organized as follows. Section 2 provides an introduction of institutional background and our data from SuperRare. Section 3 describes the bubble and speculative activities in the market. Section 4 and 5 provide evidence for investors' loss aversion and extrapolative beliefs, respectively. We address the role of selection-neglect in Section 6. Section 7 concludes.

2. Institutional Background and Data

NFTs are digital assets that are tokenized on a blockchain which records all transactions in a way that makes it hard to modify, hack, or cheat. These assets are connected to "smart contracts," which confirm the underlying assets as individually unique, traceable, and verifiable.[8] NFTs are traded in a number of online marketplaces, including that of our data provider, SuperRare. SuperRare is a curated platform that screens for digital artworks of high quality. It was launched in April 2018 and as of June 2022, there were 35,211 NFTs created by 2,018 artists and owned by 5,510 investors. In this platform, bids, asks, and transactions are denominated and executed in the cryptocurrency Ethereum (ETH).

SuperRare selects and invites digital artists to create their work on the platform. After an artwork is created, it can be listed for sale. Owners can set a listing price (asking price) or wait for bids by prospective buyers. SuperRare charges a 15% commission for the first sale of the work. Buyers can resale the work by listing it again and the subsequent buyers can simply accept the asking price or offer lower bids. Prospective buyers can withdraw bids up to the moment of acceptance by sellers. Alternatively, sellers can set up a traditional scheduled auction or a reserve auction which is automatically triggered by a sufficiently high bid price, and this typically lasts for 24 hours.[9]

Data provided by SuperRare contain all events on their platform from April 2018 to June 2022. Events include NFT creation, listings, bids, and sales. For each event, we observe the unique identifier and the name of the NFT, the time of event, the price (in Ethereum), and the digital wallet addresses of the agents involved in the events. In most cases, we also have a short description and an image of the artwork.

2.1. Information Set

It is important for our analysis to understand what market participants are able to observe about current and past prices when they were making decisions. SuperRare provides considerable information in real time on their website. Figure 1 is a screenshot from the SuperRare platform. It includes a picture of the artwork, its name, a detailed technical description, the pseudonym of the artist who created it, the pseudonym of the current owner, the asking price in cryptocurrency (if there is one), the date of creation of the work, and a history of the transactions and transfers. For example, the transaction history for *The Remote Welder* is fairly simple. It was created i.e. "minted" on January 9, 2023 by artist Iñigo Bilbao

[8] https://www.forbes.com/sites/nicolesilver/2021/11/02/the-history-and-future-of-nfts/
[9] https://help.SuperRare.com/en/articles/4576436-how-auctions-work-on-SuperRare

[@ibl3d]. A link to the artist's site within SuperRare tells how many followers and how many collectors the artist has. The artwork was purchased on February 20th for .25 ETH, then ownership was transferred on the Ethereum blockchain that day to the current owner @juancruzeth. As of the time of writing, it was listed for sale for 69 ETH or the current equivalent of $115,300.

Figure 2 shows another type of sale: an auction with a reserve. The artist, Alex CGbans held a 24-hour auction on November 23, 2022. The reserve was met with a bid of 1 ETH, two bidders, @collin and @lorenerl, competed and one won with a bid of 1.49 ETH ($2,400). The site provides information about both of these accounts. Lorene R Earl has 5 followers and is following 8. @lorenerl has one NFT in her collection which last sold for 0.1 ETH. In contrast, @collin has 96 NFTs in his collection. The highest price he paid for any of these was 12.136 ETH or $57,500 on November 12, 2021. Sometime after he bought *The Frog King* he listed it for approximately double his purchase price: 3 ETH ($5,000) and it is currently listed for $516,500 [309 ETH].

Finally, Figure 3 shows the image and transaction history of *VR.Girl* by a widely collected and high-priced artist @hackatoo. This transaction history documents a series of offers, rather than a formal auction. Minted on December 16th, 2019, it attracted multiple offers and accepted one for $1,500 [12.5 ETH] on December 18th. *VR. Girl* sold May 25, 2021, for $297,100 [115 ETH].

These three examples give an idea of the real time information available to NFT market participants. The prominent display of past prices and bid history is particularly salient – while time on the market may be less so -- although it may be inferred by the time since last purchase. The fact that @billywhistler made a return of 19,700 in less than two years on her/his investment of $1,500 *VR.Girl* is an example of the stratospheric returns that some speculators in the NFT market realized during the peak of the bubble.

2.2. Summary Statistics

Table 1 Panel A provides summary statistics of event prices including bids, asks, and sale prices. All prices are converted into US dollars at prevailing exchange rates for comparison over time. First note that all prices are quite volatile and positively skewed: they can be less than $100 or as high as several million dollars. We also observe that the asking prices are generally substantially higher than bids, suggesting high seller reserves. The sale prices typically lie between the bids and the asks, as expected.

Table 1 Panel B reports the numbers of NFTs created and held by artists and investors on June 30, 2022, the last day in our sample period. Artists were heterogeneous in their productivity. The median artist created six artworks while some created more than 40. Artists held relatively few NFTs, suggesting that most artists were successful in the primary sales. The last row shows the distribution of NFT inventories among buyers. More than three quarters of the investors held only one or two NFTs. But if we account for the prices of these NFTs, the inventories are much larger than they appear.

Finally, Table 1 Panel C presents summary statistics of repeat-sale pairs. The log returns are highly right-skewed and more than three quarters of the pairs realized gains. The median log return is 1.75, corresponding to a nearly five-fold increase in the purchase price. The mean and median holding periods are 8.5 months and 6 months, respectively, a sign of speculative activity which we next explore.

3. Evidence of the Bubble and Speculative Activity

The existing literature has documented several common features of asset bubbles. These include sharply rising prices, a lead-lag relationship between prices and volume (DeFusco et al., 2022), an increase in price volatility (Xiong and Yu, 2011), and a dispersion in beliefs (cf. Barberis 2018 for a comprehensive discussion). Speculative activities of various forms, such as shorter holding periods and higher turnover (Bayer et al., 2020), and "riding the bubble" by sophisticated investors also rise during a bubble episode (Brunnermeier and Nagel, 2004, Temin & Voth, 2004). In this section we provide evidence that the NFT bubble displayed many of the classic attributes of an asset bubble. This suggests that inference from tests about agent's beliefs and behavior in this particular bubble episode may be relevant to explaining bubbles in general.

3.1. Price and Volume Dynamics

As a metric for a survival-biased index we construct a price index of NFTs using the baseline RSR. Specifically, we estimate equation (1) with OLS,

$$r_{i,b\to s} = \sum_{t=b_i+1}^{s_i} \mu_t + \varepsilon_{i,b\to s}, \tag{1}$$

where subscript i indexes repeat-sale pairs and subscript t the months. $r_{i,b\to s}$ is the log return of repeat-sale pair i purchased at month b_i and sold at month s_i. μ_t is the market index return in month t, while $\varepsilon_{i,b\to s}$ is the idiosyncratic return of repeat-sale pair i during the holding period.

Figure 4a shows the estimated monthly RSR price index of NFTs (the blue line) for the SuperRare repeat-sales observations. Before October 2019, the index grows at a relatively slow pace of about 3% a month. The NFT prices then increase rapidly to a peak in October 2021. During this boom period, the average return of the index is approximately 25% a month (or 300% annually), which is higher and lasts longer than returns to investment in Bitcoin (BTC) and Ethereum (ETH). After October 2021, the NFT prices start to fall, along with other digital assets and S&P500. The NFT price index thus demonstrate a clear boom and bust pattern that is related to other trends in digital assets as well as to the equity markets.

This co-movement suggests that these disparate markets may have been subject to similar macroeconomic trends affecting speculators and investors across different assets.

One of the limitations of the RSR for purposes of testing theories of market rationality is that it incorporates future information in estimating returns in any given period. To address this we consider the effects of using only backward-looking repeat-sale pairs. Figure 4b presents a series of indexes estimated with repeat-sale pairs available up-to each month, which constitute the information set that investors would have been able to rely on when forming their expectations about the future prices in the month. These indexes, which we call “up-to” indexes, are free of forward-looking bias and do not include price information of NFTs remaining unsold up to the month. The figure suggests that the up-to indexes are consistently higher than those with the richest information (the 2022Q2 index) and may be even higher than the unbiased index since the 2022Q2 index is still unable to account for all the untraded NFTs and is thus estimated with an upward bias. Van de Minne et al. (2020) observe that downward revisions in repeat-sale indexes are suggestive evidence of upward bias in observed prices.

DeFusco et al. (2022), among others, suggest bubble episodes are characterized by a lead-lag relationship between prices and volume. The prices of an asset and its trading volume comove at a relatively low frequency while, at a higher frequency, existing research documents that volume often declines earlier than prices as the bubble approaches collapse. Similar patterns are found in the SuperRare NFT market. Figure 5a plots the NFT price index and the monthly number of transactions. Generally speaking, the price and the volume share the same increasing, then declining, trend before and after late 2021, respectively. Note that volume peaked in March 2021, seven months earlier than the NFT price index, which remained high until October.

We follow DeFusco et al. (2022) and regress monthly log price p_t on lagged log volume v_{t-k}. The normalized coefficients of log volume, that is, the coefficients of correlation between the two variables, are shown in Figure 5b. The figure shows the correlation coefficients up to one year of leads ($k = -12$) and two years of lags ($k = 24$). The correlations are positive at most lags with a maximum at around 12 months, suggesting that the trading volume generally leads the price changes by one year.

Another widely recognized feature of a bubble is a rise in price dispersion. Increasing price volatility reflects greater heterogeneity of investor beliefs about future valuations, which can be a result of different extrapolative efforts (Barberis et al., 2018) or overconfidence on private information (Scheinkman and Xiong, 2003). Figure 6 illustrates the trends in price dispersion measured by the cross-sectional volatility of observed transaction prices in the month. Figure 6a and Figure 6b calculate the standard deviation of prices denominated in ETH and in USD, respectively. They show similar patterns: the volatility jumped in November 2019, maintained a rising trend until the peak in early 2022, and dropped afterwards. This pattern suggests an increase in price disagreement while the bubble was growing.

In the Appendix, we also show further evidence of dispersed beliefs using data on bid and ask prices. For any listed NFTs, potential buyers can bid by offering a price which the owner can decide to accept it or not. Dispersion in bid prices for the same NFT within a short period of time is a potential proxy for belief disagreement.[10] Appendix Figure 1 shows the average distance between the winning bid and the second largest or median bid for the same NFTs offered within a month. Due to a limited number of bids in the early stage of the market, the time series starts from January 2019. The figures show that the gap between the maximum bid and lower bids became larger and more volatile after 2021.

The SuperRare database also contains information about NFT asking prices, which can be set when owners post a work for sale. The asking price is thus an upper bound for the seller reserve, and the ratio of the asking price to the purchase price is an aspirational proxy for expected return. Appendix Figure 2 shows the average, median, 25th, and 75th percentiles of the log asking-to-purchase price ratios over the sample period and Appendix Figure 3 shows the interquartile range (i.e. the difference between the third and the first quantiles). These figures suggest a higher and more widespread mark-up in asking prices during the boom, indicating more optimistic and more dispersed beliefs about NFT valuation during the bubble.

3.2. Speculative Activity

This section documents the trading behavior of speculators -- agents characterized by frequent turnover. Following Bayer et al. (2020), we define speculative transactions as "short-term" sales or listings of NFTs which happen within one month of the most recent purchase. Since some set of NFTs are sold or listed for sale within a short time, it is more likely that these buyers purchased them for re-sale, instead of collection. Short-term speculators are thus defined as investors who conduct more than seven (the median) short-term speculative transactions during the sample period.

Figure 7 shows intensive speculative activity during the bubble episode. Figure 7a examines the number of purchases by short-term speculators as a fraction of total transactions in the market. At the beginning, purchases by short-term speculators constitute about 10-20% of market volume. But in mid-2020, more than half of the purchases in the market are made by speculators. The fraction drops quickly afterwards with two spikes around March and July 2021 due to a small number of very active investors. Some short-term speculators keep buying NFTs even after the market downturn in 2022. Figure 7a suggests that short-term speculators are relatively more active during the boom.

[10] This bidding procedure is different from a standard auction in the sense that the bidders can withdraw their offers any time as long as the owner has not accepted it, and there is no time limit in bidding and accepting the offers. So, the dispersion in bid prices may also be affected by sellers' reserve and their willingness and patience to wait for higher offers. Nonetheless, the differences in bid prices may still be a useful proxy for belief dispersion.

Figure 7b shows the fraction of speculative transactions among all transactions, i.e., the fraction of NFTs that are purchased and then sold or listed within one month.[11] The figure also suggests a rise in speculative activity in the period of rapid growth. Interestingly, as the observed market price index starts to turn down, excess speculative activity – as opposed to collector activity -- also declines. This may suggest entry of collectors who regard the price decline as an opportunity for long-term investment. – or of "patient" investors willing to bet on an eventual recovery.

4. Selection Bias in Successful Transactions

4.1. Selection Bias-Adjusted Price Index

As in real estate markets and art markets (Goetzmann and Peng, 2006; Korteweg et al., 2016), many sellers in the NFT markets set reservation prices or have implicit reserves in mind. Since reservation prices can be different from consensus values because of liquidity constraints and/or loss aversion and the prices will not be observed until the reservation prices are met, the presence of seller reserves can distort the information about the average market trend in transaction prices without appropriate adjustments.

To see how transaction prices can be biased, we first utilize the bids and asks in our data to estimate latent transaction prices for untraded NFTs. Specifically, in each month, for each untraded NFT whose transaction price is unobserved, we take the average of the most recent bid and the most recent ask as a proxy for the latent transaction price of that NFT. If there is only an ask from the seller or bid from potential buyers, that price is used as the proxy. If the NFT is neither listed nor bid on by any investors, no information about it is available in that month. As noted above, we then combine actual and proxied transaction prices and estimate a bid-ask augmented repeat-sale regression again as in equation (1).[12]

Figure 8 reports the indexed for the baseline RSR (red line) and the bid-ask augmented RSR (green line). The figure shows that the price index of NFTs shrinks after accounting for the unobserved prices. For example, the peak of the baseline RSR index is reached in October 2021 at around 46,000 while bid-ask augmented RSR index is only about 6,600 – almost seven times lower. This difference is mainly due to missing transaction prices – NFTs offered for sale at a high reserve relative to demand. The spread clearly suggests that the baseline RSR index requires adjustments for selection bias.

[11] It is different from Figure 4a as some speculative transactions are implemented by investors who speculate less frequently, and not all purchases by speculators are eventually defined as speculative transactions.

[12] Here we restrict our sample to the same NFTs as in the repeat-sale sample so any differences between the standard RSR index and the new index come solely from the unobservability of prices outside the repeat-sale intervals. This gives us a clean understanding about the effect of selection bias. In unreported results, we also include NFTs that are only traded once and that are never again traded. The corresponding RSR index has a similar trend and is also lower than the standard index. However, since the difference also comes from the expansion of the sample this test is less informative.

Several econometric approaches have been proposed to adjust for selection bias in the RSR. For example, Goetzmann and Peng (2006) construct a three-step maximum likelihood estimation to account for variations in seller reserve and the resulting bias. Vecco et al. (2022) use the Heckman (1976) two-step method and estimate the selection equation which includes selection-related artwork characteristics. Kortweg et al. (2016) use a data-augmented Markov chain Monte Carlo (MCMC) method that is most useful to our setting. In the presence of loss aversion, the probability of a sale depends on potential gains and losses, which are unobservable unless a transaction occurs. Hence, given priors about the data generating process, Kortweg et al. (2016) recursively impute all the latent prices based on their law of motion and observed transactions, and then update the selection equation with imputed potential gains and losses. This procedure is repeated until convergences of all parameters are reached.

Specifically, the data generating process consists of the three following equations:

$$p_i(t) = p_i(t-1) + \delta(t) + \varepsilon_i(t) \tag{2}$$

$$\begin{aligned} w_i(t) = \alpha_0 &+ I\big(r_i^p(t) \le 0\big)\big[r_i^p(t)\cdot\alpha_1 + r_i^p(t)^2\cdot\alpha_2\big] \\ &+ I\big(r_i^p(t) > 0\big)\big[\alpha_3 + r_i^p(t)\cdot\alpha_4 + r_i^p(t)^2\cdot\alpha_5\big] + h_i(t)\cdot\alpha_6 + h_i(t)^2\cdot\alpha_7 \\ &+ \eta_i(t) \end{aligned} \tag{3}$$

$$p_i^*(t) = \begin{cases} p_i(t), & w_i(t) \ge 0 \\ unobserved, & w_i(t) < 0 \end{cases}. \tag{4}$$

Equation (2) describes the law of motion for asset prices, in which $p_i(t)$ is the latent price of NFT i in month t, $\delta(t)$ is the common return factor (i.e., the market return of NFTs), and $\varepsilon_i(t)$ is the idiosyncratic risk deviation from the common return factor. Equation (3) is the selection equation, which is a function of potential holding period return $r_i^p(t) = p_i(t) - p_i(s)$, with s the time of last purchase. Thus the time interval since the last sale is defined as $h_i(t) = t - s$.[13] Nonlinear terms of $r_i^p(t)$ are included, capturing the disposition effect. $h_i(t)$ and its square are exogeneous variables for identification; otherwise, the model is identified only from the distribution assumptions.[14] Finally, as in a probit model, equation (4) specifies the selection rule. Price $p_i^*(t)$ is observed if and only if the latent selection variable $w_i(t)$ is no less than zero.

