Selection-Neglect in the NFT Bubble*

by

Dong Huang[a] and William N. Goetzmann[a]

Abstract

Using transaction data from a large non-fungible token (NFT) trading platform, this paper examines how selection-neglect interacts with extrapolative beliefs, possibly accelerating the boom and delaying the crash in the recent NFT bubble. We document extrapolative beliefs about increasing prices and an upward sample selection bias in observed transaction prices. We show that agents prone to selection-neglect formed more optimistic beliefs and traded more aggressively during the boom. When market liquidity declined, selection bias intensified and delayed the crash onset. Finally, investors with sophisticated bidding behavior showed less selection-neglect, responded more strongly to declining market conditions, and performed better.

JEL classification: G11; G12; G41; Z11

Keywords: Non-fungible token; Selection bias; Selection-neglect; Extrapolative belief; Bubble; MCMC

* 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, the 12th annual Portuguese Financial Network conference, and Yale SOM for helpful comments. All errors are the responsibility of the authors. Please direct correspondence to william.goetzmann@yale.edu (W.N. Goetzmann).
[a] Yale School of Management, Yale University, 165 Whitney Avenue, New Haven, CT, USA, 06511.

Introduction

An NFT is a digital collectible registered and transferable on a blockchain.[1] NFTs gained attention in March 2021 with the multi-million-dollar auction of a work by Mike Winkleman (Beeple), introducing a new market for speculation in digital collectibles. Observed prices in the NFT art marketplace rose by over 10,000% from summer 2019 to autumn 2021, only to decline spectacularly by summer 2022. The parallels to Dutch Tulipmania are unmistakable, as NFT artworks generate no economic dividends and reflect what collectors are willing to pay for ownership, aesthetic enjoyment, and speculative future resale value. The NFT bubble offers economists a rare opportunity to study a speculative bubble in detail, with blockchain records providing rich material to test theories about market dynamics.

In this paper, we use a database of bids, offers, and transactions from the SuperRare platform to test theories about expectations and trading behavior in the NFT market during the COVID-19 period. We test for behavioral heuristics that could support a bubble in asset prices: extrapolative beliefs and the disposition effect. We investigate whether these heuristics were exacerbated by selection-neglect—the mistaken belief that observed transaction prices represent the population mean. Finally, we examine the heterogeneity in behavioral biases and find that a subset of astute NFT traders made significantly higher profits during the bubble.

Selection bias is a concern in markets where transactions are rare and dependent on various conditioning variables.[2] This conditioning makes observed prices biased measures of mean population value and price trends biased measures of unrealized capital appreciation. Despite this,

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

[2] For example, selection biases may arise in the housing market (Goetzmann and Peng, 2006; Andersen et al., 2022), venture capital evaluation (Korteweg and Sorensen, 2010), art investment (Korteweg et al., 2016) , and wine markets (Van Tillo and Verdickt, 2023). Selection bias is also relevant in markets with information transmission friction, for example, mutual fund advertising (Koehler and Mercer, 2009), saving and consumption decisions (Han et al., 2019), and media biases (Enke, 2020).

market transaction data may be mistakenly interpreted as representative of unrealized values and returns.[3] We find evidence in the SuperRare NFT marketplace that participants conditioned their expectations and transactions on trends estimated from censored data, which played a significant role in bubble formation.

Economic experiments have shown that decision-makers can be prone to selection-neglect and misled by observed information, potentially leading to market dislocation.[4] For example, Hirshleifer and Plotkin (2021) find that a combination of salience and selection-neglect can explain investment booms. Non-experimental evidence about selection-neglect is limited, with few exceptions such as Van Tillo and Verdickt (2024) in the wine market. Records from the SuperRare marketplace over 2018-2022 allow us to observe or infer the beliefs of individual agents during a high-activity period. Using rich bid and ask data, we estimate expected speculative profits and document classic signs of a speculative bubble in the NFT market during 2021-2022: a rapid rise in asset prices followed by a dramatic decline, with a significant lead-lag relationship between trading volume and market prices. The cross-sectional volatility of transaction prices increased, suggesting greater disagreement among market participants during the bubble's growth phase.

Seller reservation prices in a market for heterogeneous goods can bias information about market conditions. Due to factors like liquidity constraints or prospect theory effects, seller reserves of non-fungible goods can differ from market consensus value. Without appropriate adjustment, transaction prices provide inadequate knowledge about true market conditions.

We test whether seller reservation prices induced a selection bias in transaction prices in two ways. First, we proxy potential transaction prices of untraded NFTs using the mid-point of the

[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), Goetzmann et al. (2012), and Glaeser and Nathanson (2017).
[4] cf. Esponda and Vespa (2018), Barron et al. (2024), Enke (2020), Ali et al., (2021), and López-Pérez et al. (2022).

latest bid and ask each month. The combined-price index is much lower than an index using only transaction prices, indicating significant positive selection bias. Second, we estimate a bias-corrected NFT index using information available as of the date of trade. Our econometric specification allows us to test whether seller prospect theory utility may have generated the selection bias. We find that sellers with potential losses set higher asking prices than those with potential gains, indicating reluctance to realize losses. This results in observed transactions having asymmetrically greater mass on realized gains.

To estimate the selection-corrected index, we adopt the data-augmented Markov Chain Monte-Carlo (MCMC) algorithm implemented by Korteweg et al. (2016). 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, with the corrected indexes declining much earlier. [5]

A key question is whether investors are aware of the selection bias and adjust their beliefs accordingly, or if 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 account for selection bias, they will form over-optimistic price expectations due to observing upward-biased prices. When true market conditions begin to deteriorate, the declining trend may be masked by selection bias, causing investors to continue purchasing NFTs and delaying the realization of a downturn in observed market prices.

[5] Using a subset of 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 major NFT collections (selected ex-post) consistent with selection bias, although they characterize it as economically insignificant (1% to 3% weekly).

We test whether NFT investors extrapolated past returns at the beginning of the bubble. We find that past returns of both the baseline RSR and the selection-adjusted index are positively correlated to market trading volume. We then test whether selection-neglect has additional explanatory power to predict trading volume. Specifically, we proxy selection-bias each month as the difference between the rolling six-month returns of a standard repeat-sale index and a selection-adjusted index. Regression results show that a one-standard-deviation increase in selection-bias amplifies volume sensitivity to positive past true returns and weakens the relationship when past returns are negative. These findings support the hypothesis that selection-neglect extrapolators tend to trade more aggressively than those attending to selection bias.

Next, we test for differing abilities among market participants to perceive and correct for selection bias using two metrics of sophisticated bidding behavior. The first is attentiveness to the broad cross-section of NFT market values, measured by bid dispersion. We hypothesize that investors bidding across a range of price tiers are more likely to understand the actual price distribution and thus be less subject to selection bias. The second measure is the fraction of NFTs purchased below the asking price, suggesting better assessment of NFT values and correction for overly optimistic beliefs. We find that investors with more NFTs purchased below asking prices are more responsive to price declines.

Finally, by adding investor 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. We find that astute investors identified by either measure outperformed their counterparts, even after including unrealized returns.

Our research contributes to the literature on investor behavior during a bubble by using the NFT market to study the role of selection-neglect in the ascending and declining stages. Our findings with market data supplement experimental evidence showing suboptimal decisions induced by selection-neglect and taking the observed sample as representative of the population distribution.[6] The results are also 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 suggest that the neglect of social assortativity can lead to misperception of income inequality and political attitude polarization.

We also provide supportive evidence for behavioral biases, including prospect theory utility 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; Liao et al., 2022), 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 proven to be critical in explaining bubbles in those markets, and our findings suggest that the NFT market is no exception. The relationship between the three behavioral biases that we document in NFT markets is potentially important in other major markets, for example, in real-estate markets. Papers by Genesove and Mayer (2001) and more recently Andersen et al. (2022) both find asymmetric patterns in house listing and selling decisions regarding potential gains and losses. In particular, Andersen et al. (2022) show that reference dependence and loss aversion result in a greater propensity for houseowners to list and sell their houses when they expect potential gains than potential losses. The distribution of observed transactions also exhibits sharp bunching at realized gains of zero, which attenuates to the right of zero. These observations

[6] cf. Koehler and Mercer, 2009; Barron et al., 2024; Enke, 2020; Ali et al., 2021; Van Tillo and Verdickt, 2023.

imply a significant overweight on realized gains than on realized losses, i.e., an upward selection bias. Given that extrapolators constitute an important force in the housing market (Glaeser and Nathanson, 2017), the existence of upward selection bias and the neglect of it can amplify their optimism, further accelerating the boom, and delaying the crash.

Current work on NFTs mainly focuses on market dynamics and market structure (cf. Kräussl and Tugenetti, 2023; Borri et al., 2023; Lommers et al., 2023). Borri et al. (2023), examine factors that may affect NFT prices, including quality, rarity, and visual characteristics. Chen et al. (2023) propose an asset pricing model of digital art and use NFT markets for calibration. A number of papers examine investor behavior in the NFT market. For instance, Oh et al. (2023) document the intensive participation of experienced investors in primary market sales of NFT collectibles. 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. Sun (2023) utilizes the random assignment of NFTs with different values among buyers in the primary sale market to examine the experience effect on investors' future market participation. However, the role of behavioral heuristics and biases and their potential contribution to the NFT bubble have not been directly and fully explored. Our paper adds to existing NFT research on these issues.

The remainder of this paper is organized as follows. Section 1 provides an introduction to the institutional background and our data from SuperRare. Section 2 describes the bubble and speculative activities in the market. Sections 3 and 4 provide evidence for investors' prospect theory utility and extrapolative beliefs, respectively. We address the role of selection-neglect in Section 5. Section 6 concludes.