[13] For expositional simplicity we have omitted subscript *i* for purchase date *s* and valuation date *t* which will be different for each repeat-sale observation *i*.

[14] The time since the last sale is a valid instrument if it affects the probability of a sale (the relevance condition), but is uncorrelated with the idiosyncratic risk $\varepsilon_i(t)$ (the exogeneity condition). The first condition is typically true in many ways. For example, a collector with a high utility for the aesthetic "dividend" of an NFT is more likely to hold it for a longer duration. The exogeneity condition is generally satisfied since the current price $p_i(t)$ has incorporated all available information, including the time since the last sale, under the efficient market hypothesis.

The estimated selection equation (3) is as follows, with standard errors in brackets below the coefficients,[15]

$$\begin{aligned} w_i(t) = \underset{[0.039]^{***}}{-1.80} &+ I\left(r_i^p(t) \le 0\right)\left[r_i^p(t) \cdot \underset{[0.043]^{***}}{0.37} + r_i^p(t)^2 \cdot \underset{[0.008]^{***}}{0.04}\right] \\ &+ I\left(r_i^p(t) > 0\right)\left[\underset{[0.043]^{***}}{0.35} + r_i^p(t) \cdot \underset{[0.028]^{***}}{0.54} + r_i^p(t)^2 \cdot \underset{[0.003]^{***}}{(-0.05)}\right] \\ &+ h_i(t) \cdot \underset{[0.004]^{***}}{(-0.10)} + h_i(t)^2 \cdot \underset{[0.0001]^{***}}{0.003} \quad \underset{[\sigma^2 = 0.405^{***}]}{+\eta_i(t),} \end{aligned} \tag{5}$$

Loosely speaking, given a prior about parameter distributions, the MCMC algorithm iteratively refines the three equations with estimations from the previous iteration until a convergence is reached. Details of the algorithm can be found in the technical notes of Kortweg et al. (2016). The red line in Figure 8 reports the selection-adjusted NFT price index estimated from the MCMC algorithm, along with the baseline RSR index and the bid-ask augmented index.[16] The adjusted index is considerably lower than the unadjusted one, consistent with a positive selection bias throughout the sample period. Also note that the magnitude of the adjusted index is similar to the bid-ask augmented RSR index -- particularly in the early periods – supporting the validity of the MCMC estimation. In addition, because bids and asks are available in real time to market participants, it would have been possible for astute investors to estimate or at least approximate the selection bias implied by using realized returns. Finally, note that the adjusted index declines substantially after September, 2021 -- much earlier and faster than the unadjusted indexes. Such a difference in the later stages of our sample period suggests a large selection bias as market liquidity dried up.

Figure 9 plots the probability of a sale with respect to potential holding returns at different horizons. Consistent with the predictions from the disposition effect, the probability of a sale jumps abruptly at zero and increases with potential positive holding returns. The sale probability is relatively low for negative predicted returns around zero. The coefficients on both holding horizons are significant, indicating strong identification power.

A useful feature of the MCMC is that it estimates the latent price series for all NFTs in the market. We can use these to characterize the magnitude of the bias – i.e. observed vs. unobserved values at any given point in time. Figure 10a shows the distribution of latent and observed prices (in logs) in August 2021, when the last peak of trading volume occurred. Note that the distribution of observed prices (red bars) are

[15] Significance level: *: p<0.1, **: p<0.05, ***: p<0.01.

[16] In other related, unpublished work we simulate a selection process and find that an intercept term captures a transaction-weighted average of selection bias over entire estimation interval but the richer model estimated by the MCMC algorithm allows for conditional time variation in the bias.

shifted to the right compared to the distribution of latent prices, with a mean difference of 82% ($\approx e^{0.6}$). Figure 10b plots the monthly time series of mean differences. The differences ("biases") are predominantly positive and increase during the market downturn.

4.2. Loss Aversion as a Source of Selection Bias

Considerable research has documented investor loss aversion and anchoring effects. These two common psychological tendencies provide a potential explanation for selection bias in NFT markets. Specifically, NFT investors prone to loss aversion are less willing to realize losses during a market downturn. Investor sales decisions are thus conditional upon (anchored on) the purchase price of the asset. In the NFT marketplace this could manifest itself in high reservation prices in a falling market relative to the distribution of bids. As a result, actual transaction prices and realized returns observed in a falling market would likely be higher than the population distribution of NFT prices, leading to a positive bias.

In the presence of loss aversion, a seller with a potential loss would be expected to set a higher reservation price than for an NFT with a potential gain. In this section we test for evidence of loss aversion among NFT investors by following the frameworks of Genesove and Mayer (2001) and Beggs and Graddy (2009). Wang (2022) uses a similar framework to document anchoring effect and loss aversion in the *Crypto Punks* market – a particular subset of NFTs. These papers use asking prices as a proxy for sellers' unobserved reservation prices and regress asking prices (or the mark-up implied by asking prices) on the potential gain or loss measured by the difference between the previous purchase price and the predicted price, that is,

$$\begin{aligned} &Log\left(AskPrice_{i,t}\right)\left(or\ AskMarkUp_{i,t}\right) \\ &\qquad = \beta_0 + \beta_1 Log\left(PredictPrice_{i,t}\right) + \beta_2 \min\left\{0, PotentialReturn_{i,t}\right\} + \beta \boldsymbol{X} + \varepsilon_{i,t}, \end{aligned} \tag{6}$$

where $AskPrice_{i,t}$ is the asking price of NFT i in month t listed by the owner, $AskMarkUp \equiv Log\left(AskPrice_{i,t}\right) - Log\left(PredictPrice_{i,t}\right)$ is the owner mark-up. $PredictPrice_{i,t}$ is calculated as the previous purchase price grossed up by the $PotentialReturn_{i,t}$ which is the NFT price index return over the holding period from purchase date to listing date t estimated with the baseline repeat-sale regression described in Section 3.1.[17] The use of the baseline RSR implicitly assumes that investors do not perceive it as an unbiased measure. $\boldsymbol{X}$ variables include a quadratic function of holding horizon (in months), artist fixed

[17] Genesove and Mayer (2001) Beggs and Graddy (2009), estimate predicted prices by hedonic regression. Since our goal is to examine the effect of non-sales on predicted prices, we use the repeat sale methodology.

effects and time fixed effects. If the sellers are subject to loss aversion, the above prediction suggests a negative β_2 coefficient. The potential gains or losses then equal the index return within the holding periods.

We also consider the possibility of a nonlinear relationship between asking prices and returns by estimating a specification that includes squared returns and a dummy for losses.

$$\begin{aligned} &Log(AskPrice_{i,t})(or\ AskMarkUp_{i,t}) \\ &\qquad = \gamma_0 + \gamma_1 Log(PredictPrice_{i,t}) + \gamma_2 1(PotentialRet_{i,t} < 0) \\ &\qquad + \gamma_3 PotentialRet_{i,t}^{+} + \gamma_4 {PotentialRet_{i,t}^{+}}^2 + \gamma_5 PotentialRet_{i,t}^{-} \\ &\qquad + \gamma_6 {PotentialRet_{i,t}^{-}}^2 + \beta \boldsymbol{X} + \varepsilon_{i,t}. \end{aligned} \tag{7}$$

With the specification in equation (7), loss aversion predicts that γ_2 should be positive and that γ_3 and γ_4 (γ_5 and γ_6) should generate a negative slope within the positive (negative) returns interval.

Table 2 reports the results of regressions of asking prices on potential gains or losses. Columns 1 to 3 use log asking prices as the dependent variable. Columns 4 to 6 use asking price mark-ups. Standard errors clustered at a calendar month level are reported in parentheses. Consistent with the loss-aversion hypothesis for NFT investors, the coefficients of the loss dummy, $1(PotentialRet_{i,t} < 0)$, are significantly positive across all specifications. Sellers with potential losses tend to set asking prices approximately 25% higher than those with potential gains. The coefficients of $PotentialRet_{i,t}^{+}$, the positive part of potential returns, are also significantly negative except for the last specification, suggesting a lower seller reserve when facing with larger potential returns. $PotentialRet_{i,t}^{-}$ and its squared are generally insignificant, which may be due to a lack of statistical power since negative returns are less common in our sample.

5. Extrapolative Beliefs

5.1. Predictive Past Returns

Extrapolation theory predicts that investor expectations about future price changes positively depend on past asset returns (Glaeser and Nathanson, 2017; Barberis et al. ,2018; Chinco, 2020; DeFusco, Nathanson, and Zwick, 2022), irrespective of statistical or fundamental evidence to the contrary. In a market that has experienced past, rapid gains, this can result in over-optimism which, in turn, increases demand for the asset.[18] If extrapolative beliefs sufficiently prevalent, past asset returns can positively predict trading volume.

As our first test for extrapolative beliefs, we regress monthly trading volume, either in the entire market or constrained to new-entry buyers, on estimated past NFT returns, as shown in equation (8). The explanatory variable $LogReturn_{t-\tau\to t-1}$, are log returns from the beginning of month $t-\tau$ to the end of month $t-1$ calculated based on the estimated price indexes from three methods in Section 4.1.

$$Log(Volume_t) = \beta_0 + \beta_1 LogReturn_{t-\tau\to t-1} + \varepsilon_t. \tag{8}$$

Table 3 shows the relationship between trading volume and lagged NFT returns. We use a six-month horizon (the median holding period) for the lagged returns in the reported results. Results with different horizons are similar.[19] The significance of the coefficients is robust to different choices of the lags. Columns 1, 3, and 5 use the baseline RSR index, the MCMC selection-adjusted index, and the "up-to" RSR index to calculate past returns, respectively. The coefficients on past return are all positive and significant. They suggest that a 10% increase in six-month log returns, which is not uncommon in NFT markets, is associated with a 53~104% increase in market trading volume. The R-squares of the regressions are also large. Panel B of the table uses purchase trading volume from new buyers as the dependent variable and produces similar results. These findings indicate strong predictability consistent with extrapolation.[20]

To further explore the relationship, the even columns of Table 3 split the log returns into positive and negative parts; that is, $LogReturn^{+} = \max\{LogReturn, 0\}$, and $LogReturn^{-} = \min\{LogReturn, 0\}$. An interesting result is that, although the coefficients for positive past returns remain large, the coefficient for the negative past returns is insignificant. The lack of statistical power due to fewer observations in the negative return period is not the sole reason: with the MCMC selection-adjusted index as in column 4, 45% (20 out of 51) of the months in our sample period have negative past returns. Investors' inability to account for selection bias in seemingly high prices when the bias-corrected market values have declined may explain the finding. This hypothesis motivates our analyses of selection-neglect in Section 6.

Finally, we regress log trading volume on past realized returns of repeat sales completed in the past six months in the market. Since the platform provides the transaction history of each NFT both in ETH and USD, recent realized returns in the market are the information that investors can most easily observe or learn from the media. The explanatory variable is past market realized return measured by rolling six-month

---

[18] Case & Shiller test for extrapolative beliefs using survey data from rising housing markets. Goetzmann, Peng and Yen (2012) test for extrapolative beliefs in the 2006 US housing market using demand and credit proxies.

[19] Standard errors are adjusted for autocorrelation using the Newey-West correction with the optimal number of lags suggested by (Newey and West, 1987, 1994).

[20] The differences between any pair of coefficients are statistically insignificant. For example, the coefficient difference between columns 1 and 3 is 0.115, which is roughly only one standard error away from each other.

mean, maximum, or minimum. To account for the investment horizon, we also use the mean of monthly realized returns, which are realized returns divided by the lengths of holding periods in months.

Table 4 tests the explanatory power of past market realized returns on trading volume. Column 1 confirms that past realized returns are also positively associated with future trading. Columns 2 and 3 compare the effect of the highest and lowest observed realized returns. The (in)significance of maximum (minimum) returns suggests that investors tend to ignore relatively negative news in the market. In column 4, we test the effect of horizon-adjusted realized returns, which also displays significance.

5.2. A Positive Feedback Loop Trigger?

An important implication of extrapolation theory is that a positive price shock can trigger the initial price spiral that results in an asset bubble (Barberis et al., 2018). Some precipitation events identified in the literature are shocks to fundamentals such as technological innovation (Nicholas, 2007, Pastor & Veronese, 2009, Acheson et. al. 2009, Quinn & Turner 2021, Kindleberger, 1996 and Griffin et al., 2011), good news about the economy (Liao et al., 2022), and innovations in trade (Frehen et al., 2013). Others are speculation-related in which the shock is the sudden appearance of a new market accompanied by narratives of huge gains, for example the Mississippi Bubble on 1719-20 which sparked the neologism *millionaire*, or more recently, news of investment killings in cryptocurrency or Chinese warrants.[21] Many historical bubbles involved the rush to invest in new property markets, for example the 1920's Florida land bubble.[22]

Monetary policy is yet another potential channel to stimulate speculation. Allen and Gale (2000) for example show how agency frictions and cheap money can lead to an asset bubble. Regarding the NFT bubble during our sample period, an increase in speculative demand due to monetary easing in late 2019 and early 2020 could have triggered the initial price growth. The central bank's Federal Open Market Committee approved a series of interest rate cuts in late 2019 as a "midcycle adjustment." Figure 11 shows the effective federal funds rate along with the RSR NFT index. The federal funds effective rate decreased by 85 basis points from 2.40% in July 2019 and reached 1.55% in November 2019. Such an interest rate cut coincides with a turn in the price index and a sharp increase in speculative activity as shown in Figure 7c. The number of purchases by short-term speculators in November 2019 is four times larger than that in the previous month. The lowering interest rates in March and April 2020 as a measure against the Covid-19 pandemic may have further fueled the market. This period experienced another peak in speculation and

[21] For the Mississippi Bubble, Velde, 2003, for cryptocurrency news and network effects, Sokin & Xiong, 2023, for Chinese arrants, DeFusco et al., 2022, Pearson et al., 2021 and Xiong, 2021.

[22] See Sokolowski, 1933 for a list of American land bubbles including Florida.

returns.[23] By the same token the interest rate increases in early 2022 were accompanied by a downturn of market prices and volume.

6. Selection-Neglect

In the sections above, we show evidence consistent with the hypothesis that investors in the NFT market displayed extrapolative expectations. Extrapolative expectations can exist in markets that are not subject to selection bias. However, we hypothesized that they may be enhanced due to the failure of agents to properly infer average population prices from selection-biased transaction records. A natural question is whether or not extrapolators rationally corrected for selection bias when making investment decisions. If not, unrepresentative high transaction prices can interact with extrapolation and lead to greater overoptimism among naïve extrapolators. Selection-neglect can then accelerate a boom, attracting more investors even when true market conditions have deteriorated, and therefore delay the crash of a bubble.

6.1. Aggregate-Level Evidence

To examine how selection bias interacts with extrapolative beliefs at the market level, we include lagged selection bias as an additional explanatory variable in the extrapolation equation (8) in Section 5.1. Specifically, we use the following specification,

$$\begin{aligned} Log(Volume_t) = \beta_0 &+ \beta_1 LogUnbiasedReturn_{t-1} + \beta_2 SelectionBias_{t-1} \\ &+ \beta_3 LogUnbiasedReturn^{+}_{t-1} \times SelectionBias_{t-1} \\ &+ \beta_4 LogUnbiasedReturn^{-}_{t-1} \times SelectionBias_{t-1} + \varepsilon_t. \end{aligned} \tag{9}$$

In equation (9), we use the unbiased market returns calculated from the MCMC selection-adjusted index in the previous six months before t to capture the extrapolation effect. $SelectionBias$ is a measure of selection bias defined as the difference between the six-month market returns implied by a baseline RSR index (without an intercept) and an MCMC selection-adjusted index. Figure 12 presents the time series of the selection bias, which resembles the pattern of mean difference between observed and latent prices as shown in Figure 10b that are mostly positive and increases towards the end of our sample period as the market liquidity deteriorates.[24]

In addition to linear terms, we interact selection bias with the unbiased returns. Since selection bias may affect extrapolation in different directions under different market conditions, we separate the positive

---

[23] The bubble is unlikely to be triggered by a positive shock in the taste for NFTs. Google Trend statistics suggests that search frequency for keyword "non-fungible tokens" remained negligible until March 2021, which is much later than the first wave in NFT transactions in the market.
[24] In this specification, the use of the RSR index estimated without an intercept is conservative, since it captures an average component of the selection bias.

and negative parts of returns. If investors can correctly account for selection bias in the market, that is, they form their beliefs based on the unbiased past market returns, $SelectionBias$ then should have no additional explanatory power, giving insignificant β_2 and β_3. Otherwise, if selection bias amplifies extrapolative beliefs, we should obtain significantly positive interaction terms.

Table 5 reports the regression results for the interactions between extrapolation and selection bias. Column 1, as a benchmark, reproduces column 3 of Table 3A. Lagged selection bias and its interactions with unbiased returns are included in column 2. For ease of interpretation, the time series of selection bias is normalized to a zero mean and a unit standard deviation. Notice that the coefficient on lagged selection bias itself is negative: a rise in selection bias coincides with a deterioration in market liquidity and declining returns. The interaction term in the positive return region is 0.736 and significantly positive, indicating that during the bubble boom, a one standard deviation increase in selection bias almost doubles the extrapolative effect. In contrast, the interaction term is -0.534 when the unbiased return is negative, meaning that market trading volume is less sensitive to price declines when selection bias is larger. Column 3 splits the unbiased returns into positive and negative parts and generates similar results. These findings suggest that selection bias may have played a role in inducing over-optimistic extrapolative beliefs, accelerating the boom and delaying the collapse.

As a robustness check, columns 4 to 6 in Table 5 replicate the regressions in the first three columns, respectively, with selection bias defined as the difference between the six-month market returns implied by "up-to" RSR index and the six-month market returns in the MCMC selection-adjusted index. As before, selection bias amplifies extrapolation with positive returns and dampens extrapolation when returns are negative.

6.2. Proxies for Astute Investors

In the following section we propose measures to identify asture investors – i.e. those less subject to selection-neglect. To do this, we use observations of investors' bid and ask prices in our data to measure investors' ability to correct for selection bias in transaction prices and estimate the impact of selection-neglect across different investors. First we construct two variables to measure how informed investors are about true market conditions and how sophisticated they can be. For one variable, we use the dispersion of investor bid prices (and transaction prices if the bids are accepted) in the previous three months as a proxy for her ability to adjust for selection bias in sales data in the current month. The corresponding formula is, for each investor i in month t,

$$BidDispr_{i,t} = \frac{P_{0.75}\left(\{Bid\}_{i,t-2\leq s\leq t}\right) - P_{0.25}\left(\{Bid\}_{i,t-2\leq s\leq t}\right)}{Med\left(\{Bid\}_{i,t-2\leq s\leq t}\right)}, \tag{10}$$

where $\{Bid\}_{i,t-2\leq s\leq t}$ is the set of bid prices offered by investor i for any NFTs in months $t-2$ to t. For each investor-month-NFT bid price observations, we choose only the last offer. $P_{0.75}$, $P_{0.25}$, and Med are the third quartile, the first quartile, and the median of the bids, respectively. Therefore, the bid price dispersion of an investor is the range of bids on any NFTs in the previous three months, scaled by the median bid from her. We use quantiles instead of means and standard deviations because quantiles are robust to extreme values.

The intuition behind the proxy is that investors who bid for a wider price range of NFTs tend to do more research and have better knowledge of the actual price distribution of NFTs in the whole market, especially for low-priced and illiquid NFTs. Hence, they are more likely to understand how upward-biased the realized transaction prices can be and adjust their expectations about future price changes accordingly. In contrast, investors who focus on a small range of NFTs and whose median bids are relatively large may tend to neglect the selection bias.

As a first step, we examine how the trading volume from high and low bid dispersion investors changes over time. Specifically, in each month, we divide the investors who have valid bid dispersion measures (i.e., investors who make at least one bid in the previous three months) equally into two groups based on bid dispersion. The total purchase volume by the bid dispersion groups is shown in Figure 13a. The first finding is that investors with higher bid dispersion traded more actively and generated more volume than their counterparts. More importantly, the volume gap between the two groups narrowed over time, especially after July 2021, when price appreciation slowed down. The volume from high-dispersion investors shrank while the volume from low-dispersion investors remains high. Figure 13b plots the ratio of purchase volume between the high- and low-dispersion groups. The ratio decreases from more than 10 times in early 2020 to around 5 times in mid-2022, suggesting that the demand of high-dispersion investors is more responsive to deterioration in market conditions than the demand of low-dispersion investors.[25]

Table 6 repeats the analysis in Table 5 by replacing the market trading volume with the purchase volume from the two types of investors differentiated by their bid price dispersion as the dependent variable. The first three columns use low-dispersion investor purchase volume while the last three columns use volume from high-dispersion investors. The relative magnitude of the coefficients suggest that low-dispersion investors are more responsive to price increases than to price declines. In addition, the coefficients on the interaction term between negative past returns and selection bias are all negative and the magnitudes are larger for low-dispersion investors, suggesting that both types of investors are subject to

[25] Instead of using month-specific cut-offs of bid dispersion to divide investors, we can also use a fixed cut-off to group the investors every month. For example, in Appendix Figure 4, we define low-dispersion group as investors with $BidDispr = 0$, i.e., only one bid in previous three months, and high-dispersion group $BidDispr > 0$ and plot purchase volume per investor every month by the two groups. Similar patterns in purchase volume and the gap are obtained.

selection bias in market downturns -- low-dispersion investors more so. The low-dispersion investors' net sensitivity to negative past returns when selection bias increases by one standard deviation is even negative (-0.259 = 0.298 + (-0.557)) though insignificant; that is, they continue purchasing even when the prices have dropped when selection bias is high. These results are consistent with the intuition that investors with low bid dispersion fail to correct for selection bias in the realized sale prices.

Another metric for investor ability to adjust for selection bias is the fraction of NFTs that investors purchase at prices below their asking prices. Specifically, in each month t our second ability measure for investor i, denoted by $BuyLow_{i,t}$, is defined as the fraction of NFTs purchased below asking prices among all NFTs purchased by investor i no later than month t.

The rationale is as follows. To buy an NFT, in addition to traditional scheduled auctions and reserve auctions, potential buyers can also simply accept the asking price set by the seller or offer a lower bid price, which the seller can decide to accept or not. Therefore, if some investors are sophisticated, they will realize that asking prices generally incorporate the most optimistic expectations in the market and cannot reflect true market expectations. These investors will hence tend to bargain with sellers and buy NFTs at prices lower than the asks. Appendix Figure 5 shows the distribution of the ratios of transaction prices to asking prices among all successful transactions. In the data, we find that about 40% of the transactions are settled at transaction prices that are less than the corresponding asking prices. Conditional of settling below asks, the magnitudes of the discounts are also large – many transactions receive around 60% to 70% off.

6.3. Selection-Bias Across Investors

Individual investor transaction information, provide further evidence on how past aggregate market returns and selection bias affects trading behavior of investors with various debiasing ability. We test our aggregate market results with individual-level data. To do so, we aggregate each investor's transaction to the monthly level and in each month we select "active" investors who made at least one bid in the previous three months so that the bid dispersion variable is well-defined. An investor-month level panel regression as follows is then estimated:

$$\begin{aligned} Trade_{i,t} = \alpha_i &+ \beta_1 LogUnbiasedReturn^{+}_{t-1} + \beta_2 LogUnbiasedReturn^{-}_{t-1} \\ &+ \beta_3 SelectionBias_{t-1} + \beta_4 LogUnbiasedReturn^{+}_{t-1} \times SelectionBias_{t-1} \\ &+ \beta_5 LogUnbiasedReturn^{-}_{t-1} \times SelectionBias_{t-1} + \varepsilon_{i,t}. \end{aligned} \tag{11}$$

where subscripts i and t represent investor and month, respectively. In the equation, $LogUnbiasedReturn^{+}_{t-1}$ and $LogUnbiasedReturn^{-}_{t-1}$ are the positive and negative parts of the six-month lagged MCMC selection-adjusted index log returns, respectively, and $SelectionBias_{t-1}$ is lagged selection bias – all of them are defined as before. The dependent variable $Trade_{i,t}$ is either the log purchase volume or an indicator of a positive purchase by investor i in month t. When the dependent variable is log purchase volume, the equation is estimated by OLS and α_i is the investor fixed effect. When the dependent variable is the trade indicator, the equation is estimated by a conditional logit regression and α_i is absorbed by conditioning on investors when calculating the conditional maximum likelihood function. Standard errors are clustered at the investor level to account for autocorrelation within investors.

Table 7 reports the regression results. For ease of interpretation, selection bias is normalized. The first three columns use log purchase volume in each investor-month as the dependent variables and the last three use the trading indicator. The coefficients of past log returns are all positive, consistent with the prior results of extrapolative beliefs among NFT investors. As in Section 6.1, the interaction of positive (negative) past returns and selection bias is positive (negative). This is further evidence that selection bias reinforces extrapolation during periods of NFT appreciation and reduces responsiveness when the market falls.[26]

To further examine how the impact of selection bias varies across investors of different selection-neglect levels, we fully interact the explanatory variables in equation (11) with the proxy for investor debiasing ability prior to the current month, $Ability_{i,t-1}$, which is either the investor's previous bid price dispersion, $BidDispr_{i,t-1}$, or the fraction of NFTs purchased below asks, $BuyLow_{i,t-1}$, as shown in equation (12) below:

$$\begin{aligned} Trade_{i,t} = \alpha_i &+ \beta_1 LogUnbiasedReturn_{t-1}^{+} + \beta_2 LogUnbiasedReturn_{t-1}^{-} + \beta_3 SelectionBias_{t-1} \\ &+ \beta_4 LogUnbiasedReturn_{t-1}^{+} \times SelectionBias_{t-1} \\ &+ \beta_5 LogUnbiasedReturn_{t-1}^{-} \times SelectionBias_{t-1} + \beta_6 Ability_{i,t-1} \\ &+ \beta_7 LogUnbiasedReturn_{t-1}^{+} \times Ability_{i,t-1} + \beta_8 LogUnbiasedReturn_{t-1}^{-} \times Ability_{i,t-1} \\ &+ \beta_9 SelectionBias_{t-1} \times Ability_{i,t-1} \\ &+ \beta_{10} LogUnbiasedReturn_{t-1}^{+} \times SelectionBias_{t-1} \times Ability_{i,t-1} \\ &+ \beta_{11} LogUnbiasedReturn_{t-1}^{-} \times SelectionBias_{t-1} \times Ability_{i,t-1} + \varepsilon_{i,t}. \end{aligned} \tag{12}$$

Again, for ease of interpretation, we normalize the measures of selection bias and investor ability. To understand the three-way interaction terms better, we can rewrite the terms that are related to $LogUnbiasedReturn_{t-1}^{+}$ in the following form as in Dawson and Richter (2006):

$$\left(\beta_1 + \beta_4 SelectionBias_{t-1} + \beta_7 Ability_{i,t-1} + \beta_{10} SelectionBias_{t-1} \times Ability_{i,t-1}\right) \times LogUnbiasedReturn_{t-1}^{+}. \tag{13}$$

This formulation suggests that the degree of extrapolative beliefs when returns are positive, i.e., the coefficient on $LogUnbiasedReturn_{t-1}^{+}$, depends jointly on selection bias and investor ability. As selection bias increases by one standard deviation, the whole coefficient changes by β_4. An increase in investor

[26] Also note that the number of observations is greatly reduced in the conditional logit regressions since the trading indicator of some investors have values of all 1's or 0's and they are excluded from the logit regressions.

ability then strengthens (weakens) the effect by $|\beta_{10}|$ when β_{10} is positive (negative). Similar manipulation can be done on terms involving $LogUnbiasedReturn_{t-1}^{-}$. Therefore, the coefficients on the triple interaction terms, β_{10} and β_{11}, capture the changes in the impact of selection bias when past market returns are positive or negative, respectively, and when investor ability increases by one standard deviation above overall average.

We report the regression results in Table 8. The odd columns use the log purchase volume as the dependent variable while the even columns use an indicator for positive purchase as the dependent variable and are estimated with conditional logit regressions. The first two columns use investors' bid price dispersion as the proxy for investor ability. First notice that the coefficients of singleton return terms are positive, indicating the existence of extrapolative beliefs. The selection bias term alone is negative as before since higher selection bias is usually associated with lower market liquidity. A positive bid dispersion term suggests that high-dispersion investors on average trade more actively, as we have seen in Figure 13. The interaction term $LogUnbiasedReturn_{t-1}^{+} \times SelectionBias_{t-1}$ is significantly positive while the triple interaction term with $BidDispr_{i,t-1}$ is insignificant. It suggests that a rise in selection bias increases the sensitivity to positive past returns, but the effects are similar among investors of high and low bid dispersion. On the contrary, the negative return interaction term $LogUnbiasedReturn_{t-1}^{-} \times SelectionBias_{t-1}$ is negative and significant and the coefficient on the triple interaction term with $BidDispr_{i,t-1}$ is positive and significant. The results suggest that an increase in selection bias reduces investors' response to negative past returns, and less so among high-dispersion investors. This finding is consistent with our previous results and our proposition that high-dispersion investors are subject to less selection bias.

Similarly, columns 3 and 4 in Table 8 use the fraction of NFTs purchased below asking prices as a measure of investor ability to correct for selection bias. Consistent with the previous results, the three-way interaction term $LogUnbiasedReturn_{t-1}^{-} \times SelectionBias_{t-1} \times Ability_{i,t-1}$ is positive, though insignificant when using log purchase volume as the dependent variable, so investors who tend to buy NFTs at lower prices are less subject to selection bias when market downturns.

In addition, the signs of the coefficients on past returns and their interactions with selection bias are the same as the corresponding coefficients in the first two columns, confirming the impacts of selection bias. The coefficients on the interactions between past returns and ability are both negative, indicating that high-ability investors extrapolate less over past returns. Finally, high-ability investors buy less aggressively under the influence of selection bias and positive past returns since the triple interaction terms of positive past returns, selection bias, and ability is negative and significant in the last column.

In sum, our results suggest that selection bias in the market prices can mask actual price decline so that investors may not fully respond to price drops by reducing NFT purchases. Investors with large bid price dispersion or with a larger fraction of NFTs purchased below asking prices tend to understand true market conditions better than their counterparts and hence are less affected by selection bias.

6.4. Performance of Sophisticated Investors

Previous sections show that astute investors, as identified by the two ability proxies are less influenced by selection bias during market decline. As a complement, we next assess the performance of these investors and test whether the ability to correct for selection bias was associated with superior investment performance.

To do so, we follow Mei and Moses (2002) and Borri et al. (2023) by adding investor characteristics at the time of purchase into standard repeat-sale regression equation. Specifically, we estimate the following equation,

$$r_{i,b\to s} = \sum_{t=b_i+1}^{s_i} \mu_t + \gamma(s_i - b_i) \times X_{i,b} + \varepsilon_{i,b\to s}, \tag{14}$$

where b_i and s_i are the time of purchase and sale for the repeat-sale pair i, respectively. In addition to time dummies μ_t's as in standard repeat-sale regression, investors' characteristics at the time of purchase, X_{i,b_i}, are also included in the regression. The characteristics are scaled by the length of holding period $(s_i - b_i)$ so that the impact of investor characteristics γ is measured at monthly basis. The first investor characteristic variables are the two ability proxies: bid price dispersion ($BidDispr$) and the fraction of NFTs purchased below asking prices ($BuyLow$). Since both measures require at least one bid or sale, an indicator for missing values among inactive or new investors is also included. Following current literature we also use the cumulative number of NFTs bought (in log, $Log(\#\ buy)$) and investors' average length of holding periods in the regression. The former is a proxy for investors' trading experience and the latter helps identify speculators (short holding periods) and collectors (long holding periods), who have different required capital returns.

Table 9 reports the relationship between investor performance and their characteristics using repeat-sale regressions. For conciseness, only the coefficients on the interaction terms are reported in the table. Panel A use completed repeat-sale pairs as in the standard setting so only realized returns are evaluated. In contrast, regressions in Panel B also include unrealized returns of NFTs that remain unsold in investors' inventory by the end of our sample period. We use the MCMC selection-corrected index returns within the corresponding holding intervals as the proxies for the unrealized returns.

The first column uses bid price dispersion in the regression. The positive and significant coefficients in both panels suggest that investors with larger bid dispersion generate significantly higher returns, even though the inclusion of unrealized returns of unsold NFTs reduces the effect by more than 80%. New or inactive investors lacking a bid dispersion measure also have higher realized returns, but the difference disappears once we account for unrealized returns. We obtain similar results when using $BuyLow$ as the proxy for ability as shown in column 2. These findings confirm that sophisticated investors as identified by the proposed ability proxies outperform other investors, consistent with their astute response to market decline in the presence of selection bias.

Second, column 3 includes investors' past purchases, a measure of investor experience, as the additional explanatory variable. In contrast to Borri et al. (2023), we find that experienced investors had significantly lower realized returns compared to unexperienced investors. The return gap remains negative but becomes insignificant after including unrealized returns. These findings suggest that, instead of experienced investors underperforming their counterpart, experienced investors are more likely to realize losses. Since there is large selection bias in successful transactions, realized returns understate the investment performance of experienced investors, on average.

Lastly, we examine the relationship between investor performance and their average holding time in column 4. Lovo and Spaenjers (2018) predict that investors who derive greater utility from the NFTs tend to hold longer and are more likely to sell their assets at lower prices. Consistent with their theoretical predictions and the finding in Borri et al. (2023), we find that investors with longer holding periods receive lower realized and unrealized returns.

7. Conclusion

In this paper we explore the extent to which extrapolation and selection-neglect biases may have contributed to the NFT bubble in 2020-2022. We use data on NFT prices, bids and offers and trading volume to test the hypothesis that investors were subject to the extrapolation heuristic, where expectations about future returns are based on past returns, and find evidence consistent with this hypothesis. At the same time, upward selection bias naturally arises in the NFT market since transactions are censored by high positive expectations by investors who extrapolated past returns in a rising market, and loss-averse behavior in a declining market.

We next test whether selection bias played a salient role in biasing the expectations of NFT market participants. We construct a proxy for selection bias and test for selection-neglect -- the inability to adjust valuations for the fact that market transactions are a biased measure of market value and investment returns. We find that selection-neglect amplified extrapolation and may have accelerated the bubble during the boom. During the market downturn, however, selection bias likely prevented investors prone to selection-neglect from recognizing that low volume signaled a market decline and thus delayed price revelation. These effects are more pronounced for investors with less knowledge about market conditions.