1. 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 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). This is a small subset of the data collected and used by Borri et. al. (2023) who analyze price data from most of the major NFT marketplaces over the bubble period. The advantage of the SuperRare data is the additional information provided other than transaction prices. 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 the 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.

SuperRare selects and invites digital artists to create their work on the platform, however the potential biasing effect of this selectivity is not obvious, other than the SuperRare NFTs are more likely to be higher -priced and/or resold. A comparison of the raw indexes constructed from SuperRare indicate that the trends and magnitudes of the bubble are similar in magnitude. [CHECK] While auction dynamics are not observable for the entire universe of NFTs, the creation, listing

[8] https://www.forbes.com/sites/nicolesilver/2021/11/02/the-history-and-future-of-nfts/

auction and transaction mechanisms of the SuperRare platform are standard. 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]

1.1. 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 [@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 our screen capture, it was listed for sale for 69 ETH or the current equivalent of $115,300.

[9] https://help.SuperRare.com/en/articles/4576436-how-auctions-work-on-SuperRare

Figure 2 shows another type of sale: an auction with a reserve. The artist, Alex CGbans [@alex_cgbans] minted *The Frog King* and set up an auction with a reserve for it in November 2022. The reserve was met with a bid of 1 ETH from bidder @collin on November 23, 2022, which triggered a 24-hour auction.[10] 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 at the time of screen capture it 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 on 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 in *VR.Girl* is an example of the stratospheric returns that some speculators in the NFT market realized during the peak of the bubble.

[10] Before the auction started, potential buyers could see the reserve prices on the webpage, and they decide whether or not to place a bid price greater than or equal to the reserve price to trigger the 24-hour auction. The triggering bid is usually the same as the reserve price. After the auction ends, only the bidding history and the new listing or reserve price is available on the website as shown in Figure 2.

1.2. 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 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.

2. 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 arise during bubble episodes (Brunnermeier and Nagel, 2004, Temin and Voth, 2004). In this section we provide evidence that the NFT bubble displayed many of these 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.

2.1. Price and Volume Dynamics

As a metric for a survival-biased index we construct a price index of NFTs using the baseline RSR. Returns are calculated in U.S. dollars. 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 thick 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, NFT prices start to fall, along with other digital assets and the S&P500. The NFT price index thus demonstrates a distinctive boom and bust pattern that is

related to other trends in digital assets as well as to the equity market. This co-movement during the COVID period suggests that 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, which is methodologically equivalent to the widely watched S&P CoreLogic Case–Shiller U.S. National Home Price Index during the global financial crisis. Figure 4b presents a series of indexes (light grey lines) estimated with repeat-sale pairs available up to each month, that is, using expanding windows, 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 first "up-to" index ends in April, 2019 so that we can have at least one year of observations for a reliable estimation. The blue solid line connects the ending index levels of the "up-to" indexes for every month, except for the index levels before April, 2019 which are from the first "up-to" index estimation. The red dashed line is the standard RSR index estimated using the full sample period, the same as the NFT RSR index in Figure 4a.

The figure suggests that the up-to indexes have quite similar trends to the full-sample index, but they are more volatile than the full-sample index, especially in the early part of our sample period. In addition, the up-to indexes are generally higher than the full-sample index, which has the richest information. This observation suggests that unsold NFTs are experienced lower (unrealized) returns than twice-traded NFTs and, therefore, as previously unsold NFTs are traded later and their prices are revealed, the new information revises the repeat-sale index downward.

Such downward revisions in repeat-sale indexes are suggestive evidence of upward bias in observed prices (Van de Minne et al., 2020). Also note that the downward revisions are smaller towards the end of our sample period, but this does not mean that there is less upward selection bias. Instead, it is because that we observe relatively less new information in later months. The revisions are incomplete – even the full-sample index here is unable to account for all the untraded NFTs and is thus estimated with an upward bias. If we were able to observe transactions after our sample period and the selection bias persisted, we might expect more downward adjustments than what is shown here.

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, price and 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 expectations (Barberis et al., 2018) or overconfidence about 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 6 calculate the standard deviation of prices denominated in ETH and in USD. They show similar patterns: the volatility jumped in November 2019, maintained a rising trend until the peak in early 2022, and dropped afterward. This pattern suggests an increase in price disagreement while the bubble was growing.

In the Online Appendix A, we 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 that the owner can decide to accept or not. Dispersion in bid prices for the same NFT within a short period of time is a potential proxy for belief disagreement.[11] Appendix Figure A1 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 A2 shows the average, median, 25th, and 75th percentiles of the log asking-to-

[11] 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.

purchase price ratios over the sample period and Appendix Figure A3 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.

2.2. 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 sets 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. In 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.

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.[12] The figure also

[12] 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.

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 the 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.

3. Selection Bias in Successful Transactions

3.1. 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 reference dependence, 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 impute 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), in which now each repeat-sale pair consists of two consecutive available transaction prices, either observed or imputed.[13]

Figure 8a reports the indexes for the baseline RSR (thin blue line) and the bid-ask augmented RSR (red dashed 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 was reached in October 2021 at around 46,000 while the 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.

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 to describe the probability of a sale. Kortweg et al. (2016) use a data-augmented Markov chain Monte Carlo (MCMC) method that is most useful to our setting. This approach can be regarded as an extension of the Heckman (1976) sample selection model to account for the sample selection problem, as well as a nonlinear Kalman filter in the state-space representation. The intuition is that, in the presence of reference dependence, the probability of a sale depends on potential gains and losses since the purchase. Hence, even though the true prices are unobservable unless a transaction occurs, the non-occurrence of transactions is still informative

[13] 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.

about the latent prices of NFTs. Kortweg et al. (2016) make use of such information and manage to impute all the latent prices given the assumptions about the joint distributions of NFT prices and the probability of a sale. The selection-adjusted NFT price index can then be estimated from the observed and imputed prices.

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 log price of NFT i in month t, $\delta(t)$ is the common return factor (i.e., the market return of NFTs) to be determined, and $\varepsilon_i(t)$ is the idiosyncratic risk deviation from the common return factor. $\varepsilon_i(t)$ are assumed to be independent and identically distributed following a normal distribution with mean 0 and variance σ^2 that need to be estimated. Equation (3) is the selection equation that relates the probability of a sale with latent NFT prices. As in a probit model, $w_i(t)$ can be viewed as the latent random "utility" of selling, which is a function of potential holding period return $r_i^p(t) = p_i(t) - p_i(b)$ (with b being the time of last purchase), the time interval since the last sale defined

as $h_i(t) = t - b$, and a standard normal random disturbance $\eta_i(t)$.[14] To capture the disposition effect, $w_i(t)$ is assumed to be a piece-wised quadratic function of $r_i^p(t)$ with a potential jump at zero, with α's being the coefficients to estimate. As in the Heckman sample selection model, the selection equation needs to include at least one regressor that does not appear in the outcome equation, i.e. equation (2); otherwise, the model is technically identified only from the distribution assumptions and not from exogeneous variation in the probability of a sale (Cameron and Trivedi, 2005).[15] In our setting, $h_i(t)$ and its square are the exogeneous variables for identification. Holding periods are relevant to investors' selling decisions but are unlikely to affect the latent prices of NFTs, which are determined by market equilibrium.[16] Finally, as in a Heckman sample selection model, equation (4) specifies the selection rule. A transaction happens and price $p_i^*(t)$ is observed if and only if the latent selection variable $w_i(t)$ is no less than zero.

Equations (2)-(4) specify the joint distribution of parameters ($\theta = \{\delta, \alpha, \sigma^2, \{p_i(t), w_i(t)\}\}$) and data ($\{p_i^*(t)\}$), which is denoted by $f(\theta, data)$. Estimating the system is challenging due to the nonlinearity of equation (3) in terms of the latent prices.[17] Kortweg et al. (2016) propose an MCMC algorithm to estimate it. Loosely speaking, given a prior about parameter distributions, the MCMC algorithm iteratively simulates a realization of one parameter (θ_j, where j indexes

[14] 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*.

[15] Recall that in estimating a Heckman sample selection model using the two-step approach, the first step is to estimate the selection equation to obtain predicted inverse Mills ratios, which are used as an additional regressor in estimating the outcome equation in the second step. Under the normality assumption, the inverse Mills ratio is a nonlinear function of the regressors and, therefore, there is no multicollinearity in the second step even if the outcome equation and the selection equation are both linear and contain the same set of explanatory variables.

[16] 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.

[17] When the selection equation is linear in the outcome variable, the system is a standard Kalman filter and can be estimated using the Forward Filtering Backwards Sampling procedure, which has an analytical solution (Korteweg and Sorensen, 2010).

parameters) at a time for all parameters based on the conditional distribution of the parameter $f(\theta_j|\theta_{-j}, data)$ conditioning on the data and the simulated realizations of other parameters (θ_{-j}) from previous simulations. The process continues until the distributions of simulated parameters reach convergence, and the series of simulated parameters after dropping the initial "burn-in" period before convergence give the posterior joint distribution of the parameters and the marginal distributions of each parameter. Details of the algorithm can be found in Online Appendix B and the technical notes of Kortweg et al. (2016).

The thick green line in Figure 8a reports the selection-adjusted NFT price index estimated from the MCMC algorithm, along with the baseline RSR index and the bid-ask augmented index. The adjusted index takes into account the unobserved prices of untraded NFTs and is considerably lower than the unadjusted index. It suggests that the observed returns are on average larger than the unobserved returns, indicating a positive sample 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 true market price index. This is similar to the experimental setting in Enke (2020), in which participants have coarse information to adjust for selection bias but fail to do so. Finally, note that the adjusted index declines substantially after September, 2021 – much earlier and faster than the unadjusted indexes, which reached a peak in October, 2021 and remained high until the end of our sample period. Such a difference in the later stages of our sample period suggests a large selection bias as market liquidity dried up.