Finally we use proxies for investor sophistication to test for clientele effects in selection-neglect. We provide evidence that the relatively more sophisticated investors were less-subject to selection-neglect, and demonstrate that this awareness resulted in positive relative investment performance during the NFT bubble.

References

Allen, F. and Gale, D., (2000). Asset Price Bubbles and Monetary Policy.

Acheson, G.G., Hickson, C.R., Turner, J.D. and Ye, Q., (2009). Rule Britannia! British stock market returns, 1825-1870. *The Journal of Economic History*, *69*(4), pp.1107-1137.

Barberis, N. (2018). Psychology-based models of asset prices and trading volume. In *Handbook of Behavioral Economics: Applications and Foundations* 1 (Vol. 1, pp. 79-175). North-Holland.

Barberis, N., Greenwood, R., Jin, L., & Shleifer, A. (2018). Extrapolation and bubbles. *Journal of Financial Economics*, 129(2), 203-227.

Barbon, A. & Ranaldo, A. (2023). NFT Bubbles. *arXiv*:2303.06051.

Barron, K., Huck, S., & Jehiel, P. (2019). Everyday econometricians: Selection neglect and overoptimism when learning from others. *WZB Discussion Paper* No. SP II 2019-301.

Bayer, P., Geissler, C., Mangum, K., & Roberts, J. W. (2020). Speculators and middlemen: The strategy and performance of investors in the housing market. *The Review of Financial Studies*, 33(11), 5212-5247.

Beggs, A., & Graddy, K. (2009). Anchoring effects: Evidence from art auctions. *American Economic Review*, 99(3), 1027-39.

Borri, N., Liu, Y., & Tsyvinski, A. (2023). The economics of non-fungible tokens. SSRN Working Paper No.4052045.

Brunnermeier, M., & Nagel, S. (2004). Hedge funds and the technology bubble. *The journal of Finance*, 59(5), 2013-2040.

Brown, S.J., Goetzmann, W.N. and Ross, S.A., 1995. Survival. *The Journal of Finance*, *50*(3), pp.853-873.

Cai, J., He, J., Jiang, W., & Xiong, W. (2021). The whack-a-mole game: Tobin taxes and trading frenzy. *The Review of Financial Studies*, 34(12), 5723-5755.

Case, K. and Shiller, R., 1988. The behavior of home buyers in boom and post-boom markets. *New England Economic Review*, (Nov), 29-46.

Case, K. E., & Shiller, R. J. (2003). Is there a bubble in the housing market? *Brookings Papers on Economic Activity, Economic Studies Program, The Brookings Institution*, *34*(2), 299–362.

Chinco, A. (forthcoming). The ex-ante likelihood of bubbles. *Management Science*.

Dawson J. F., Richter, A.W. (2006). Probing three-way interactions in moderated multiple regression: development and application of a slope difference test. Journal of Applied Psychology. 91(4), 917-26.

DeFusco, A. A., Nathanson, C. G., & Zwick, E. (2022). Speculative dynamics of prices and volume. *Journal of Financial Economics*, 146(1), 205-229.

Enke, B. (2020). What you see is all there is. *The Quarterly Journal of Economics*, 135(3), 1363-1398.

Esponda, I., & Vespa, E. (2018). Endogenous sample selection: A laboratory study. *Quantitative Economics*, 9(1), 183-216.

Frehen, R. G., Goetzmann, W. N., & Rouwenhorst, K. G. (2013). New evidence on the first financial bubble. *Journal of Financial Economics*, 108(3), 585-607.

Frick, M., Iijima, R., & Ishii, Y. (2022). Dispersed behavior and perceptions in assortative societies. *American Economic Review*, 112(9), 3063-3105.

Genesove, D., & Mayer, C. (2001). Loss aversion and seller behavior: Evidence from the housing market. *The Quarterly Journal of Economics*, 116(4), 1233-1260.

Gill N., Mathur, A. and Conde, M.V., 2022. A Brief Overview of AI Governance for Responsible Machine Learning Systems. *arXiv preprint arXiv:2211.13130*.

Glaeser, E. L., & Nathanson, C. G. (2017). An extrapolative model of house price dynamics. *Journal of Financial Economics*, 126(1), 147-170.

Goetzmann, W. N. (1992). The accuracy of real estate indices: Repeat sale estimators. *The Journal of Real Estate Finance and Economics*, 5(1), 5-53.

Goetzmann, W., & Peng, L. (2006). Estimating house price indexes in the presence of seller reservation prices. *Review of Economics and Statistics*, 88(1), 100-112.

Goetzmann, W.N., Peng, L. and Yen, J., 2012. The subprime crisis and house price appreciation. *The Journal of Real Estate Finance and Economics*, *44*, pp.36-66.

Griffin, J. M., Harris, J. H., Shu, T., & Topaloglu, S. (2011). Who drove and burst the tech bubble?. *The Journal of Finance*, 66(4), 1251-1290.

Han, B., Hirshleifer, D., & Walden, J. (2019). Visibility bias in the transmission of consumption beliefs and undersaving. *NBER Working Paper* No. w25566.

Han, B., Hirshleifer, D. and Walden, J., 2022. Social transmission bias and investor behavior. *Journal of Financial and Quantitative Analysis*, *57*(1), pp.390-412.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of Economic and Social Measurement*, volume 5, number 4 (pp. 475-492). NBER.

Hirshleifer, D. and Plotkin, J.B., 2021. Moonshots, investment booms, and selection bias in the transmission of cultural traits. *Proceedings of the National Academy of Sciences*, *118*(26), p.e2015571118.

Koehler, J. J., & Mercer, M. (2009). Selection neglect in mutual fund advertisements. *Management Science*, 55(7), 1107-1121.

Korteweg, A., Kräussl, R., & Verwijmeren, P. (2016). Does it pay to invest in art? A selection-corrected returns perspective. *The Review of Financial Studies*, 29(4), 1007-1038.

Korteweg, A., & Sorensen, M. (2010). Risk and return characteristics of venture capital-backed entrepreneurial companies. *The Review of Financial Studies*, 23(10), 3738-3772.

Kräussl, R., & Tugnetti, A. (2022). Non-Fungible Tokens (NFTs): A Review of Pricing Determinants, Applications and Opportunities. *SSRN Working Paper* No. 4112429.

Liao, J., Peng, C., & Zhu, N. (2022). Extrapolative bubbles and trading volume. *The Review of Financial Studies*, 35(4), 1682-1722.

López-Pérez, R., Pintér, Á., & Sánchez-Mangas, R. (2022). Some conditions (not) affecting selection neglect: Evidence from the lab. *Journal of Economic Behavior & Organization*, 195, 140-157.

Lovo, S., & Spaenjers, C. (2018). A model of trading in the art market. *American Economic Review*, 108(3), 744-774.

Mei, J., & Moses, M. (2002). Art as an investment and the underperformance of masterpieces. *American Economic Review*, 92(5), 1656-1668.

Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3), 703–708.

Newey, W. K., & West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, 61(4), 631-653.

Odean, T. (1998). Are investors reluctant to realize their losses?. *The Journal of Finance*, 53(5), 1775-1798.

Oh, S., Rosen, S., & Zhang, A. L. (2022). Investor experience matters: Evidence from generative art collections on the blockchain. *SSRN Working Paper* No. 4042901.

Pearson, N. D., Yang, Z., & Zhang, Q. (2021). The Chinese warrants bubble: Evidence from brokerage account records. *The Review of Financial Studies,* 34(1), 264-312.

Scheinkman, J. A., & Xiong, W. (2003). Overconfidence and speculative bubbles. *Journal of Political Economy*, 111(6), 1183-1220.

Temin, P. and Voth, H.J. (2004). Riding the south sea bubble. *American Economic Review*, *94*(5), pp.1654-1668.

van de Minne, A., Francke, M., Geltner, D., & White, R. (2020). Using Revisions as a Measure of Price Index Quality in Repeat-Sales Models. *Journal of Real Estate and Financial Economics,* 60, 514–553.

Vecco, M., Chang, S., & Zanola, R. (2022). The more you know, the better: A Heckman repeat-sales price index. *The Quarterly Review of Economics and Finance,* 85, 194-199.

Wang, Y. (2022). Anchoring Effect and Loss Aversion: Evidence from the Non-Fungible Token Market. *SSRN Working Paper* No. 4097185.

Whitaker, A. & Burnett Abrams, N. (2023) *The Story of NFTs: Artists, Technology, and Democracy*, Rizzoli Electa.

Xiong, W., & Yu, J. (2011). The Chinese warrants bubble. *American Economic Review*, 101(6), 2723-53.

Figures

Figure 1. Sample NFT listing NFT.

This figure presents an NFT artwork named *The Remote Welder* that was listed on the SuperRare platform on February 21, 2023. The listing includes a history of transfers and prices on the SuperRare platform, artist and owner identifiers.

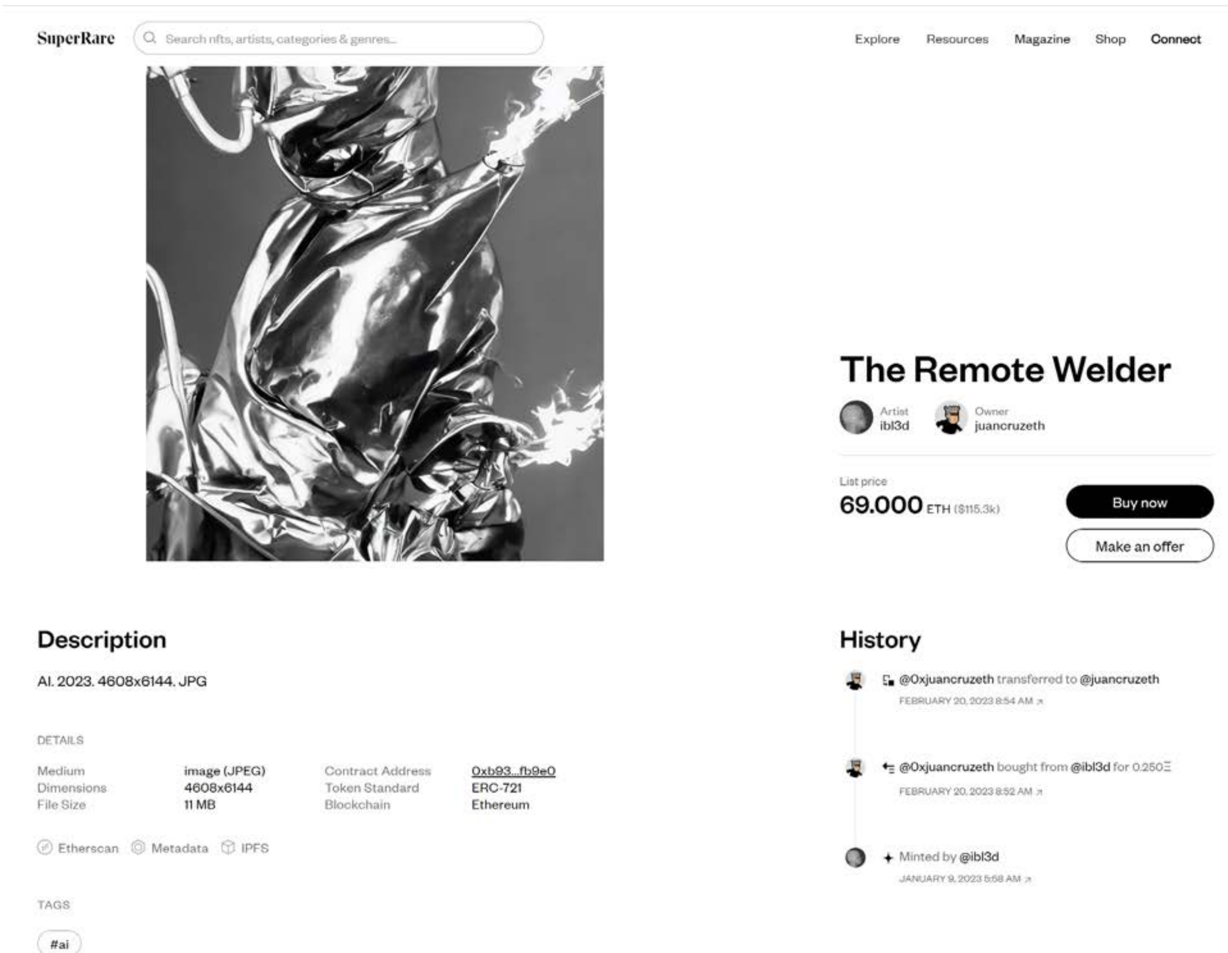


Figure 2. Sample an NFT auction.

This figure presents an NFT artwork named *The Frog King* that was listed on the SuperRare platform on February 21, 2023. Besides the same information as in Figure 1, the "History" part also shows a complete bidding history of an auction happened in November 2022.

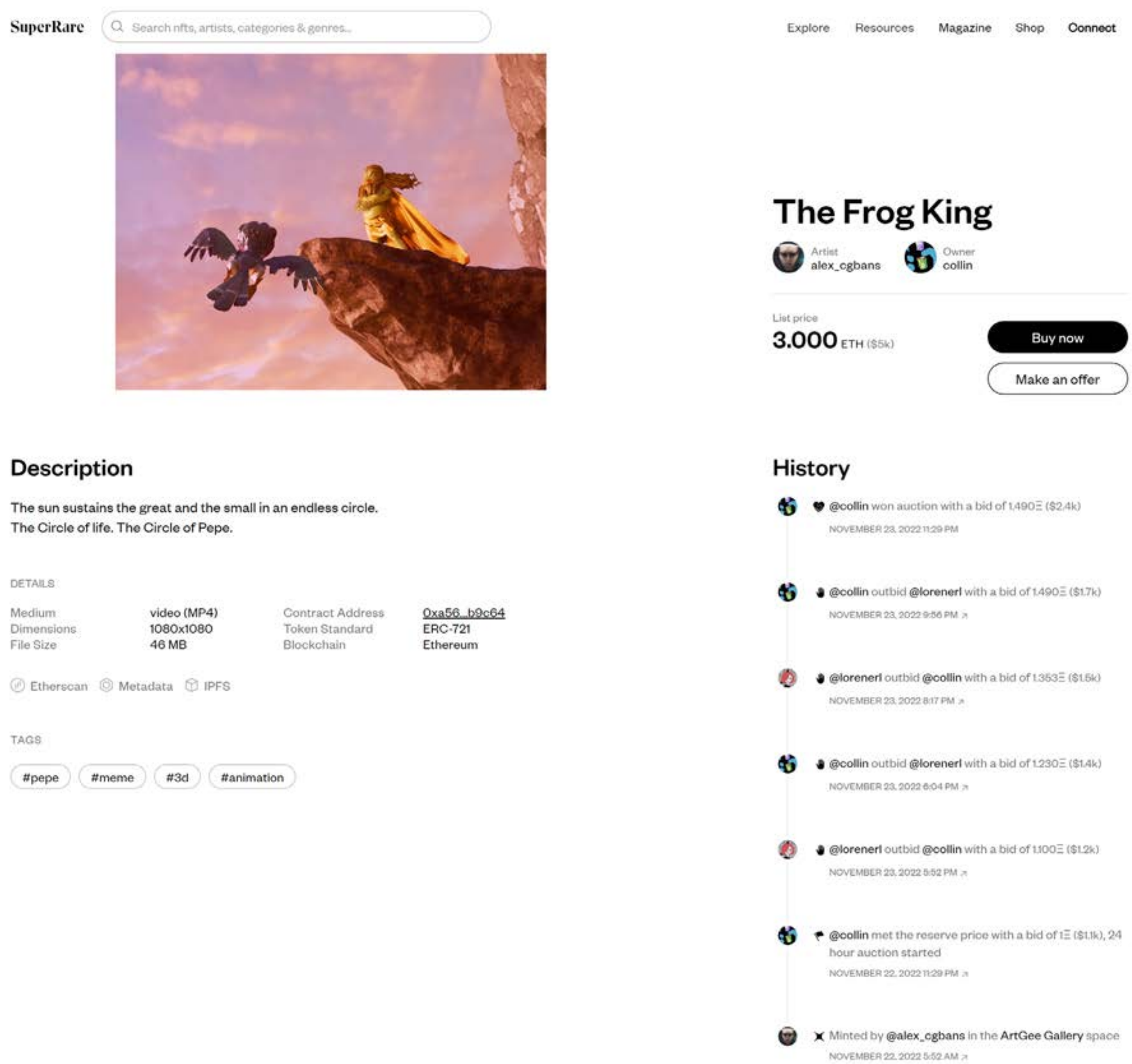
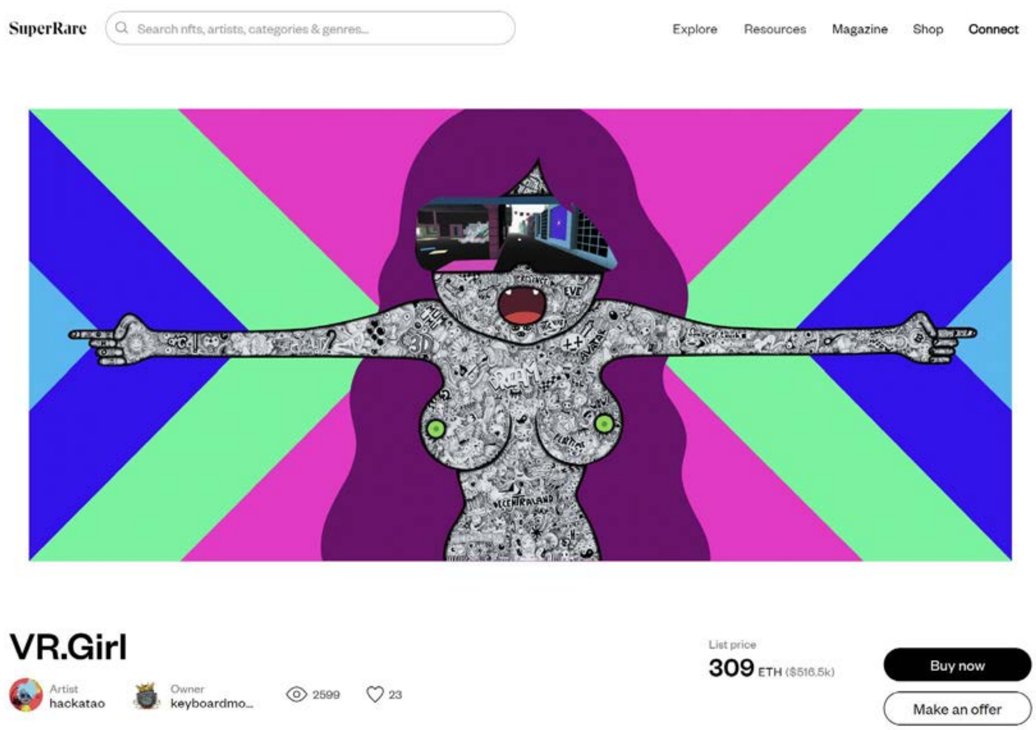


Figure 3. VR Girl by hackatao. Example of capital appreciation of an NFT over the period 12/2019- to 5/2021.

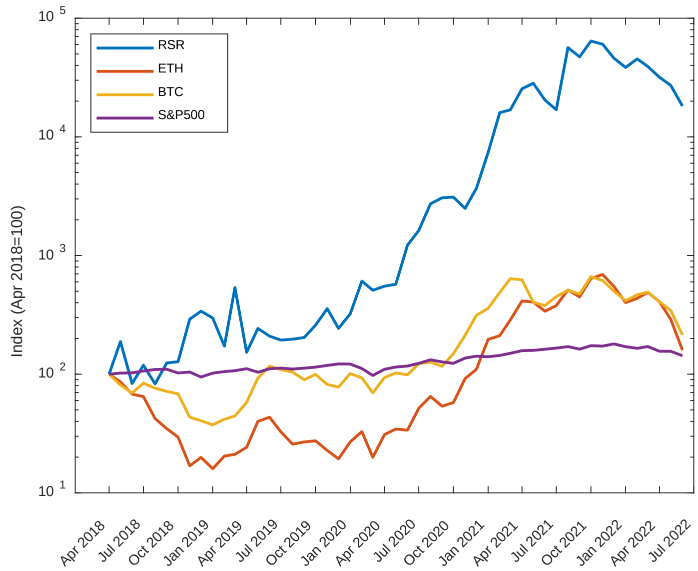


transferred
@6529fundszn3 transferred to @keyboardmonkey3
DECEMBER 24, 2022 6:22 AM
transferred
@keyboardmonkey3 transferred to @6529fundszn3
AUGUST 26, 2022 3:20 PM
transferred
@itsyaboy transferred to @keyboardmonkey3
MAY 25, 2021 1:12 PM
accept bid
@billywhistler accepted an offer of 115Ξ ($297.1k)from @itsyaboy
MAY 25, 2021 1:05 PM
bid
@itsyaboy made an offer of 115Ξ
MAY 25, 2021 1:02 PM
bid
@itsyaboy made an offer of 60Ξ
MAY 15, 2021 9:18 AM
bid
@mansamusa made an offer of 12.500Ξ
MARCH 3, 2021 3:12 PM
bid
@emeris made an offer of 24Ξ
SEPTEMBER 26, 2020 9:07 PM
bid
@jim made an offer of 12.350Ξ
DECEMBER 18, 2019 1:06 PM
accept bid
@hackatao accepted an offer of 12.200Ξ ($1.5k)from @billywhistler
DECEMBER 18, 2019 1:05 PM
bid
@billywhistler made an offer of 12.200Ξ
DECEMBER 18, 2019 1:03 PM
bid
@jim made an offer of 11.750Ξ
DECEMBER 18, 2019 1:02 PM
bid
@billywhistler made an offer of 11.500Ξ
DECEMBER 18, 2019 12:56 PM
bid
@jim made an offer of 11Ξ
DECEMBER 18, 2019 12:54 PM
bid
@billywhistler made an offer of 10.500Ξ
DECEMBER 18, 2019 12:46 PM
bid
@jim made an offer of 10.400Ξ
DECEMBER 18, 2019 10:28 AM
bid
@billywhistler made an offer of 10.200Ξ
DECEMBER 17, 2019 11:46 AM
bid
@jim made an offer of 10Ξ
DECEMBER 17, 2019 3:03 AM
bid
@billywhistler made an offer of 9.500Ξ
DECEMBER 17, 2019 2:41 AM
bid
@jim made an offer of 9Ξ
DECEMBER 16, 2019 7:37 PM
bid
@whaleshark made an offer of 8Ξ
DECEMBER 16, 2019 7:13 PM
bid
@billywhistler made an offer of 7Ξ
DECEMBER 16, 2019 2:06 PM
bid
@conlan made an offer of 6.500Ξ
DECEMBER 16, 2019 1:29 PM
bid
@moderatsart made an offer of 5.500Ξ
DECEMBER 16, 2019 12:53 PM
creation
Minted by @hackatao
DECEMBER 16, 2019 12:49 PM

Figure 4. NFT price index estimated from a repeat-sale regression

This figure presents the monthly price indexes of NFTs estimated from standard repeat-sale regressions (RSR). Figure (a) presents the RSR index (blue) estimated with the whole sample period from April 2018 to June 2022. Price in April 2018 is normalized to 100. The figure also plots the prices of Ethereum (ETH, red), Bitcoin (BTC, yellow), and S&P500 index (purple) within the same period, all normalized to 100 in April 2018. Figure (b) presents the RSR indexes estimated with repeat-sale pairs observed up-to each month, which are called the "up-to" indexes. For clarity, only indexes ending at quarters on and after 2019Q4 are shown here.

(a) RSR index estimated with the whole repeat-sale sample



(b) RSR indexes estimated with repeat-sale pairs observed up-to each month

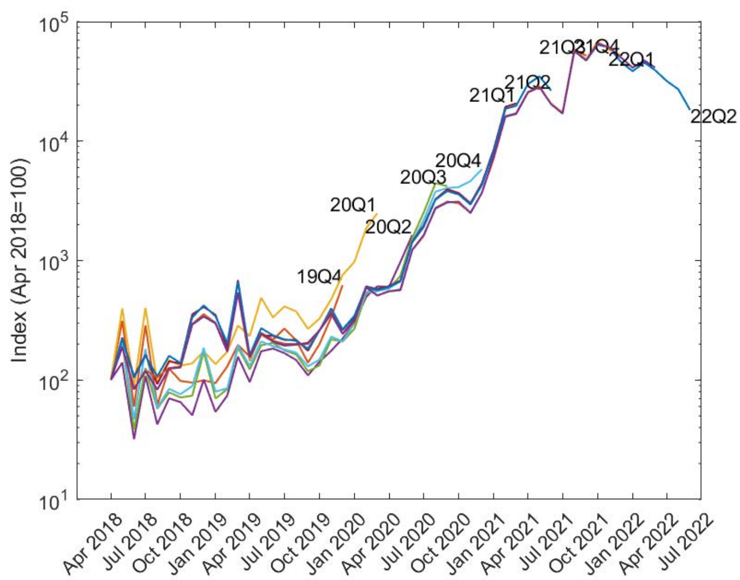
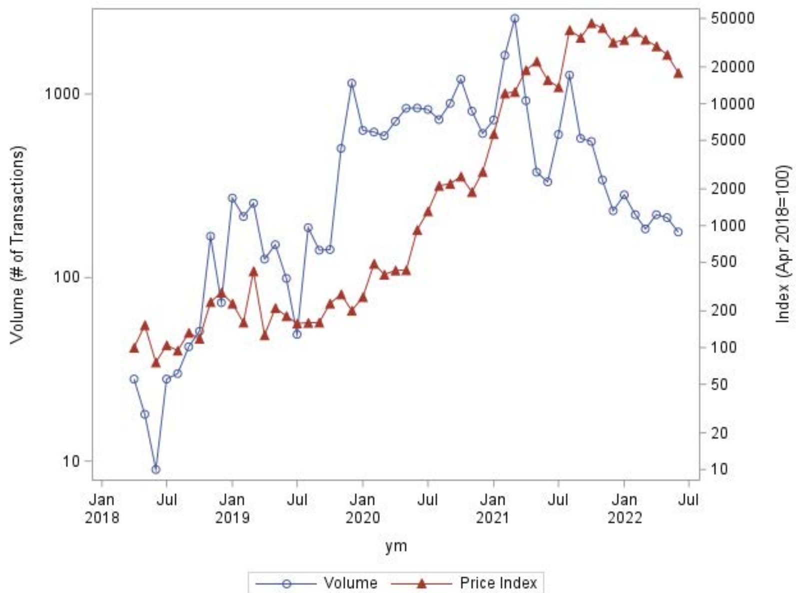


Figure 5. Dynamics between prices and volume

This figure illustrates the relationship between NFT prices and trading volume. Figure (a) plots the trading volume (blue line, left axis, log scale) and the RSR price index (red line, right axis, log scale) together. Figure (b) presents the correlation coefficients between log index prices p_t and log trading volume v_{t-k} of different leads ($k < 0$) or lags ($k > 0$).

(a) Co-movement between prices and market trading volume



(b) Correlations between log price p_t and lagged log volume v_{t-k}

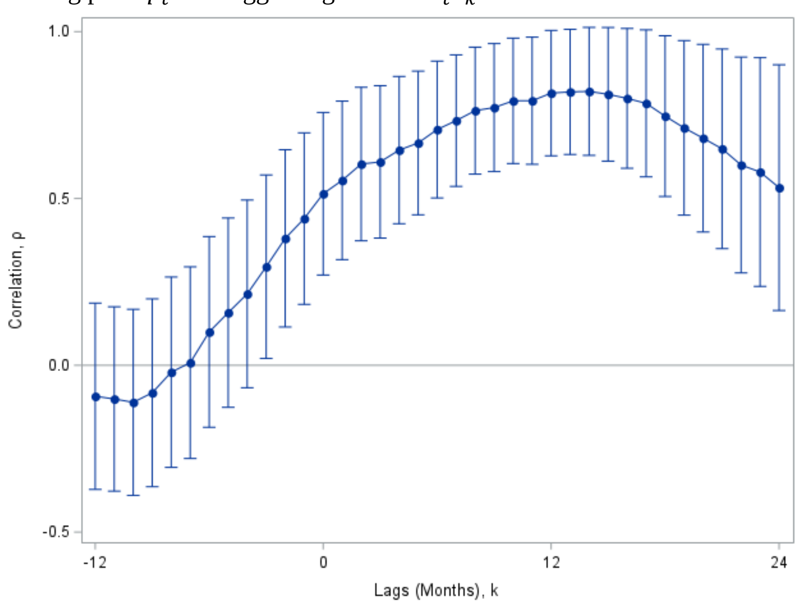
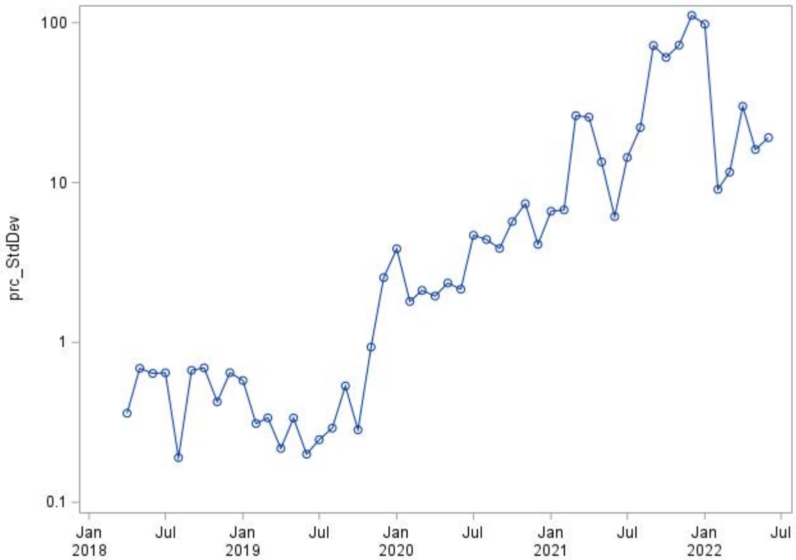


Figure 6. Price dispersion in the market

This figure presents the time-series of NFT price dispersion. Price dispersion in a month is defined as the cross-sectional standard deviation of observed transaction prices in that month. Figure (a) presents the dispersion of prices denominated in ETH, and Figure (b) in USD.

(a) ETH denominated



(b) USD denominated

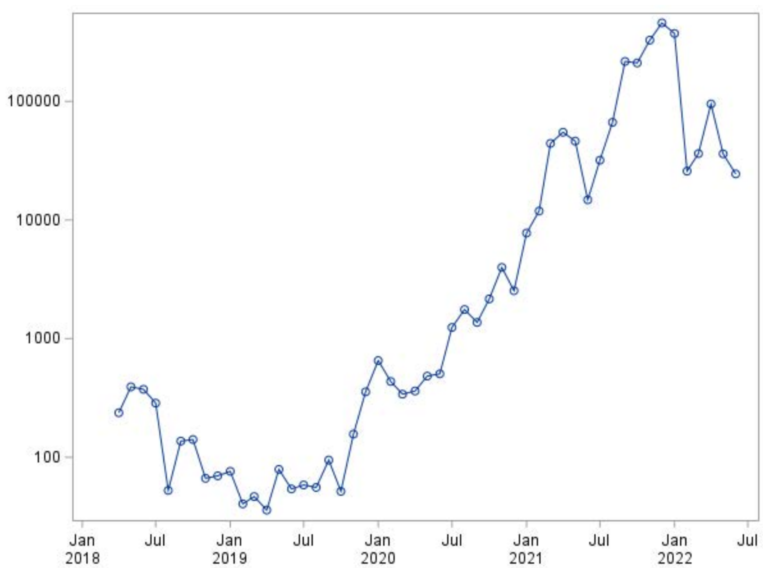
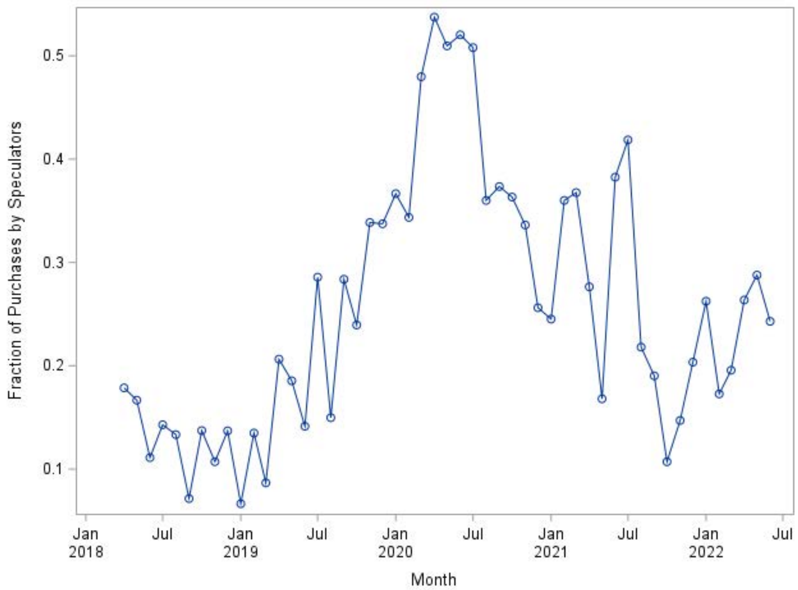


Figure 7. Trading behavior of short-term speculators

This figure presents evidence of short-term activities and speculators in the market. Speculative transactions are defined as "short-term" sales or listings of NFTs which happen within one month of the most recent purchases. Short-term speculators are then defined as the investors who conduct more than seven (the median) short-term speculative transactions during our sample period. Figure (a) presents the number of purchases by short-term speculators, as a fraction of total transactions in the market. Figure (b) shows in each month the fraction of NFTs that are purchased and then sold or listed within one month.