Table 2 column 1 reports the estimated coefficients of the selection equation (3) and Figure 9 plots the probability of a sale with respect to potential holding returns at different horizons

according to the estimated coefficients.[18] First note that the probability of a sale is quite low, around 6% for a zero one-month holding return, compared to other asset classes. Hence, the NFT market was thin and illiquid. Second, consistent with predictions from the disposition effect, the probability of a sale jumps abruptly at zero, as suggested by a significantly positive coefficient on the gain indicator $I\left(r_i^p(t) > 0\right)$ which corresponds to the doubling of the sale probability at the origin as in Figure 9 at multiple holding horizons. In addition, the linear and quadratic terms of holding period returns in the positive part, $I\left(r_i^p(t) > 0\right) \cdot r_i^p(t)$ and $I\left(r_i^p(t) > 0\right) \cdot r_i^p(t)^2$, respectively, are both positive and significant, suggesting that the sale probability increases with potential gains – also a feature of the disposition effect. NFTs with potential losses are unlikely to be traded as the sale probability is relatively low for negative potential returns. This reflects the observation that some NFTs experiencing large negative returns lose the attention of market participants and become almost worthless, remaining untraded for an extended period. Finally, the coefficients on both holding horizons are significant, indicating strong identification power.

Since NFTs are denominated in ETH, the returns of ETH could be related to investor trading decisions (Borri et al., 2023). Therefore, in Table 2 column 2 we add ETH returns as an additional covariate in the selection equation. The coefficient on ETH return is positive, suggesting that the probability of a sale is increasing in ETH returns. Also, people's attention to NFT markets may affect their trading, so in column 3 we add Google search volume index (SVI) as a proxy for attention, whose coefficient turns out to be positive as expected. In column 4, we add both covariates in the selection equation and obtain similar coefficients as before. In all specifications, the coefficients on holding returns are similar to column 1. In Online Appendix Figure A4, we

[18] In Online Appendix B, we implement several diagnostics following Korteweg and Sorensen (2010) and Gelman et al. (2013) to confirm the convergence of the Markov chains.

present the estimated MCMC selection-adjusted indexes from different specifications of the selection equation. The indexes are quite close to each other, so the trends and the index levels are robust to the inclusion of the two additional covariates.

As in the case of the repeat-sale index, we also estimate an "up-to" version of the MCMC selection-adjusted index, that is, we estimate a series of MCMC indexes with repeat-sale pairs available up to each month. Figure 8b presents the results. The light grey lines are "up-to" indexes ending in each month. We require at least one year of observations for a reliable estimation, so the first index ends in April, 2019. The blue solid line connects all ending index levels, and the red dashed line is the full-sample estimation. Since the "up-to" indexes can only use limited information, they are more volatile than the full-sample index.

The up-to index highlights the decline of NFT prices in late 2019, when NFTs still received little attention. There were many NFTs listed in the market but only a few were traded. Without additional information, sophisticated investors at that time would therefore conclude that the actual NFT prices were quite low. Also, the "up-to" index levels during that time fluctuated considerably, which is another evidence of limited information and large uncertainty. For later periods, the up-to indexes track the full-sample index closely. In contrast to the RSR index where the up-to version is generally above the full-sample index, the up-to MCMC indexes do not show systematic one-way deviations from the full-sample MCMC index. This finding suggests that our MCMC adjustment is able to account for the latent prices of untraded NFTs so that the realizations of new transaction prices do not lead to consistent downward revisions of the old up-to indexes.

A useful feature of the MCMC algorithm 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. First note that observed prices are much fewer than latent prices, indicating that transactions are rare events in this market. The figure also shows that the distribution of observed prices (white bars with red borders) is shifted to the right compared to the distribution of latent prices (filled blue bars), with a mean difference of 141% ($\approx e^{0.88} - 1$). Figure 10b plots the monthly time series of mean differences. The differences, i.e. the sample selection biases, are predominantly positive and increase sharply during the market downturn.

3.2. Prospect Theory as a Source of Selection Bias

Considerable research has documented prospect theory effects in investor behavior, including reference dependence and diminishing sensitivity, both of which are implications of the theory. Reference dependence suggests that an agent's utility function is defined over gains and losses relative to a reference point, while diminishing sensitivity posits that agents are risk averse over moderate-probability gains and risk-seeking over moderate-probability losses (Barberis et al., 2021). These two common psychological tendencies provide a potential explanation for the disposition effect and, hence, selection bias in NFT markets.[19] Specifically, investors make sale decisions conditioning upon (anchoring on) the purchase price of the asset and evaluate their utility of a sale according to the potential holding returns. Investors with diminishing sensitivity are more willing to realize gains and less willing to realize losses during a market downturn. 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

[19] A third element of prospect theory is loss aversion, which implies that people have a greater sensitivity to losses than to gains. It suggests that agent's utility function has a kink at its origin. Although loss aversion was long believed to be the cause of the disposition effect, recent research suggests that the other two aspects of prospect theory, that is, reference dependence and diminishing sensitivity, are more relevant to the disposition effect (Barberis and Xiong, 2009). In addition, realization utility alone can also explain the disposition effect (Barberis and Xiong, 2012).

market would likely be higher than the population distribution of NFT prices, leading to a positive bias.

In the presence of the prospect theory effect, a seller with a potential loss would be expected to set a higher reservation price than the seller of an NFT with a potential gain. In this section we test for evidence of the prospect theory effect 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 effects 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,

$$Log\big(AskPrice_{i,t}\big)\big(or\ AskMarkUp_{i,t}\big) = \beta_0 + \beta_1 Log\big(PredictPrice_{i,t}\big) + \beta_2 \min\{0, PotentialReturn_{i,t}\} + \beta \boldsymbol{X} + \varepsilon_{i,t}, \quad (5)$$

where $AskPrice_{i,t}$ is the asking price of NFT i in month t listed by the owner, $AskMarkUp \equiv Log\big(AskPrice_{i,t}\big) - Log\big(PredictPrice_{i,t}\big)$ 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 2.1.[20] The use of the baseline RSR implicitly assumes that investors do not perceive it as an biased measure. $\boldsymbol{X}$ variables include a quadratic function of

[20] 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. The RSR is equivalent to restricting the sample to sales for which there is perfect, rather than estimated hedonic control.

holding horizon (in months), artist fixed effects and time fixed effects. If the sellers are subject to diminishing sensitivity, 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.

$$\begin{aligned} &Log\big(AskPrice_{i,t}\big)\big(or\ AskMarkUp_{i,t}\big) \\ &\quad = \gamma_0 + \gamma_1 Log\big(PredictPrice_{i,t}\big) + \gamma_2 1\big(PotentialRet_{i,t} < 0\big) \\ &\quad + \gamma_3 PotentialRet_{i,t}^{+} + \gamma_4 PotentialRet_{i,t}^{+\,2} + \gamma_5 PotentialRet_{i,t}^{-} \\ &\quad + \gamma_6 PotentialRet_{i,t}^{-\,2} + \beta \boldsymbol{X} + \varepsilon_{i,t}. \end{aligned} \tag{6}$$

With the specification in equation (6), diminishing sensitivity 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 3 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 on the loss dummy, $1\big(PotentialRet_{i,t} < 0\big)$, 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 on $PotentialRet_{i,t}^{+}$, 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.

4. Extrapolative Beliefs

4.1. 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, 2023; DeFusco et al., 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.[21] If extrapolative beliefs are 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 (7). 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 the three methods in Section 3.1.

$$Log(Volume_t) = \beta_0 + \beta_1 LogReturn_{t-\tau \to t-1} + \varepsilon_t. \quad (7)$$

Table 4 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.[22] All specifications include past six-month Ethereum returns as an additional covariate. Columns 1, 3, 5, and 7 use the standard RSR index, the MCMC selection-adjusted index, the "up-to" RSR index, and the "up-to" MCMC selection-adjusted index

[21] Case and Shiller (1988) 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.
[22] Standard errors are adjusted for autocorrelation using the Newey-West correction with the optimal number of lags suggested by (Newey and West, 1987, 1994). The significance of the coefficients is robust to different choices of the lags.

to calculate past returns, respectively. The coefficients on past returns are mostly positive and significant at the 1% level in most specifications.[23] They suggest that a 100% six-month log return, which is not uncommon in NFT markets, is associated with a 25~82% 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. Interestingly, after accounting for NFT returns, the ETH returns have insignificant explanatory power on trading volume.

To further explore the relationship, the even columns of Table 4 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 on negative past returns is insignificant and has the opposite sign. 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 44) 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 5.

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 SuperRare website and other media. The explanatory variable

[23] The coefficient of the NFT returns with the "up-to" RSR index (Table 4 column 5) is positive but insignificant because the trading volume's sensitivity to the positive and negative parts of past market index returns has comparable magnitudes but opposite signs as suggested by Table 4 column 6, leading to average sensitivity that is closed to zero. Also, the strong correlation between the "up-to" RSR index returns and ETH returns could inflate the standard errors. Nevertheless, the positive signs are consistent with the extrapolative beliefs.

is past market realized return measured by rolling six-month mean, maximum, or minimum returns. 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 5 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 focus on positive news and ignore relatively negative news in the market. In column 4, we test the effect of horizon-adjusted realized returns, which also displays significance.

4.2. 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 and Veronese, 2009; Acheson et. al. 2009; Quinn and Turner 2021; Kindleberger, 1996; 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 in 1719-20 which sparked the neologism *millionaire*, or more recently, news of investment killings in cryptocurrency or Chinese warrants.[24] Many historical bubbles involved the rush to invest in new property markets, for example, the 1920's Florida land bubble.[25]

[24] For the Mississippi Bubble, Velde (2003), for cryptocurrency news and network effects, Sockin and Xiong (2023), for Chinese warrants, Pearson et al. (2021) and Xiong and Yu (2011).
[25] See Sakolski (1932) for a list of American land bubbles including Florida.