(a) Fraction of purchases by short-term speculators



(b) Fraction of purchased NFTs that are sold or listed within one month

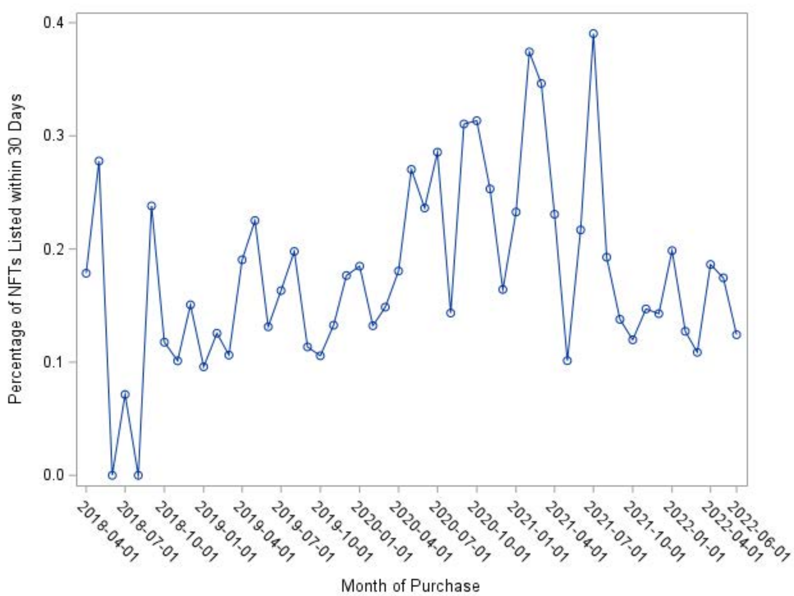


Figure 8. RSR price index with bids and asks and selection-adjusted price index from MCMC

This figure shows the RSR price indexes that account for selection bias in the observed transaction prices. The blue line is the standard RSR index. For the RSR index with bids and asks (the green line), the observed transaction prices are complemented with the mid-points of the latest bid and ask when the corresponding NFTs are untraded in the month. If there is only an ask from the seller or a bid from the potential buyers, that price is used instead. The red line shows the selection-adjusted price index using Markov Chain Monte Carlo algorithm following Kortweg et al. (2016).

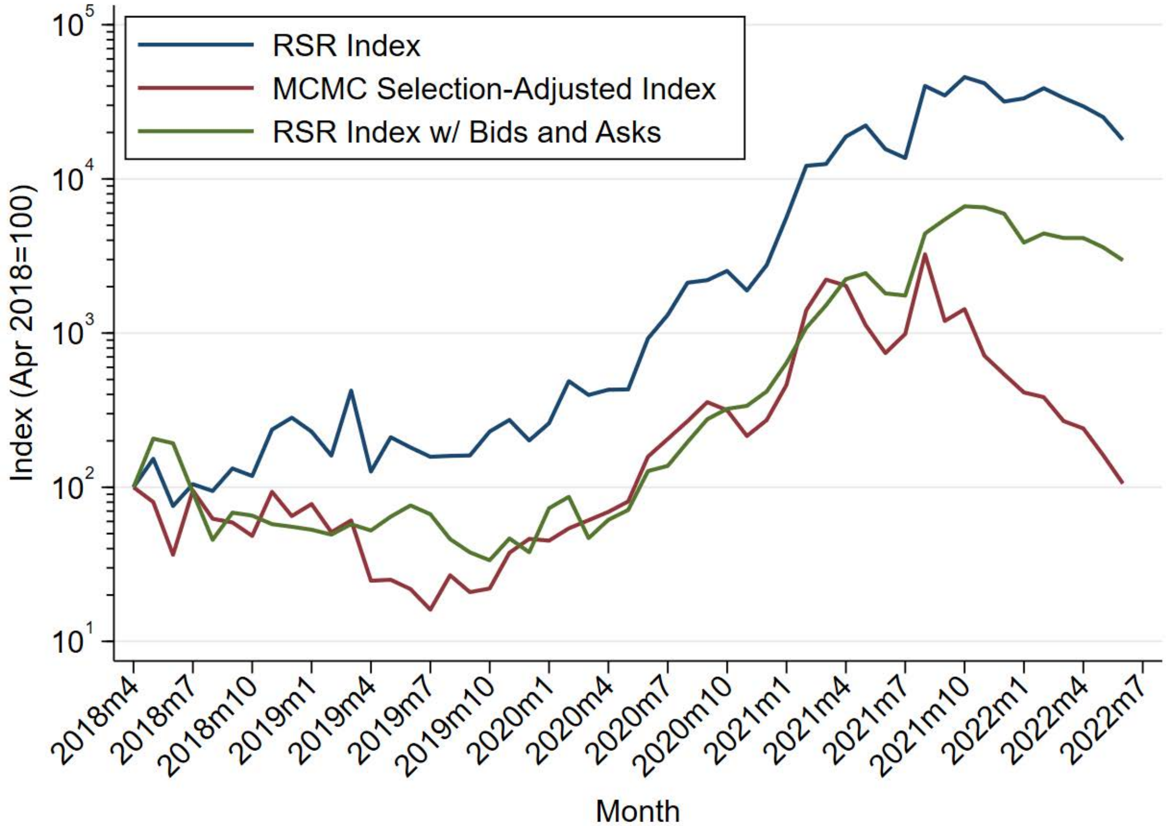


Figure 9. The probability of a sale

This figure plots the probability of a sale with respect to potential holding returns at different horizons, estimated by the MCMC algorithm.

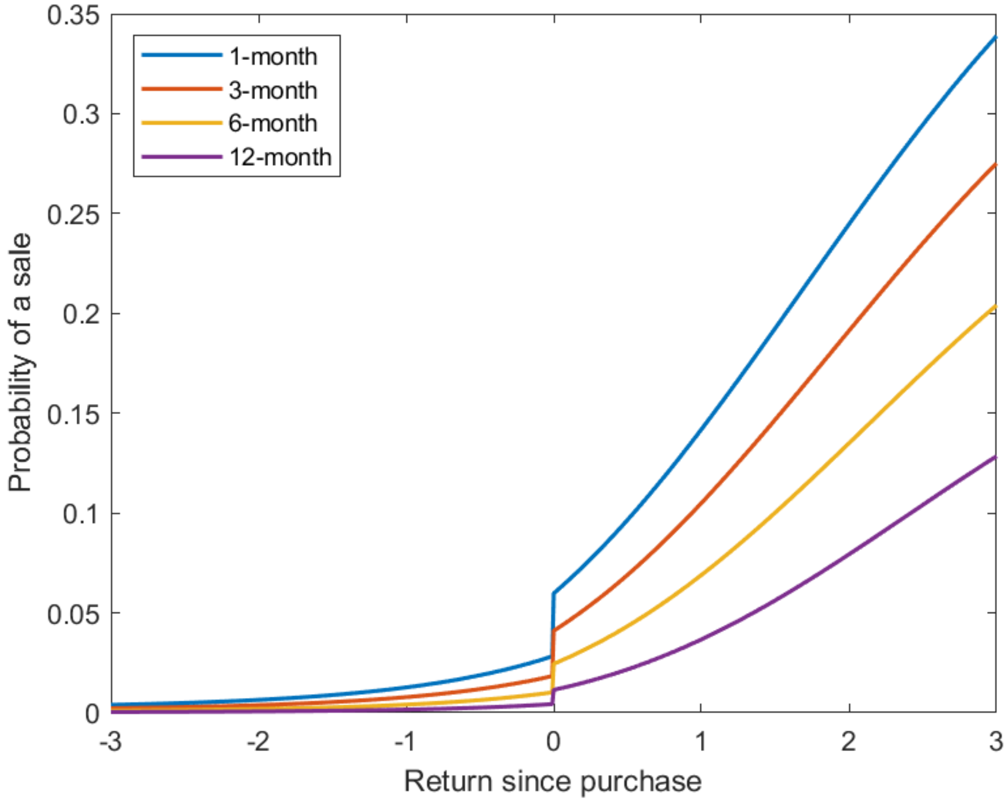
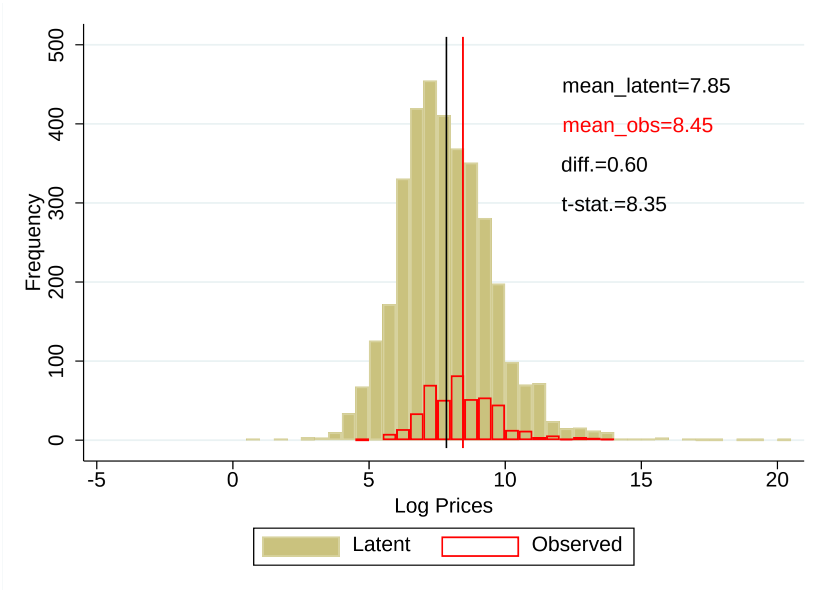


Figure 10. Distribution of latent and observed prices and returns

This figure shows the distribution of observed transaction prices and the latent prices of untraded NFTs estimated by the MCMC selection-adjustment procedure. Figure (a) shows a histogram of latent and observed prices in September 2021 as an example indicating the significant differences between observed and unobserved prices. Figure (b) plots the differences in average latent and observed prices over time.

(a) Histogram of latent and observed prices in September 2021



(b) Mean differences between observed and latent prices

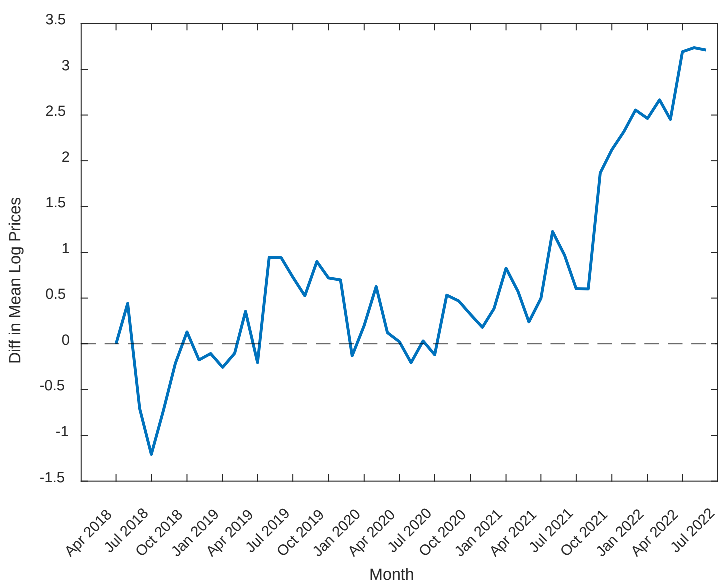


Figure 11. The effective federal funds rate and NFT price index

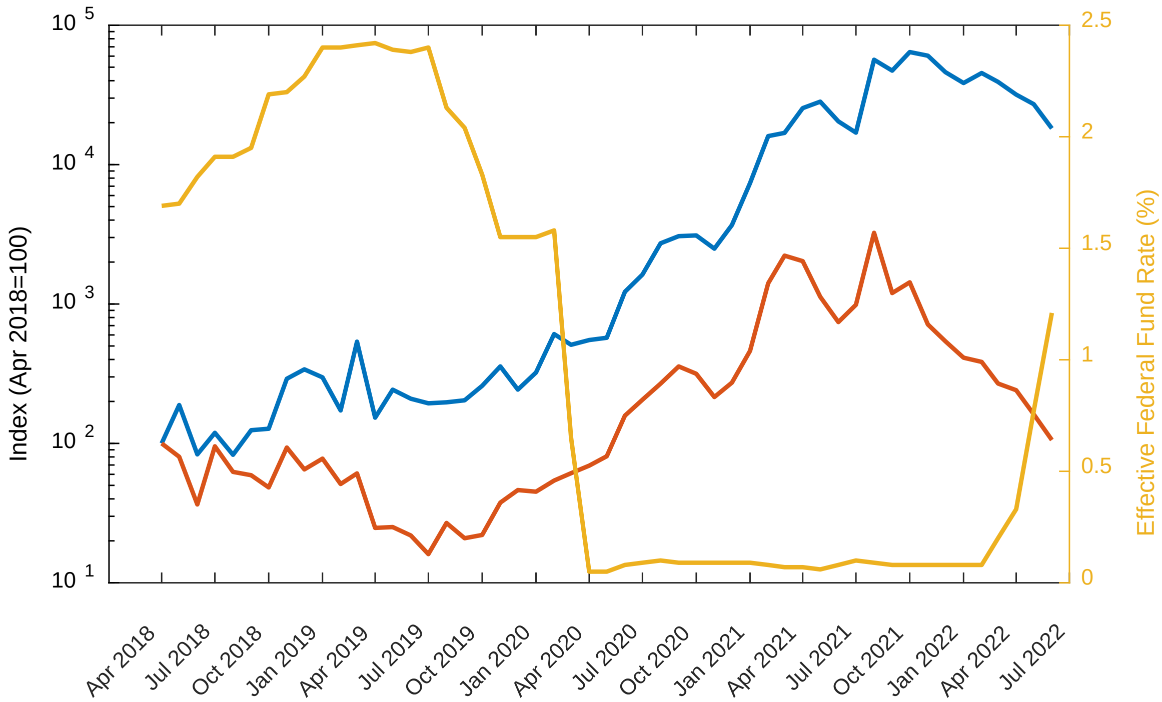


Figure 12. Selection bias as the difference between RSR index returns and MCMC-adjusted index returns

This figure shows the time-series of selection bias used for regressions in Section 6. The selection bias is defined as the difference between the six-month market returns implied by a standard RSR index (without an intercept) and an MCMC selection-adjusted index. The lag of six months is chosen based on the median holding period of investors.

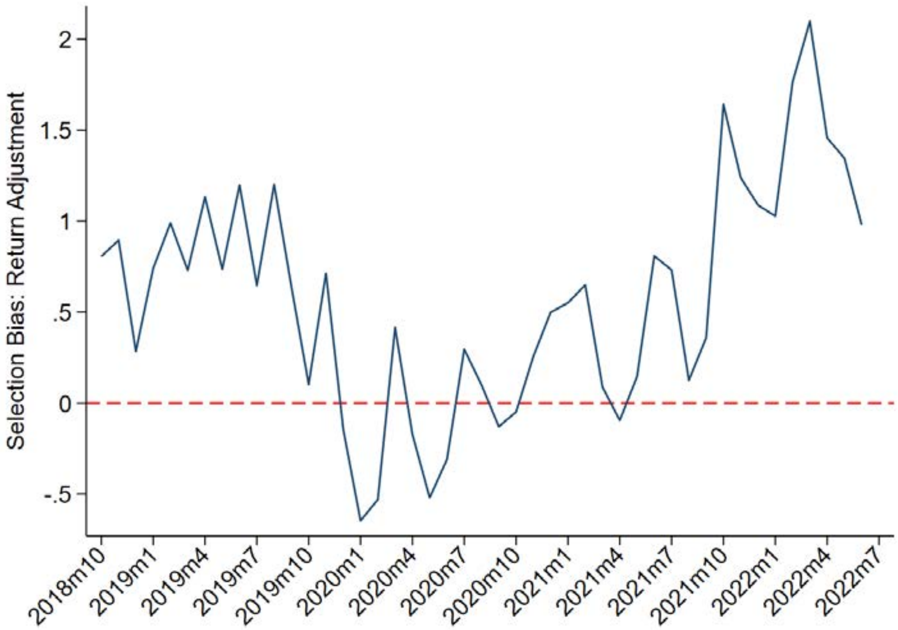
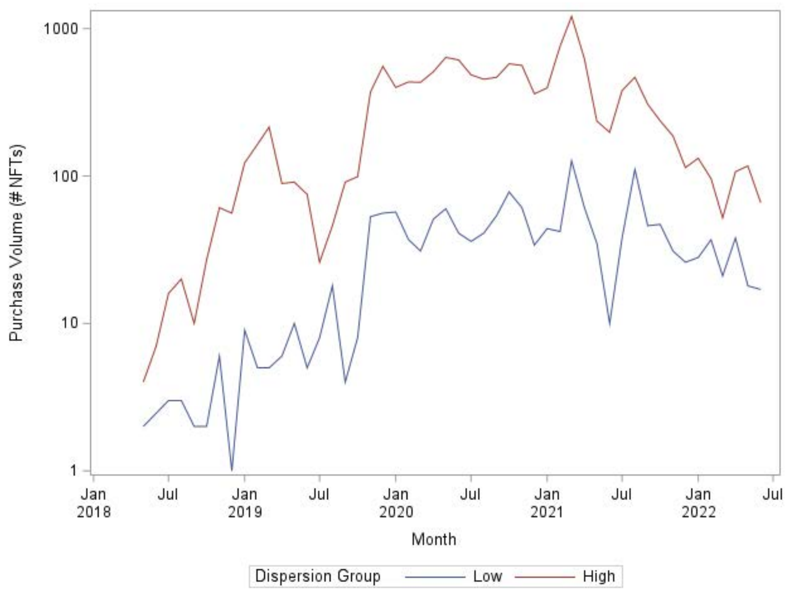


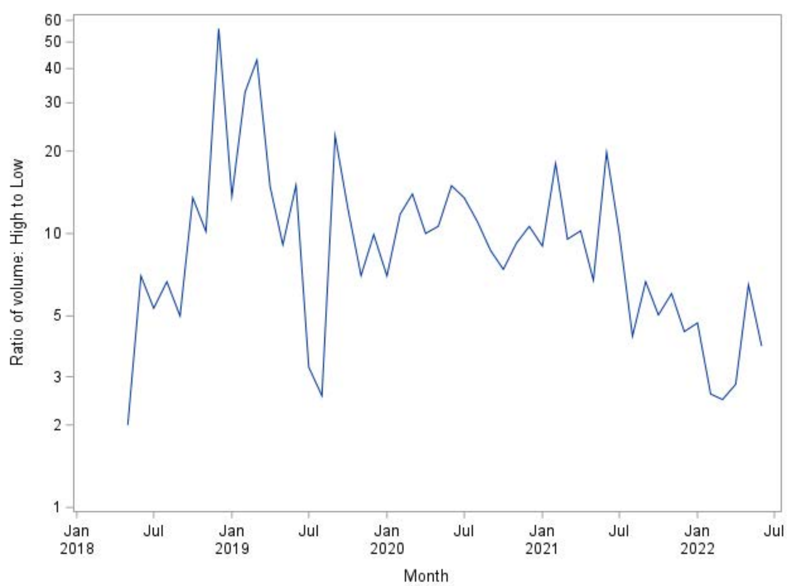
Figure 13. Purchase volume from high- and low-bid dispersion investors over time

This figure shows the purchase trading volume from high- and low-bid dispersion investors over time. Bid dispersion is defined as the range of her logarithmic bid prices scaled by their median in the previous three months for each investor in each month. Investors of high (low) bid dispersion in a month are those whose bid dispersion is above (below) the median within the month. Figure (a) plots the number of purchases by bid dispersion groups and figure (b) shows the ratio of the volume between the two groups.

(a) Purchase volume



(b) Ratio of purchase volume



Tables

Table 1. Summary Statistics

(a) Summary statistics of event prices – Prices in US dollars are converted based on the prices of ETH/USD at the end of the transaction date. The ask price refers to the list price (or the transaction price if the list price is not available) immediately after creation. List prices do not exist for all NFTs.

	P10	P25	P50	P75	P90	Mean	StdDev	Max	N
Ask (ETH)	0.39	0.96	2	5	10	41.93	1261.16	88888.89	14500
Ask (USD)	106.23	283.03	1228.99	5604.15	17068.48	34721.2	626489.5	50583875	14500
Bid (ETH)	0.07	0.2	0.55	1.75	4.24	2.4	15.78	1630	71165
Bid (USD)	23.41	76.45	384.39	1919.53	6783.39	4581.65	55225.49	6591705	71165
Sale (ETH)	0.15	0.3	0.8	2	5.1	3.47	25.87	1630	24424
Sale (USD)	33.94	100.5	586.03	2867.57	9561.22	7521.16	89778.89	6591705	24424

(b) Summary statistics of investors (as of June 30, 2022) – Artists are market participants who have ever created NFTs. Owners are investors who hold at least one NFT at the end of June 30, 2022.

	P10	P25	P50	P75	P90	Mean	StdDev	Max	N
Artists:									
# NFTs create	1	2	6	18	40	17.45	33.18	397	2018
# NFTs hold	0	0	1	3	6	2.91	8.36	143	2018
Owners:									
# NFTs hold	1	1	1	2	7	4.5	17.8	509	5510

(c) Summary statistics of repeat-sale pairs – Buy and sell prices have been converted to US dollars based on the prices of ETH/USD at the end of the transaction date. Log returns are the log differences between sell and buy prices in USD. For “Buy date” and “Sell date”, the sample period (April 1, 2018 to June 30, 2022) is normalized to [0, 1].

	Mean	Std Dev	Min	P25	Median	P75	Max	N
Sell price (USD)	11981.94	155297.84	2.61	313.96	1090.57	3805.40	6591705.47	4281
Buy price (USD)	2251.54	45909.43	0.04	37.29	112.93	606.08	2925565.67	4281
Log return	2.01	1.52	-1.39	0.85	1.75	2.99	6.64	4281
Holding period (in months)	8.52	7.56	1.00	3.00	6.00	12.00	40.00	4281
Buy date	0.50	0.18	0.00	0.40	0.52	0.64	0.97	4281
Sell date	0.67	0.14	0.06	0.59	0.69	0.77	1.00	4281

Table 2. Regressions of asking prices on potential NFT returns

This table reports the estimations of equation (7), showing the relationship between sellers' reservation prices and potential gains or losses. For columns 1-3, the dependent variable, $AskPrice_{i,t}$, are the listing prices of NFTs and used as a proxy for sellers' reservation prices. $PredictPrice_{i,t}$ is the estimated true value of NFTs estimated by standard repeat-sale index returns. $PotentialRet_{i,t}$ is the market index return within the holding period. T is the length of holding period. Columns 4-6 use $AskMarkUp \equiv Log(AskPrice_{i,t}) - Log(PredictPrice_{i,t})$ as the dependent variables. Standard errors clustered by calendar months are reported in parentheses. Significance level: *: p<0.1, **: p<0.05, ***: p<0.01.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	$Log(AskPrice_{i,t})$			$AskMarkUp_{i,t}$		
$Log(PredictPrice_{i,t})$	0.438***	0.436***	0.443***			
	(0.0210)	(0.0198)	(0.0211)			
$1(PotentialRet_{i,t} < 0)$	0.275***	0.232***	0.269***	0.317***	0.332***	0.339***
	(0.0766)	(0.0840)	(0.0839)	(0.0788)	(0.0976)	(0.0959)
$PotentialRet^{+}_{i,t}$	-0.217***	-0.454***	-0.354***	-0.0625***	-0.127**	-0.124
	(0.0219)	(0.0439)	(0.0902)	(0.0164)	(0.0552)	(0.110)
${PotentialRet^{+}_{i,t}}^2$		0.0602***	0.0291*		0.0165	0.00339
		(0.0101)	(0.0167)		(0.0125)	(0.0150)
$PotentialRet^{-}_{i,t}$	-0.267	0.0153	0.0131	-0.549**	-0.251	-0.260
	(0.274)	(0.396)	(0.420)	(0.252)	(0.449)	(0.444)
${PotentialRet^{-}_{i,t}}^2$		0.369	0.385		0.349	0.338
		(0.447)	(0.436)		(0.408)	(0.408)
T			-0.0260			-3.89e-05
			(0.0218)			(0.0237)
T^2			0.00157*			0.000620
			(0.000813)			(0.000751)
Constant	1.337***	1.424***	1.432***	1.562***	1.585***	1.583***
	(0.0178)	(0.0241)	(0.0272)	(0.0175)	(0.0279)	(0.0289)
Observations	20,104	20,104	20,104	19,984	19,984	19,984
R-squared	0.569	0.572	0.573	0.217	0.217	0.218
Artist FEs	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Month FEs	Yes	Yes	Yes	Yes	Yes	Yes

Table 3. The predictability of past returns on trading volume

This table reports the results for regressions of monthly log trading volumes on past six-month market log returns. Market returns are calculated based on the price indexes estimated by standard RSR, MCMC selection-adjustment algorithm, or "up-to" indexes as described in the text. Besides regressing on past returns, we also regress log volume on the positive and negative parts of the returns. Panel A and B use market trading volumes and those from new buyers as the dependent variables, respectively. Newey-West autocorrelation adjusted standard errors are reported in parentheses. Significance level: *: p<0.1, **: p<0.05, ***: p<0.01.

Panel A. Dependent variable: market trading volume (in log)

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: $Log(Volume_t)$					
Return Estimated from Index:	RSR	RSR	MCMC	MCMC	Up-to	Up-to
Past Log Return	0.715***		0.600***		0.422**	
	(0.160)		(0.100)		(0.156)	
Past Log Return^{+}		1.223***		1.468***		0.991***
		(0.395)		(0.452)		(0.310)
Past Log Return^{-}		-0.422		-0.319		-1.447
		(0.920)		(0.435)		(0.992)
Constant	5.369***	4.671***	5.802***	4.726***	5.539***	4.605***
	(0.236)	(0.517)	(0.144)	(0.558)	(0.205)	(0.453)
Observations	45	45	45	45	39	39
R-squared	0.365	0.426	0.525	0.470	0.133	0.386

Panel B. Dependent variable: purchase trading volume from new buyers (in log)

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: $Log(Volume_t)$ from new buyers					
Return Estimated from Index:	RSR	RSR	MCMC	MCMC	Up-to	Up-to
Past Log Return	0.739***		0.587***		0.452**	
	(0.189)		(0.107)		(0.183)	
Past Log Return^{+}		1.146***		1.399***		0.909***
		(0.374)		(0.403)		(0.272)
Past Log Return^{-}		-0.202		-0.283		-0.842
		(0.779)		(0.371)		(0.839)
Constant	4.091***	3.534***	4.545***	3.568***	4.243***	3.491***
	(0.225)	(0.445)	(0.156)	(0.476)	(0.238)	(0.379)
Observations	45	45	45	45	39	39
R-squared	0.306	0.376	0.395	0.423	0.117	0.323

Table 4. The predictability of realized returns on trading volume

This table reports the results for regressions of monthly log trading volumes on the mean, maximum, and minimum of past six-month realized log returns. For columns 1-3, realized returns are not adjusted for the holding horizons while in column 4, realized returns are scaled to monthly returns. Newey-West autocorrelation adjusted standard errors are reported in parentheses. Significance level: *: p<0.1, **: p<0.05, ***: p<0.01.

	(1)	(2)	(3)	(4)
VARIABLES	Log(Volume)	Log(Volume)	Log(Volume)	Log(Volume)
Realized Return Mean	0.597**			
	(0.230)			
Realized Return Max		0.346***		
		(0.105)		
Realized Return Min			-0.495	
			(0.604)	
Monthly Realized Return Mean				4.649***
				(1.377)
Constant	4.967***	3.982***	5.318***	4.397***
	(0.399)	(0.590)	(0.745)	(0.515)
Observations	47	47	47	42
R-squared	0.149	0.304	0.044	0.213

Table 5. Interaction between extrapolation and selection bias

This table reports regression results for equation (9), testing the interactions between extrapolation and selection bias. $LogUnbiasedReturn_{t-1}$ are the past six-month market returns calculated with the MCMC selection-adjusted index. $SelectionBias_{t-1}$ is defined as the difference between the standard RSR index return and the MCMC index return, both in the six-month horizon. Columns 4 to 6 repeat the regressions in the first three columns, respectively, using $SelectionBias_{t-1}$ that is defined as the difference between the "up-to" RSR index return and the MCMC index return, both in the six-month horizon. Newey-West autocorrelation adjusted standard errors are reported in parentheses. Significance level: *: p<0.1, **: p<0.05, ***: p<0.01.

	(1)	(2)	(3)	(4)	(5)	(6)
Selection bias calculated with:	MCMC index returns			"Up-to" RSR index returns		
	Dependent Variable: $Log(Volume_t)$					
$LogUnbiasedReturn_{t-1}$	0.600***	0.783***		0.600***	0.707***	
	(0.100)	(0.114)		(0.100)	(0.129)	
$LogUnbiasedReturn^{+}_{t-1}$			0.900***			0.769***
			(0.240)			(0.254)
$LogUnbiasedReturn^{-}_{t-1}$			0.619**			0.639***
			(0.253)			(0.221)
$SelectionBias_{t-1}$		-0.564**	-0.560**		-0.558***	-0.542***
		(0.213)	(0.230)		(0.119)	(0.0978)
$LogUnbiasedReturn^{+}_{t-1} \times SelectionBias_{t-1}$		0.736**	0.776**		0.851***	0.869***
		(0.305)	(0.314)		(0.146)	(0.181)
$LogUnbiasedReturn^{-}_{t-1} \times SelectionBias_{t-1}$		-0.534***	-0.471***		-0.354**	-0.322***
		(0.175)	(0.160)		(0.135)	(0.107)
Constant	5.802***	5.906***	5.814***	5.802***	6.134***	6.085***
	(0.144)	(0.166)	(0.284)	(0.144)	(0.119)	(0.185)
Observations	45	44	44	45	38	38
R-squared	0.525	0.639	0.643	0.525	0.745	0.746

Table 6. Interaction between extrapolation and selection bias: high- versus low-dispersion investors

This table tests the interactions between extrapolation and selection bias by investors of different bid price dispersion. In each investor-month, the bid price dispersion is defined as the range of the investor's bid prices in the previous three months scaled by the median bid price. The first three columns use log purchase volume from low-dispersion investors as the dependent variables while the last three columns from high-dispersion investors. $LogUnbiasedReturn_{t-1}$ are the past six-month market returns calculated with the MCMC selection-adjusted index. $SelectionBias_{t-1}$ is defined as the difference between the standard RSR index return and the MCMC index return, both in the six-month horizon, and is normalized. Newey-West autocorrelation adjusted standard errors are reported in parentheses. Significance level: *: *p*<0.1, **: *p*<0.05, ***: *p*<0.01.

	(1)	(2)	(3)	(4)	(5)	(6)
Investor type:		Low dispersion			High dispersion	
	Dependent Variable: $Log(Volume_t)$					
$LogUnbiasedReturn_{t-1}$	0.442***	0.666***		0.688***	0.794***	
	(0.144)	(0.126)		(0.0888)	(0.100)	
$LogUnbiasedReturn^{+}_{t-1}$			0.930**			0.706***
			(0.379)			(0.202)
$LogUnbiasedReturn^{-}_{t-1}$			0.298			0.916***
			(0.411)			(0.209)
$SelectionBias_{t-1}$		-0.523	-0.515		-0.545**	-0.548**
		(0.366)	(0.408)		(0.218)	(0.205)
$LogUnbiasedReturn^{+}_{t-1} \times SelectionBias_{t-1}$		0.562	0.653		0.648**	0.617**
		(0.377)	(0.413)		(0.280)	(0.275)
$LogUnbiasedReturn^{-}_{t-1} \times SelectionBias_{t-1}$		-0.699**	-0.557*		-0.432**	-0.479***
		(0.315)	(0.277)		(0.166)	(0.150)
Constant	3.118***	3.080***	2.874***	5.208***	5.329***	5.397***
	(0.236)	(0.291)	(0.515)	(0.127)	(0.139)	(0.242)
Observations	45	44	44	45	38	38
R-squared	0.251	0.376	0.394	0.636	0.726	0.728

Table 7. Interactions between extrapolation and selection bias: Individual-level regressions

This table tests the interactions between extrapolation and selection bias using transaction data at the investor-month levels. In each investor-month, the bid price dispersion $BidDispr_{i,t-1}$ is defined as the range of the investor's bid prices in the previous three months scaled by the median bid price. The first three columns use log purchase volume as the dependent variables while the last three columns use an indicator of positive purchases and are estimated with conditional logit regressions. $LogUnbiasedReturn_{t-1}$ are the past six-month market returns calculated with the MCMC selection-adjusted index. $SelectionBias_{t-1}$ is defined as the difference between the standard RSR index return and the MCMC index return, both in the six-month horizon, and is normalized. Cluster robust standard errors clustered by investors are reported in parentheses. Significance level: *: *p*<0.1, **: *p*<0.05, ***: *p*<0.01.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	$Log(Volume_{i,t})$			$1(Volume_{i,t} > 0)$		
$LogUnbiasedReturn_{t-1}$	0.0612***			0.388***		
	(0.00603)			(0.0321)		
$LogUnbiasedReturn^{+}_{t-1}$		0.0266***	0.0190		0.116**	0.236***
		(0.00854)	(0.0118)		(0.0504)	(0.0647)
$LogUnbiasedReturn^{-}_{t-1}$		0.111***	0.154***		0.865***	0.714***
		(0.0100)	(0.0167)		(0.0773)	(0.122)
$SelectionBias_{t-1}$			-0.0874***			-0.604***
			(0.00880)			(0.0646)
$LogUnbiasedReturn^{+}_{t-1} \times SelectionBias_{t-1}$			0.0728***			0.569***
			(0.00655)			(0.0604)
$LogUnbiasedReturn^{-}_{t-1} \times SelectionBias_{t-1}$			-0.0624***			-0.248***
			(0.00856)			(0.0761)
Investor fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	38,246	38,426	38,188	12,986	13,075	12,950

Table 8. Interactions between extrapolation and selection bias at the investor-level: heterogeneity over investor ability

This table tests the interactions between extrapolation and selection bias using transaction data at the investor-month levels and how the effects vary over investors' ability to correct for selection bias. Investor ability, $Ability$, is proxied by either bid price dispersion ($BidDispr$, the first two columns) or the fraction of NFTs purchased below asking prices ($BuyLow$, the last two columns). The odd columns use log purchase volume as the dependent variables while the even columns use an indicator of positive purchases and are estimated with conditional logit regressions. $LogUnbiasedReturn_{t-1}$ are the past six-month market returns calculated with the MCMC selection-adjusted index. $SelectionBias_{t-1}$ is defined as the difference between the standard RSR index return and the MCMC index return, both in the six-month horizon, and is normalized. Cluster robust standard errors clustered by investors are reported in parentheses. Significance level: *: p<0.1, **: p<0.05, ***: p<0.01.