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. This interest rate cut coincides with a turn in the price index and a sharp increase in speculative activity as shown in Figure 7. The number of purchases by short-term speculators in November 2019 is four times larger than that in the previous month. The decision to lower interest rates in March and April 2020 as a measure against the COVID pandemic may have further fueled the market. This period experienced another peak in speculation and returns.[26] By the same token the interest rate increases in early 2022 were accompanied by a downturn in market prices and volume.

5. 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

[26] 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 when the famous NFT *Everydays: the First 5000 Days* by Mike Winkelmann (a.k.a. Beepler) sold at auction for $69 million -- much later than the first wave in NFT transactions in the market.

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.

5.1. 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 (7) in Section 4.1. Specifically, we use the following specification,

$$\begin{aligned} Log(Volume_t) &= \beta_0 + \beta_1 LogUnbiasedReturn_{t-1} + \beta_2 SelectionBias_{t-1} \\ &+ \beta_3 LogUnbiasedReturn^{+}_{t-1} \times SelectionBias_{t-1} \\ &+ \beta_4 LogUnbiasedReturn^{-}_{t-1} \times SelectionBias_{t-1} + \varepsilon_t. \end{aligned} \tag{8}$$

In equation (8), we use the past six-month unbiased market returns calculated from the full-sample MCMC selection-adjusted index before t to capture the extrapolation effect. Hence, the benchmark is that investors can form their beliefs based on true unbiased returns, or "as-if" they could observe real-time prices. $SelectionBias$ is a measure of selection bias defined as the difference between the six-month observed market returns implied by the baseline RSR index and the six-month true market returns from the MCMC selection-adjusted index. Figure 12 presents the time series of selection bias, which resembles the pattern of mean difference between observed and latent prices as shown in Figure 10b that are mostly positive and increased towards the end of our sample period as market liquidity deteriorated.

In addition to singular 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 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 6 reports the regression results for the interactions between extrapolation and selection bias. Column 1, as a benchmark, reproduces column 3 of Table 4A. 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: elevated selection bias coincides with a deterioration in market liquidity and declining returns. The interaction term in the positive return region is 0.676 and modestly significant, 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 -0.409 when the unbiased return is negative, meaning that market trading volume is less sensitive to price declines when the selection bias is larger. The row “Impact Diff.” tests the difference of the coefficients on the interaction terms ("positive” minus “negative”), that is, the difference of selection bias’s impacts on positive and negative true market returns. The difference is 1.085 and significant at the 5% level. Column 3 splits the unbiased returns into positive and negative parts and generates quantitatively 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 6 replicate the regressions in the first three columns, respectively, with selection bias defined as the difference between the six-month market returns implied by the "up-to" RSR indexes and the six-month true market returns in the MCMC selection-adjusted index. We also change $LogUnbiasedReturn$ to the past six-month "up-to" MCMC selection-adjusted index returns. Such a modification is conservative in the sense that it assumes that even sophisticated investors can only condition their beliefs on observed past transactions and know no extra information that may predict future transactions. As before, the interaction terms with positive (negative) returns have positive (negative) coefficients and are significant at the 5% level or above, indicating that selection bias amplifies extrapolation with positive returns and dampens extrapolation when returns are negative. The difference in the impact across the two regions of returns is significant and of similar magnitude, as demonstrated in columns 5 and 6.

5.2. Proxies for Astute Investors

In the following section, we propose measures to identify astute 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,

$$BidDispr_{i,t} = \frac{P_{0.75}\left(\{Bid\}_{i,t-2\leq s\leq t}\right) - P_{0.25}\left(\{Bid\}_{i,t-2\leq s\leq t}\right)}{Med\left(\{Bid\}_{i,t-2\leq s\leq t}\right)}, \quad (9)$$

where $\{Bid\}_{i,t-2\leq s\leq t}$ is the set of log bid prices offered by investor i for any NFTs in months $t-2$ to t. For each investor-month-NFT, 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 log 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.[27]

Table 7 repeats the analysis in Table 6 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 suggests 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 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 (-0.139 = 0.292 + (-0.431)) 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 may 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.

[27] 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 A5, 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. We use "*per investor*" volume since the number of investors in the two groups are different and the volume cannot compare directly. Similar patterns in purchase volume and the gap are obtained.

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 ask. Appendix Figure A6 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 ask, the magnitudes of the discounts are also large – many transactions receive around 60% to 70% off.

5.3. Selection-Neglect Across Investors

Individual investor transaction information provides further evidence on how past aggregate market returns and selection bias affects the trading behavior of investors with various debiasing abilities. 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:

$$\begin{aligned} Trade_{i,t} = \alpha_i &+ \beta_1 LogUnbiasedReturn^{+}_{t-1} + \beta_2 LogUnbiasedReturn^{-}_{t-1} \\ &+ \beta_3 SelectionBias_{t-1} \\ &+ \beta_4 LogUnbiasedReturn^{+}_{t-1} \times SelectionBias_{t-1} \\ &+ \beta_5 LogUnbiasedReturn^{-}_{t-1} \times SelectionBias_{t-1} + \varepsilon_{i,t}. \end{aligned} \tag{10}$$

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 past six-month 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 the 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. Both specifications include past six-month Ethereum returns. Standard errors are clustered at the investor level to account for residual autocorrelation within investors.

Table 8 reports the regression results. For ease of interpretation, selection bias is normalized. The first three columns use log purchase volume ($Log\big(Buy_{i,t}+1\big)$) in each investor-month as the dependent variable and the last three use the purchase indicator ($1\big(Buy_{i,t}>0\big)$). The coefficients of past log returns are all positive, consistent with the prior results of extrapolative beliefs among NFT investors. As in Section 5.1, 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.[28]

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 (10) with the proxy

[28] The coefficients on Ethereum returns are negative, as opposed to the results in the aggregate-level regressions. This may be due to the fact that the individual-level regressions include investor fixed effects, so they essentially capture variations within individuals who stay active in the market. The individual-level regressions fail to account for the trading from newly-entry investors, who may be attracted to NFT markets by large ETH returns.

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 (11) below:

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

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):

$$\begin{aligned} &\left(\beta_1 + \beta_4 SelectionBias_{t-1} + \beta_7 Ability_{i,t-1} + \beta_{10} SelectionBias_{t-1} \times Ability_{i,t-1}\right) \\ &\qquad \times LogUnbiasedReturn^{+}_{t-1}. \end{aligned} \tag{12}$$

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 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 9. 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 9 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 and significant, so investors who tend to buy NFTs at lower prices are less subject to selection bias in 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 impact of selection bias. The coefficients on the interactions between past returns and ability are both negative, indicating that high-ability investors extrapolate less from past returns.

Lastly, as a robustness check, Online Appendix Table A1 and Table A2 re-estimate the equations in Table 8 and Table 9, respectively, using "up-to" indexes as used in the last three columns of Table 6 and obtain qualitatively similar results. It suggests that the impacts of selection bias still exist even if we assume that the benchmark sophisticated investors can adjust the returns using transaction information only up to the current month.

In sum, our results suggest that selection bias in realized 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 appear more likely to understand true market conditions better than their counterparts and hence are less affected by selection bias. Such heterogeneity among investors' ability to account for selection bias is consistent with experimental findings from Ali et al. (2021).

5.4. Performance of Sophisticated Investors

Previous sections show that astute investors, as identified by the two ability proxies are less influenced by selection bias during a 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 a standard repeat-sale regression equation. Specifically, we estimate the following equation,

$$r_{i,b\to s} = \sum_{t=b_i+1}^{s_i} \mu_t + \gamma(s_i - b_i) \times X_{i,b} + \varepsilon_{i,b\to s}, \quad (13)$$

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 on a 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 10 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 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 inexperienced investors. The return gap remains negative but becomes insignificant after including unrealized returns. These findings suggest that, instead of experienced investors underperforming their counterparts, experienced investors are more likely to realize losses. Since there is a large selection bias in successful transactions, realized returns understate the relative investment performance of experienced investors, on average.

Finally, in column 4, we examine the relationship between investor performance and average holding time. Lovo and Spaenjers (2018) predict that investors who derive greater utility from owning an aesthetic asset such as an NFT 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.

6. 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

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

Ali, S. N., Mihm, M., Siga, L., Tergiman, C., 2021. Adverse and advantageous selection in the laboratory. *American Economic Review*, 111(7), 2152-2178.

Allen, F., Gale, D., 2000. Asset price bubbles and monetary policy. In: *Global Governance and Financial Crises*. Routledge, London, pp. 33-56.

Andersen, S., Badarinza, C., Liu, L., Marx, J., Ramadorai, T., 2022. Reference Dependence in the Housing Market. *American Economic Review*, 112(10), 3398-3440.

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

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

Barberis, N., Jin, L. J., Wang, B., 2021. Prospect theory and stock market anomalies. *The Journal of Finance*, 76(5), 2639-2687.

Barberis, N., Xiong, W., 2009. What drives the disposition effect? An analysis of a long-standing preference-based explanation. *The Journal of Finance*, 64(2), 751-784.

Barberis, N., Xiong, W., 2012. Realization utility. *Journal of Financial Economics*, 104(2), 251-271.

Barbon, A., Ranaldo, A., 2023. NFT bubbles. arXiv:2303.06051.

Barron, K., Huck, S., Jehiel, P., 2024. Everyday econometricians: selection neglect and overoptimism when learning from others. *American Economic Journal: Microeconomics (forthcoming)*.

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), 1027-1039.

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., Ross, S.A., 1995. Survival. *The Journal of Finance*, 50(3), 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.

Cameron, A. C., Trivedi, P. K., 2005. *Microeconometrics: Methods and applications*. Cambridge University Press, Cambridge, United Kingdom.

Case, K., 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.

Chen, C. Y. H., Kraussl, R., Verwijmeren, P., 2023. The pricing of digital art. *Center for Financial Studies Working Paper* No.716.

Chinco, A., 2023. The ex-ante likelihood of bubbles. *Management Science*, 69(2), 1222-1244.

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.

Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., 2013. *Bayesian Data Analysis*, 3rd Edition. Chapman and Hall/CRC, Boca Raton.

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. 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. N., 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., Yen, J., 2012. The subprime crisis and house price appreciation. *The Journal of Real Estate Finance and Economics*, 44, 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., Walden, J., 2022. Social transmission bias and investor behavior. *Journal of Financial and Quantitative Analysis*, 57(1), 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. NBER, Cambridge, MA, USA, pp. 475-492.

Hirshleifer, D. 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., 2023. Non-Fungible Tokens (NFTs): A review of pricing determinants, applications and opportunities. *Journal of Economic Surveys*, 38(2), 555-574.

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., 2023. Digital Veblen goods. *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.

Sakolski, A.M., 1932. *The Great American Land Bubble: The Amazing Story of Land-Grabbing, Speculations, and Booms from Colonial Days to the Present Time*. Harper & Brothers Publishers, New York.

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

Sockin, M., Xiong, W., 2023. Decentralization through tokenization. *The Journal of Finance*, 78(1), 247-299.

Sun, C., 2023. Personal experience effects across markets: Evidence from NFT and cryptocurrency investing. Unpublished working paper. University of North Carolina at Chapel Hill.

Temin, P., 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.

Van Tillo, R., Verdickt, G., 2024. Selection neglect and the cross-section of wine returns. *SSRN Working Paper* No.4637208.

Velde, F. R., 2003. Government equity and money: John Law's system in 1720 France. *SSRN Working Paper* No. 486983.

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, New York.

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

Figures

Figure 1. Example of an NFT listing.

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.

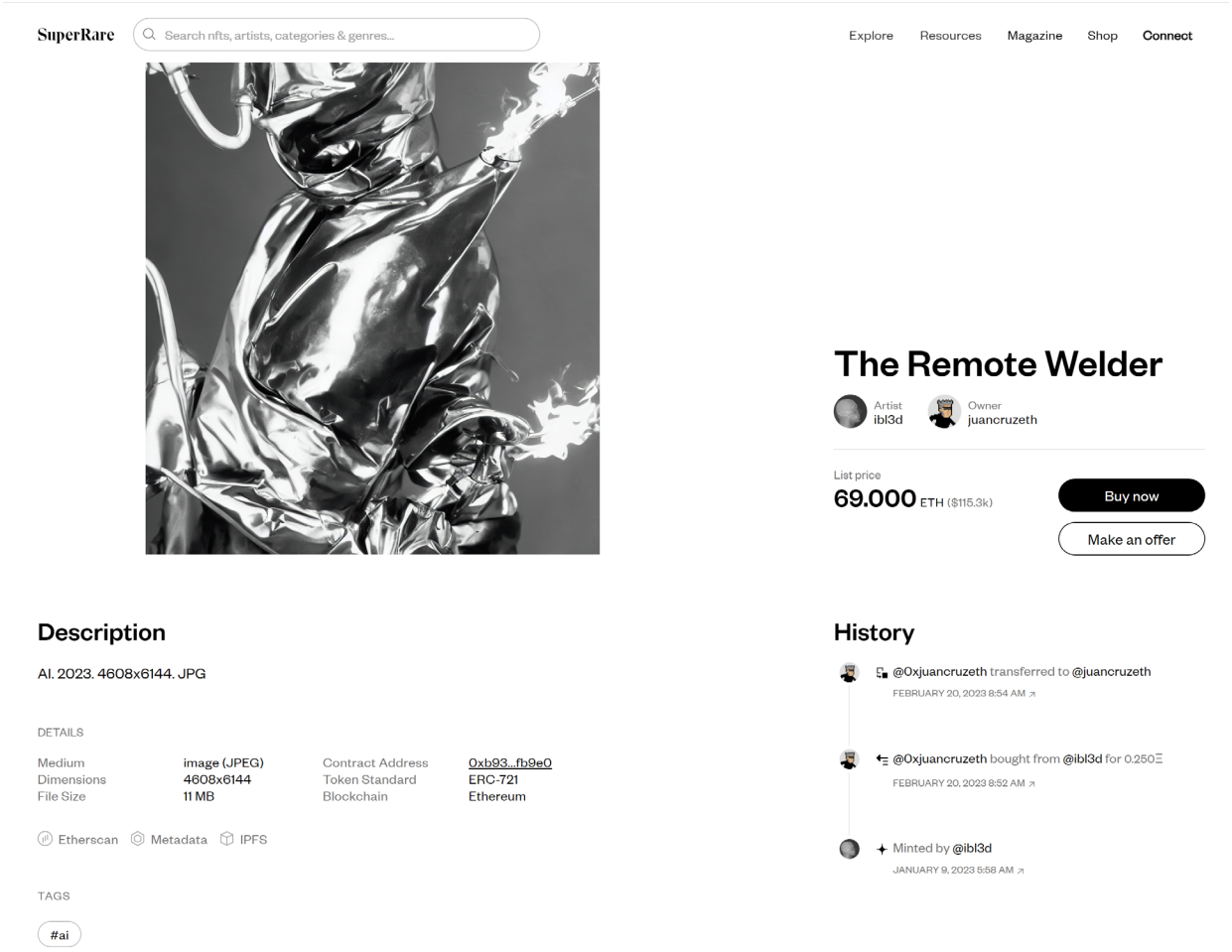


Figure 2. Example of an NFT auction.

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 (zoomed-in) also shows a complete bidding history of an auction happened in November 2022.

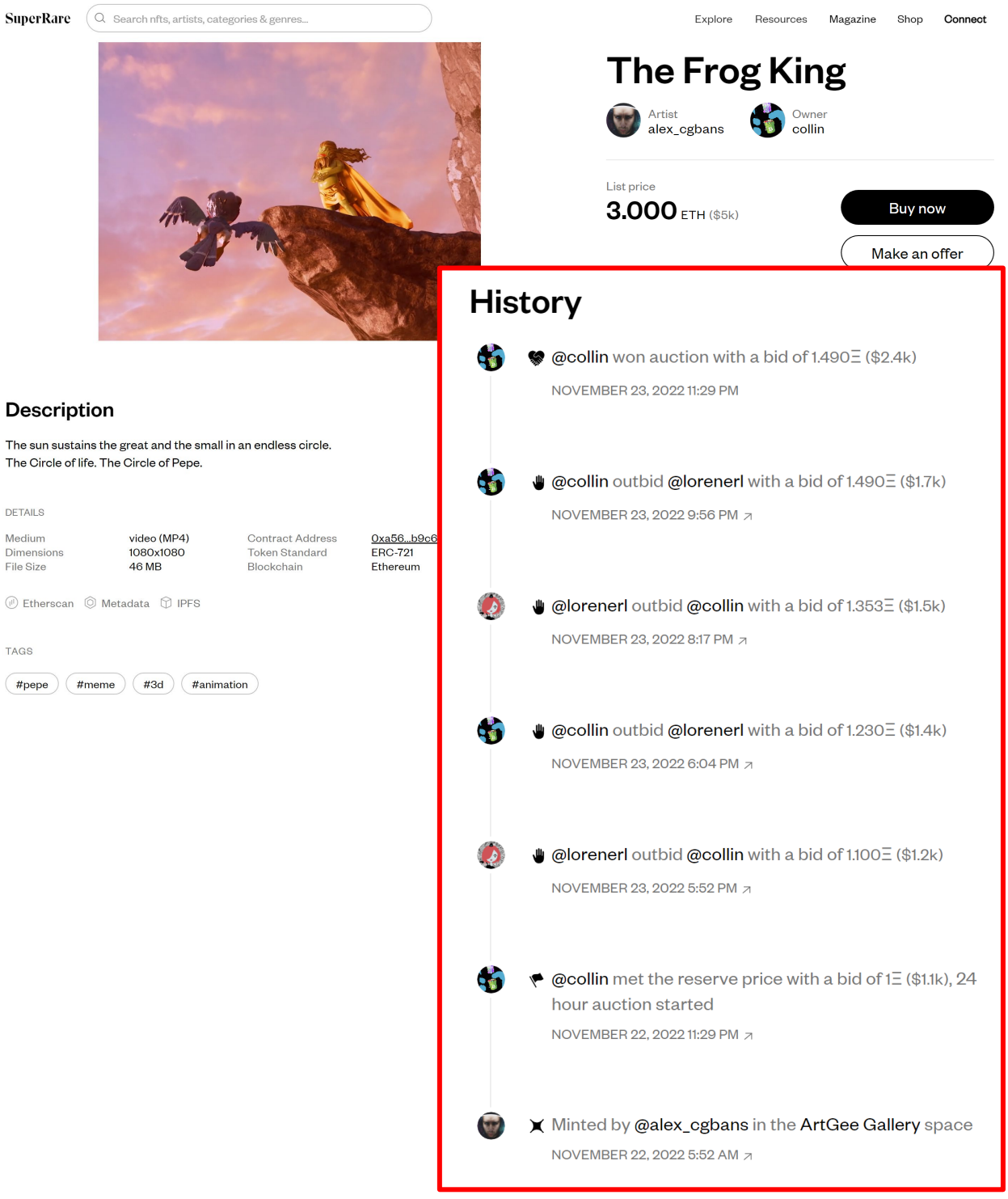
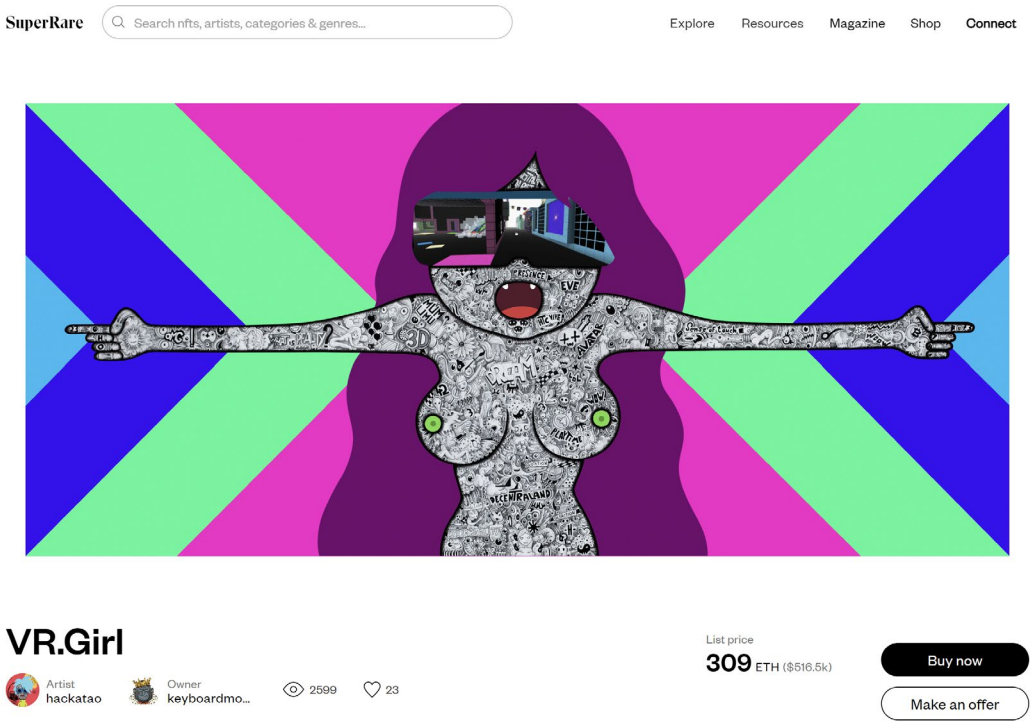