	(1)	(2)	(3)	(4)
Proxy for investor ability:	$BidDispr$		$BuyLow$	
Dependent variable:	$Log(Buy_{i,t}+1)$	$1(Buy_{i,t}>0)$	$Log(Buy_{i,t}+1)$	$1(Buy_{i,t}>0)$
$LogUnbiasedReturn^{+}_{t-1}$	0.0216*	0.203***	0.00309	0.254***
	(0.0116)	(0.0691)	(0.0180)	(0.0878)
$LogUnbiasedReturn^{-}_{t-1}$	0.141***	0.744***	0.155***	1.088***
	(0.0159)	(0.130)	(0.0190)	(0.145)
$SelectionBias_{t-1}$	-0.0809***	-0.578***	-0.147***	-0.927***
	(0.00830)	(0.0675)	(0.0145)	(0.0848)
$LogUnbiasedReturn^{+}_{t-1} \times SelectionBias_{t-1}$	0.0726***	0.553***	0.0974***	0.710***
	(0.00654)	(0.0625)	(0.0117)	(0.0784)
$LogUnbiasedReturn^{-}_{t-1} \times SelectionBias_{t-1}$	-0.0457***	-0.255***	-0.0954***	-0.454***
	(0.00887)	(0.0786)	(0.0123)	(0.0940)
$Ability_{i,t-1}$	0.0669***	0.433	0.152	0.0831
	(0.0244)	(0.266)	(0.0974)	(0.298)
$LogUnbiasedReturn^{+}_{t-1} \times Ability_{i,t-1}$	0.0558**	0.484*	-0.0540***	-0.461***
	(0.0230)	(0.252)	(0.0147)	(0.101)
$LogUnbiasedReturn^{-}_{t-1} \times Ability_{i,t-1}$	-0.0232	-0.588*	-0.00662	-0.402**
	(0.0316)	(0.321)	(0.0167)	(0.156)
$SelectionBias_{t-1} \times Ability_{i,t-1}$	-0.00737	0.0251	-0.0206*	0.0104
	(0.0251)	(0.208)	(0.0124)	(0.0998)
$LogUnbiasedReturn^{+}_{t-1} \times SelectionBias_{t-1} \times Ability_{i,t-1}$	0.0260	0.344	-0.0137	-0.219**
	(0.0217)	(0.217)	(0.0101)	(0.0929)
$LogUnbiasedReturn^{-}_{t-1} \times SelectionBias_{t-1} \times Ability_{i,t-1}$	0.118***	0.631**	0.00300	0.252**
	(0.0459)	(0.313)	(0.0112)	(0.106)
Investor fixed effects	Yes	Yes	Yes	Yes
Observations	38,188	12,950	15,327	10,979

Table 9. Performance of sophisticated investors

This table reports the relationship between NFT investment returns and investors' characteristics using repeat-sale regressions (RSR). The regression equation is,

$$r_{i,b\to s} = \sum_{t=b_i+1}^{s_i} \mu_t + \gamma(s_i - b_i) \times X_{i,b} + \varepsilon_{i,b\to s},$$

where b and s are the time of purchase and sale, respectively. In addition to time dummies μ_t's in standard RSR, investors' characteristics at the time of purchase, $X_{i,b}$, interacted with the length of holding period $(s - b)$ are also included in the regressions. Investor characteristics include bid price dispersion ($BidDispr$), fraction of NFTs purchased below asking prices ($BuyLow$), cumulative number of NFTs bought (in log, $Log(\#\ buy)$), and the average length of holding periods. For the first two characteristics about investor ability, an indicator for missing values among inactive or new investors are also included. For conciseness, only the coefficients on the interaction terms are reported in the table. Panel A uses completed repeat-sale pairs as in standard RSR and hence the dependent variable $r_{i,b\to s}$ includes only realized returns. Panel B also includes unrealized returns on NFTs that remain unsold in investors' inventory by the end of our sample period. The unrealized returns are proxied by the MCMC selection-corrected returns between the time of purchase and the end of sample period. Standard errors are reported in parentheses. Significance level: *: p<0.1, **: p<0.05, ***: p<0.01.

Panel A. Realized returns (in percentage)

	(1)	(2)	(3)	(4)
BidDisper	0.452***			
	(0.04)			
Missing BidDisper Dummy	0.854**			
	(0.30)			
BuyLow		7.107***		
		(0.66)		
Missing BuyLow Dummy		3.634***		
		(0.46)		
Log(# buy)			-0.278***	
			(0.08)	
Avg Hold Time				-0.439***
				(0.02)
Observations	4,281	4,281	4,281	4,281
R-Squared	0.82	0.82	0.82	0.83

Panel B. Realized and unrealized returns (in percentage)

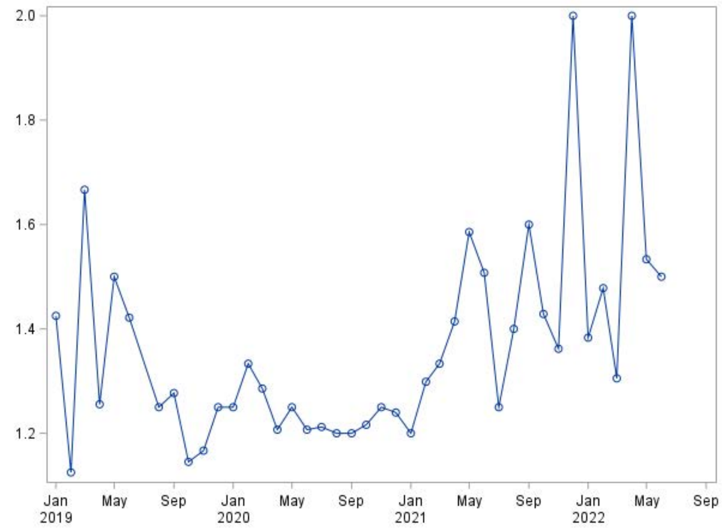
	(1)	(2)	(3)	(4)
BidDisper	0.072***			
	(0.009)			
Missing BidDisper Dummy	0.009			
	(0.04)			
BuyLow		0.204***		

		(0.07)		
Missing BuyLow Dummy		0.079		
		(0.05)		
Log(# buy)			-0.007	
			(0.009)	
Avg Hold Time				-0.067***
				(0.003)
Observations	23,610	23,610	23,610	23,610
R-Squared	0.94	0.94	0.94	0.94

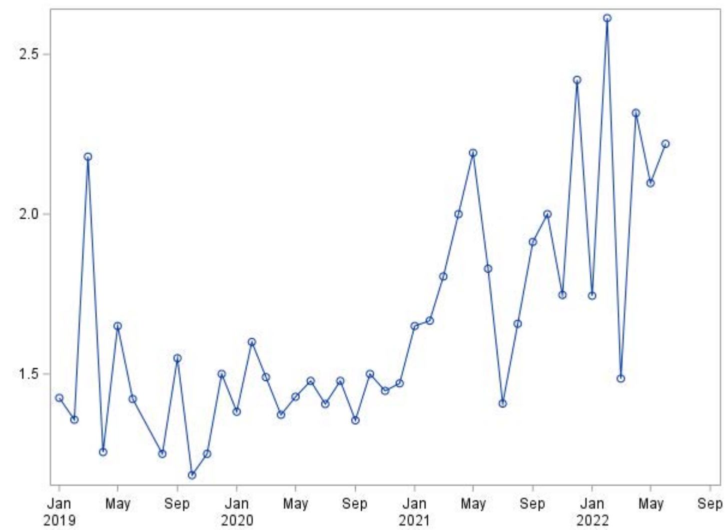
Appendix

Appendix Figure 1. Dispersion in prices offered to the same NFT

(a) The ratio of maximum (the winning bid) to second largest offer

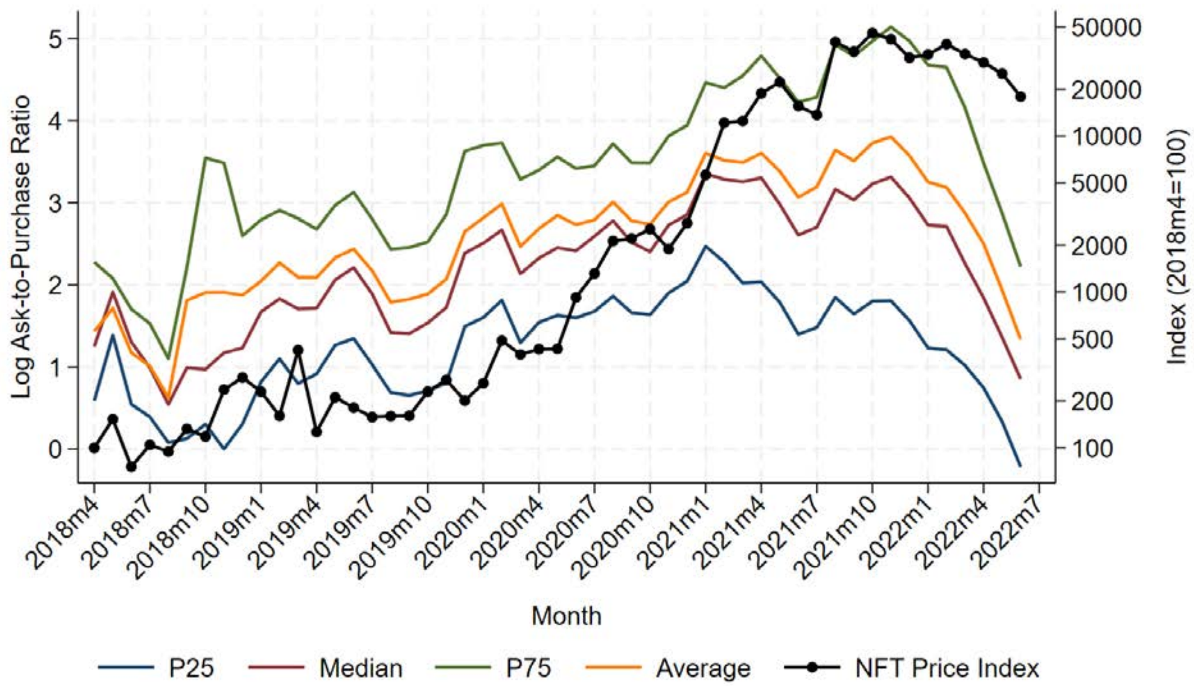


(b) The ratio of maximum (the winning bid) to median offer



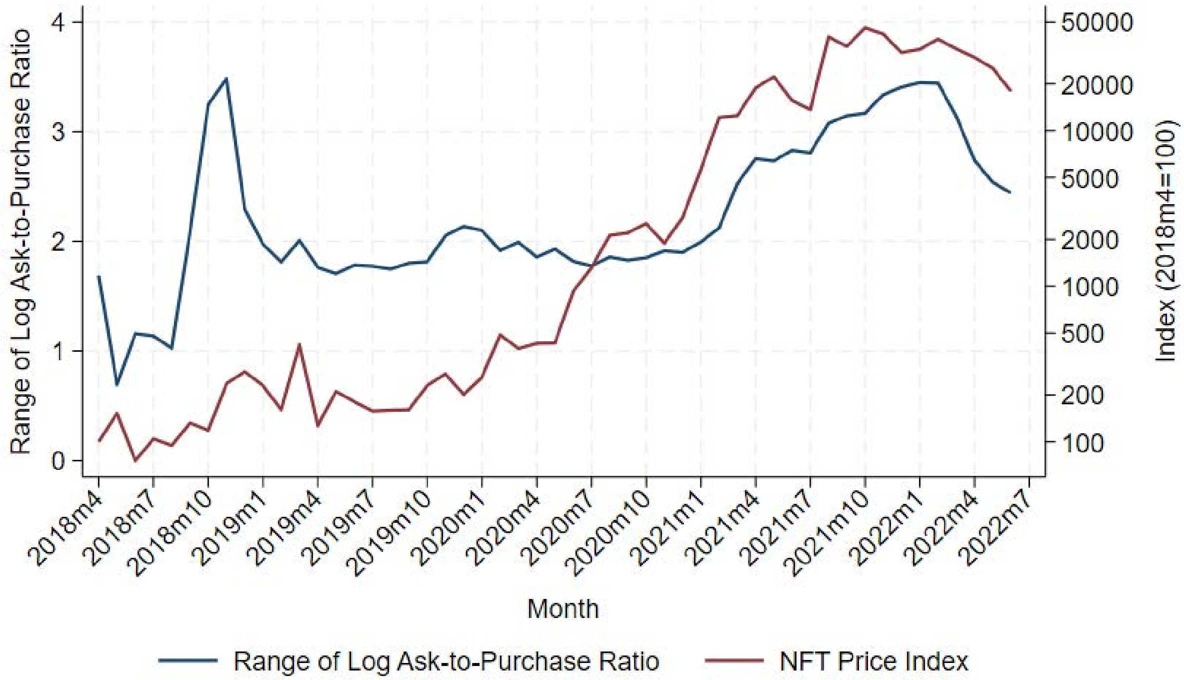
A larger gap between the winning bid and the second largest (or median) bid can be a sign of dispersed beliefs. But the seller's reservation price and her propensity to accept the bid may also matter since this is not an auction setting and the bidders can withdraw their bids any time as long as they are not accepted. Some sellers may keep waiting for higher bids while others may accept the bid as soon as it is higher than her reservation price.

Appendix Figure 2. The mean, median, and quantiles of the ratio of asking price to purchase price



A rise in ask-to-purchase price ratio during the boom, indicating more optimistic valuation.

Appendix Figure 3. The interquartile range of the ask-to-purchase price ratio

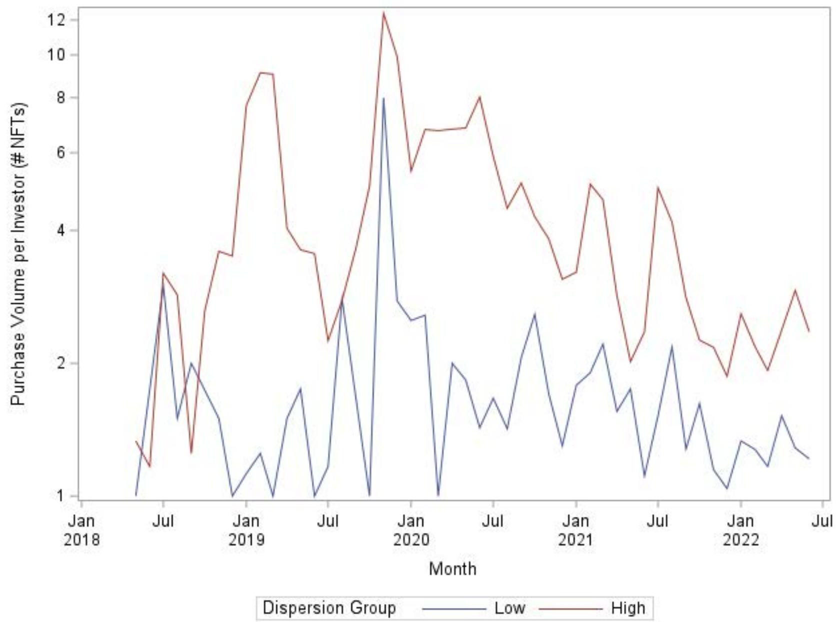


More dispersed expectations, as measured by the ask-to-purchase price ratio, during the boom. Investor expectations converged as the bubble collapsed.

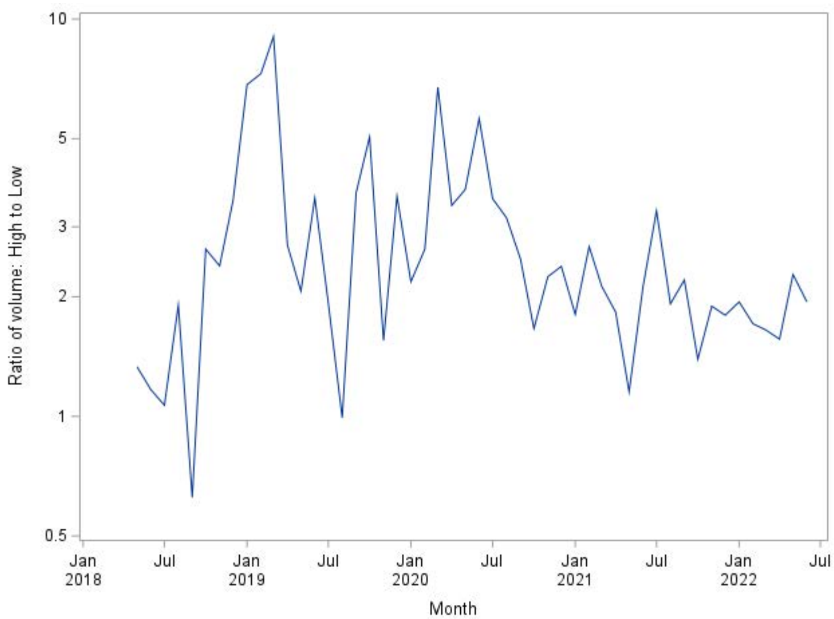
Appendix Figure 4. Purchase volume from high- and low-bid dispersion investors over time: fixed cut-off

This figure shows the purchase trading volume of high- and low-bid dispersion investors using a fixed cut-off in all months. Specifically, we define low-dispersion group as investors with $BidDispr = 0$, i.e., only one bid in previous three months, and high-dispersion group $BidDispr > 0$ and plot the monthly time-series of purchase volume per investor by the two groups.

(a) Purchase volume



(b) Ratio of volume



Appendix Figure 5. Distribution of the ratios of transaction prices to asking prices

This figure shows the distribution of the ratios of transaction prices to asking prices among all successful transactions.