Figure 3. VR Girl by @hackatao. Example of capital appreciation of an NFT over the period 12/2019 to 5/2021.

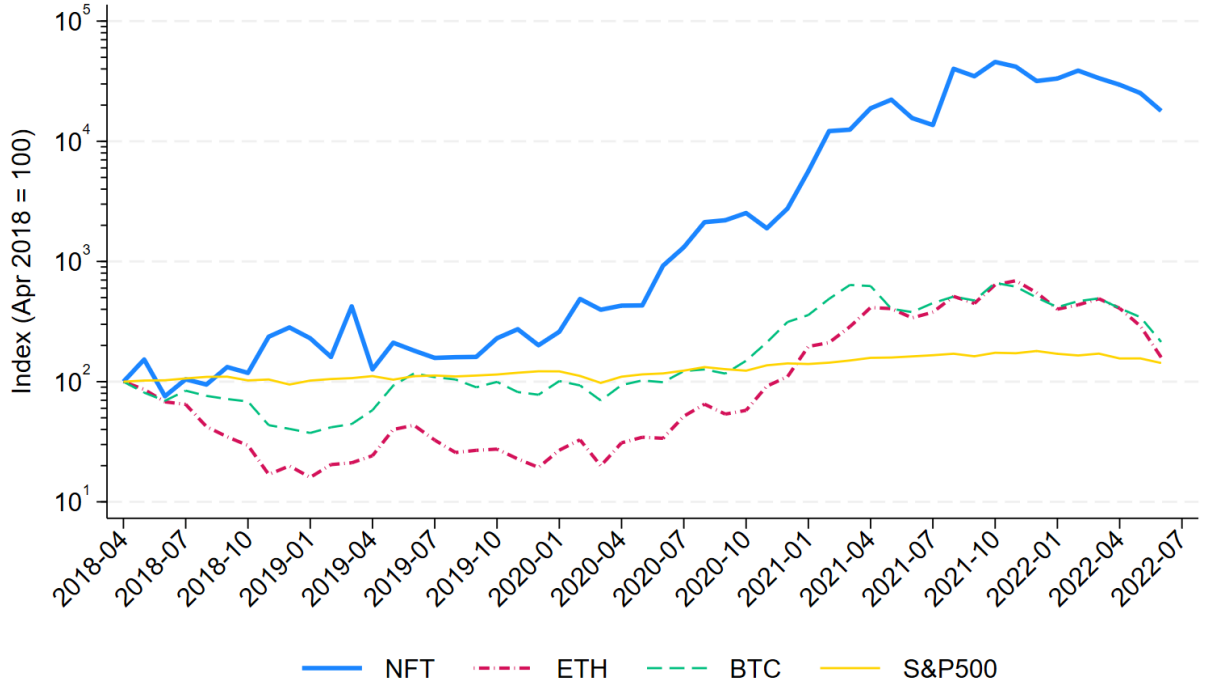


transferred
@6529fundszn3 transferred to @keyboardmonkey3
DECEMBER 24, 2022 6:22 AM
transferred
@keyboardmonkey3 transferred to @6529fundszn3
AUGUST 26, 2022 3:20 PM
transferred
@itsyaboy transferred to @keyboardmonkey3
MAY 25, 2021 1:12 PM
accept bid
@billywhistler accepted an offer of 115Ξ ($297.1k)from @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 NFT price index (thick blue line) estimated with RSR using the whole sample period from April 2018 to June 2022. The price in April 2018 is normalized to 100. The figure also plots the prices of Ethereum (ETH, red dot-dashed line), Bitcoin (BTC, green dashed line), and S&P500 index (thin yellow line) 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 (i.e., using expanding windows), which are called the "up-to" indexes. The light grey lines are period-by-period RSR indexes with the first index ending in April 2019, one year from the beginning of the sample period, and the last index ending in June 2022. The blue solid line connects the ending index levels of the grey lines for every month except that the index levels before April 2019 are from the first "up-to" index ending in April 2019. The red dashed line plots the full-sample RSR index, the same as the NFT index in Figure (a).

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



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

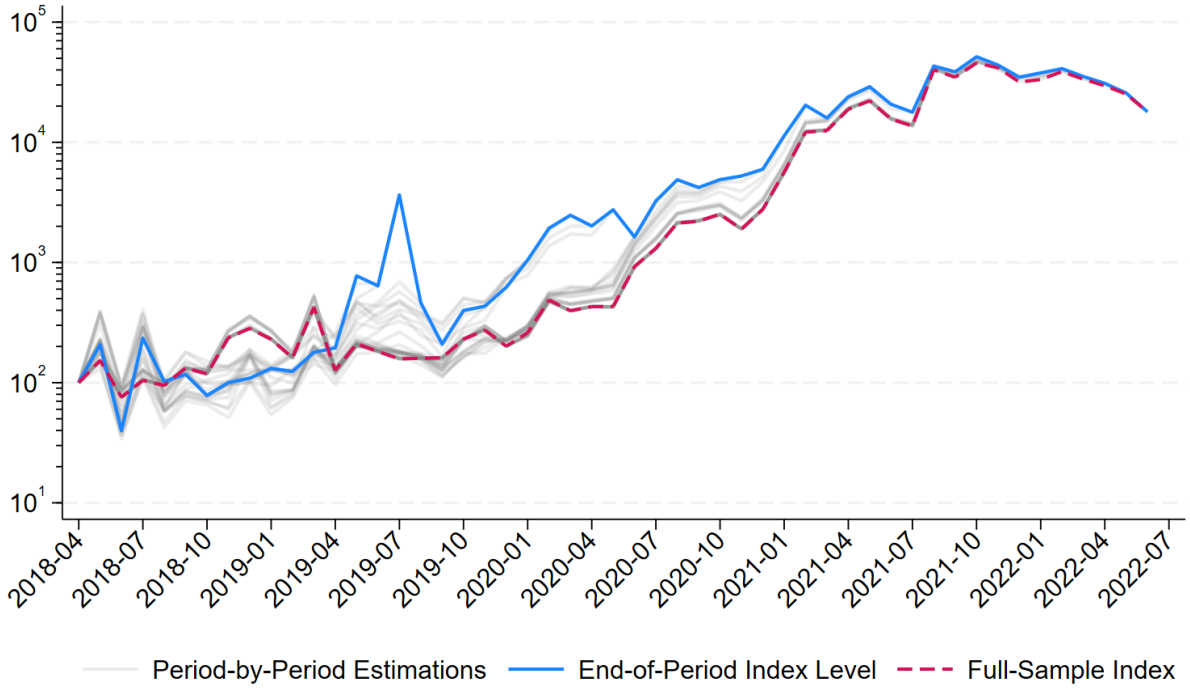
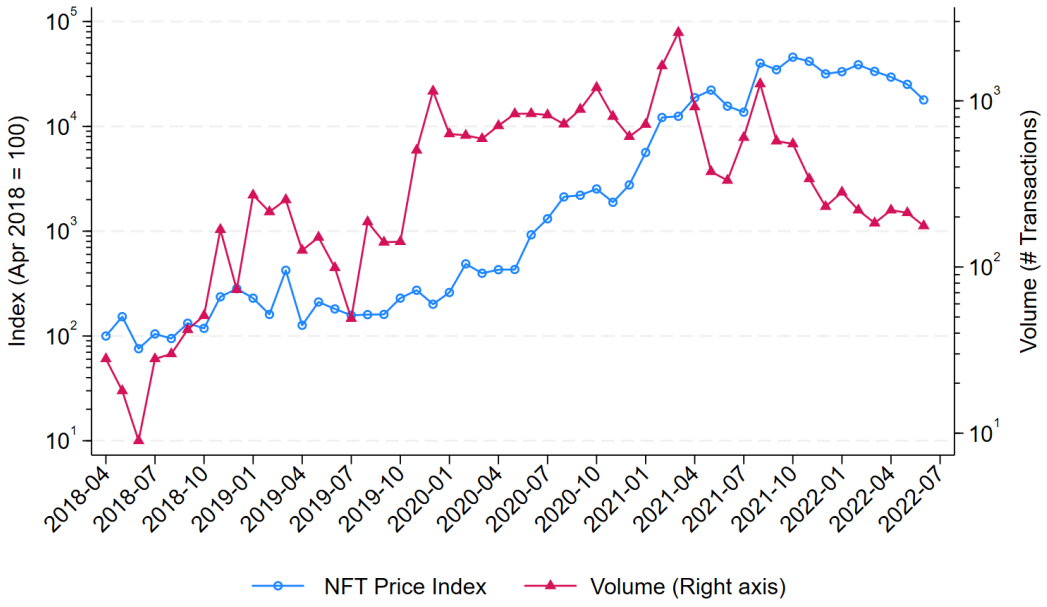


Figure 5. Dynamics between prices and volume.

This figure illustrates the relationship between NFT prices and trading volume. Figure (a) plots the RSR price index (blue line with circle markers, left axis, log scale) and the trading volume (red line with triangle markers, 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}.

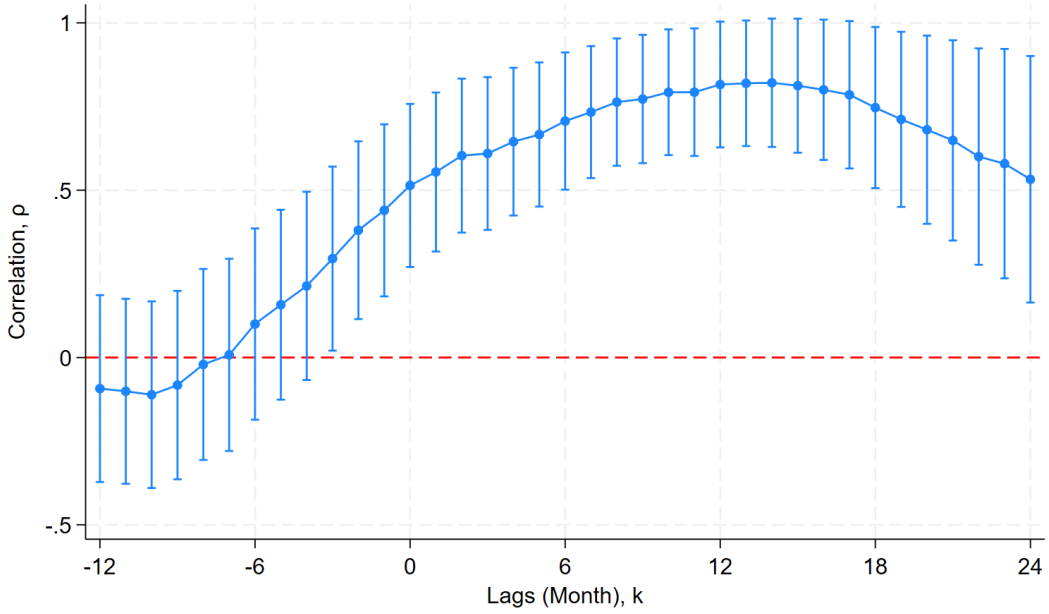


Figure 6. Price dispersion in the market.

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. The solid blue line presents the dispersion of prices denominated in ETH and the dashed red line in USD (right axis).

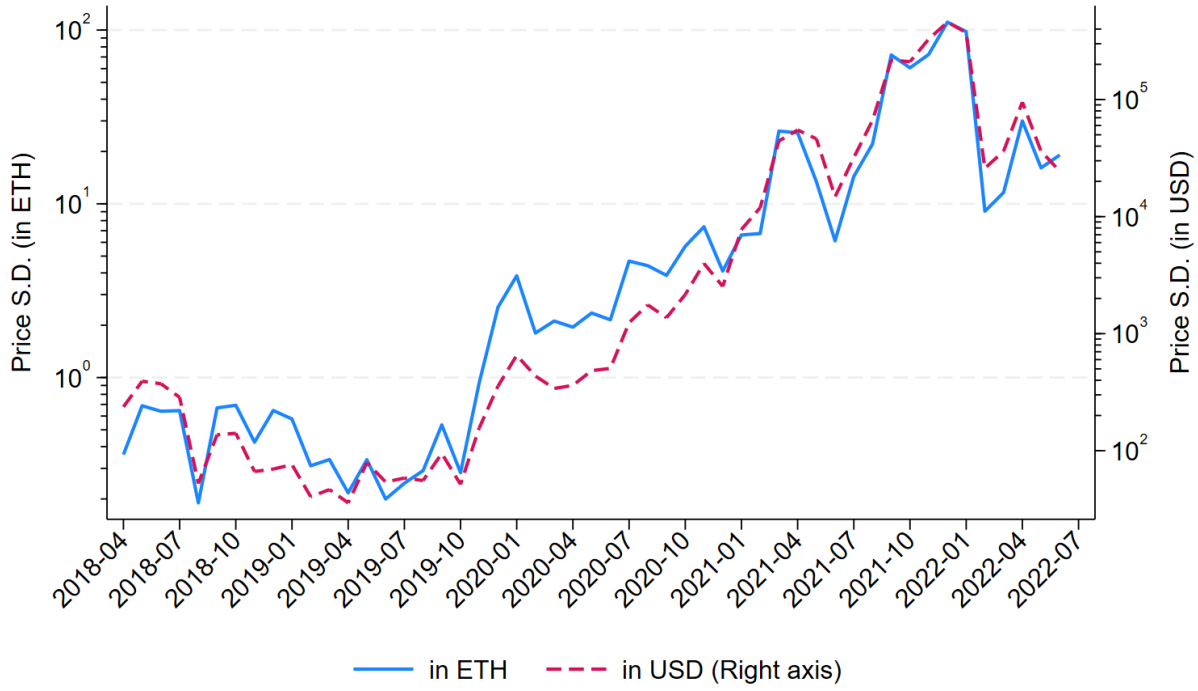
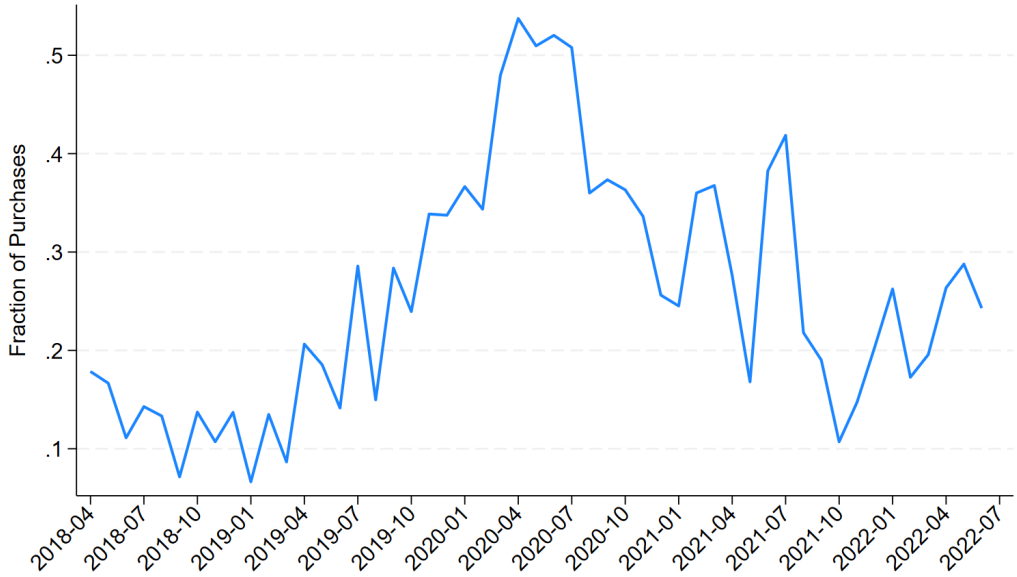


Figure 7. Trading behavior of short-term speculators.

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 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.

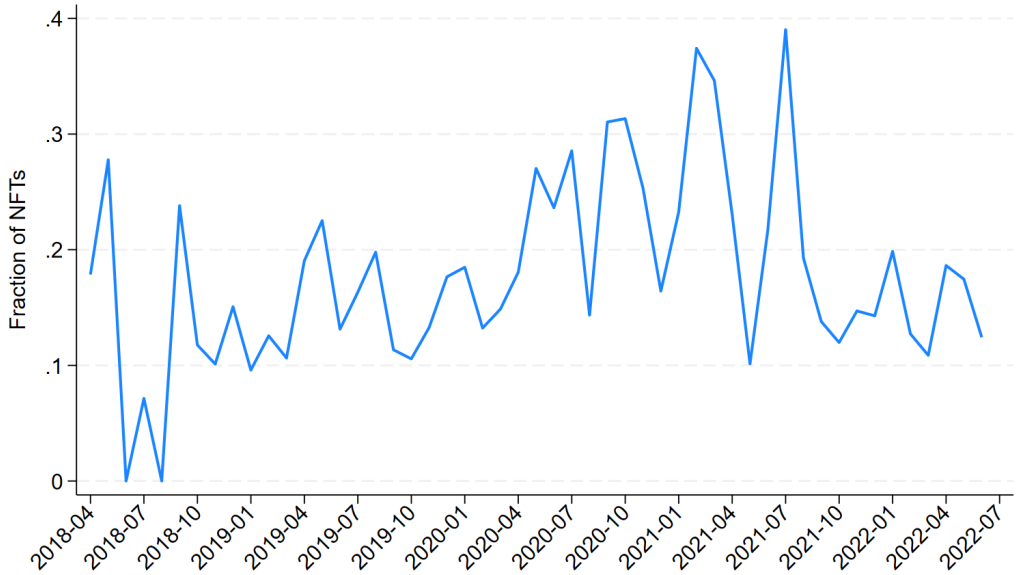
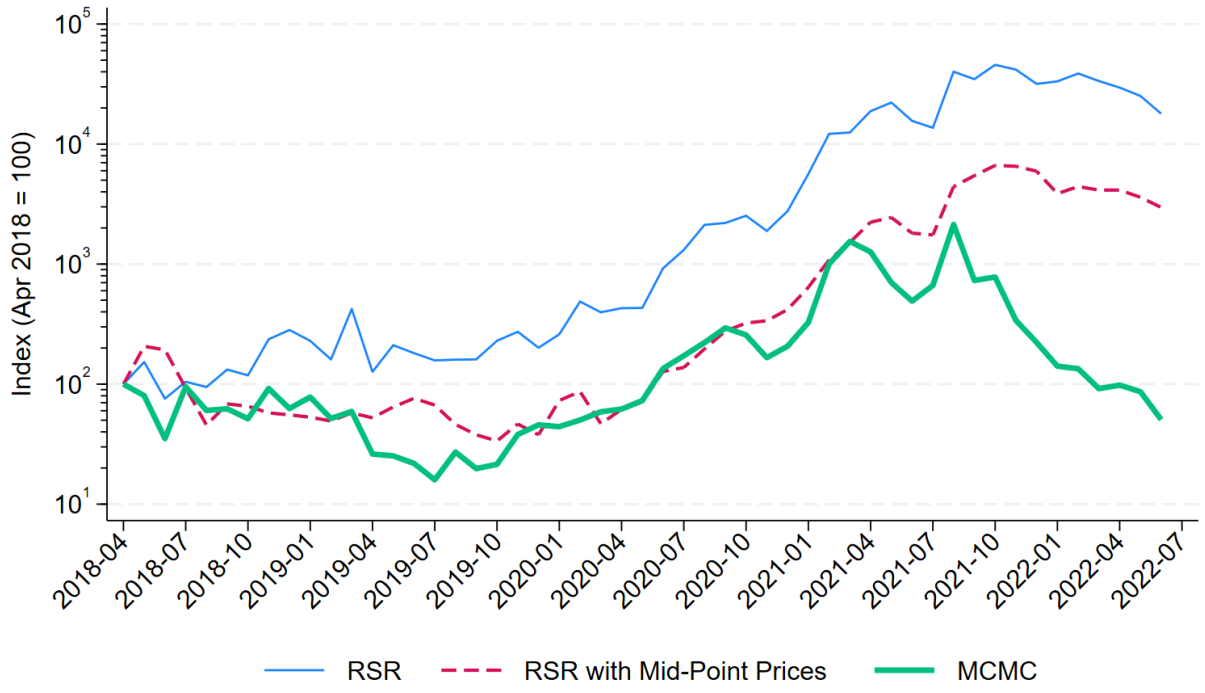


Figure 8. RSR price index with bids and asks and MCMC selection-adjusted price index.

This figure shows the RSR price indexes that account for selection bias in the observed transaction prices. Panel (a) reports the indexes estimated with the full sample. The thin blue line is the standard RSR index as a reference. For the RSR index with mid-point prices (the red dashed line), the observed transaction prices are augmented 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 thick green line shows the selection-adjusted price index estimated by the Markov Chain Monte Carlo (MCMC) algorithm following Kortweg et al. (2016). Panel (b) shows the MCMC selection-adjusted indexes estimated with repeat-sale pairs observed up to each month (i.e., using expanding windows), which are called the "up-to" indexes. The light grey lines are period-by-period "up-to" MCMC indexes with the first index ending in April 2019, one year from the beginning of the sample period, and the last index ending in June 2022. The blue solid line connects the ending index levels of the grey lines for every month except that the index levels before April 2019 are from the first "up-to" index ending in April 2019. The red dashed line plots the full-sample MCMC index, the same as the MCMC index in Figure (a).

(a) Selection-adjusted indexes with the full sample.



(b) MCMC selection-adjusted indexes estimated with repeat-sale pairs observed up to each month.

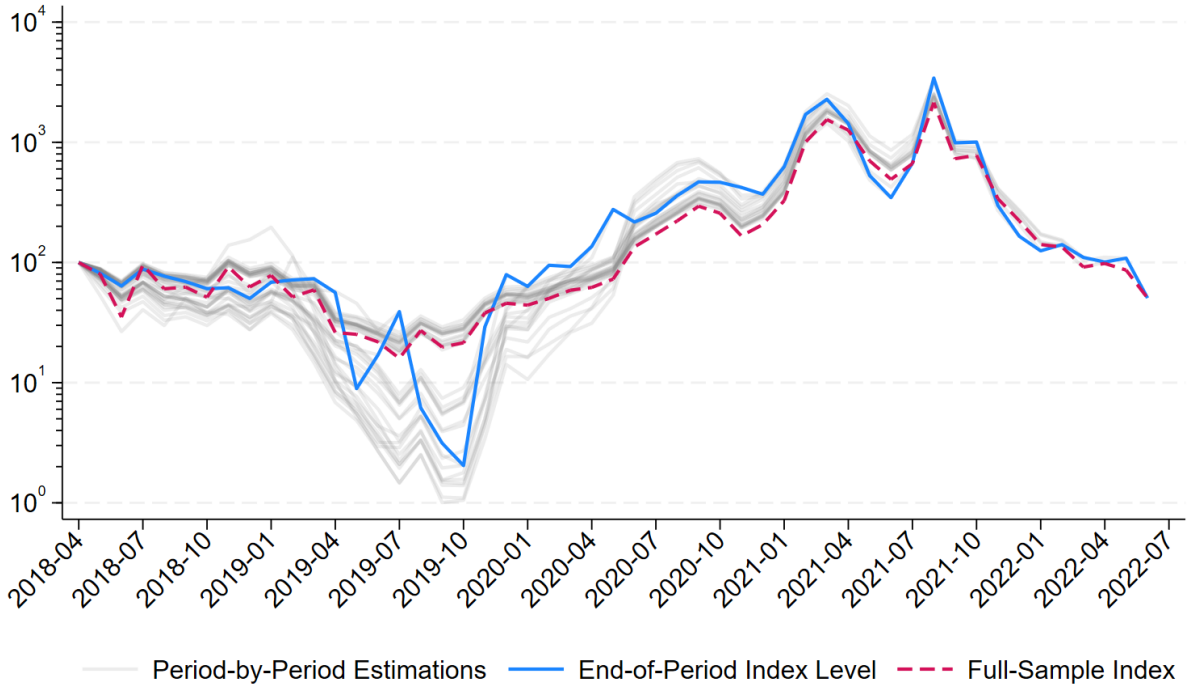


Figure 9. The probability of a sale.

This figure plots the probability of a sale with respect to potential holding returns at different horizons, calculated from the selection equation estimated by the MCMC algorithm in Table 2 Column 1.

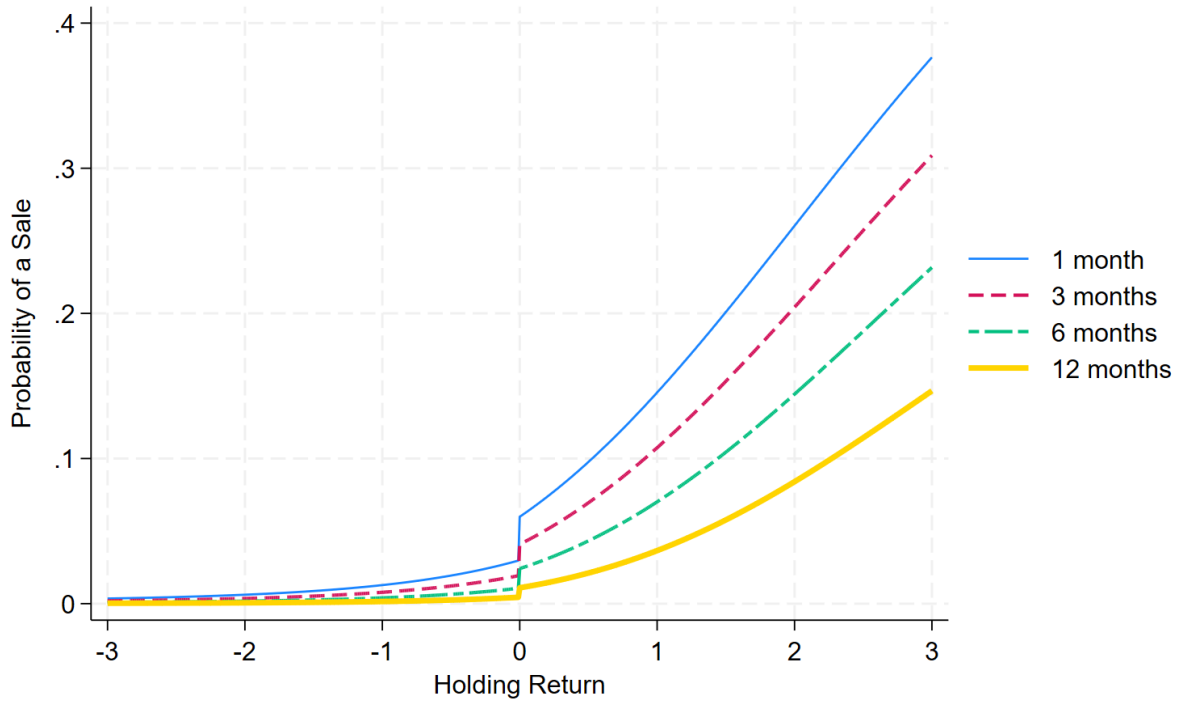
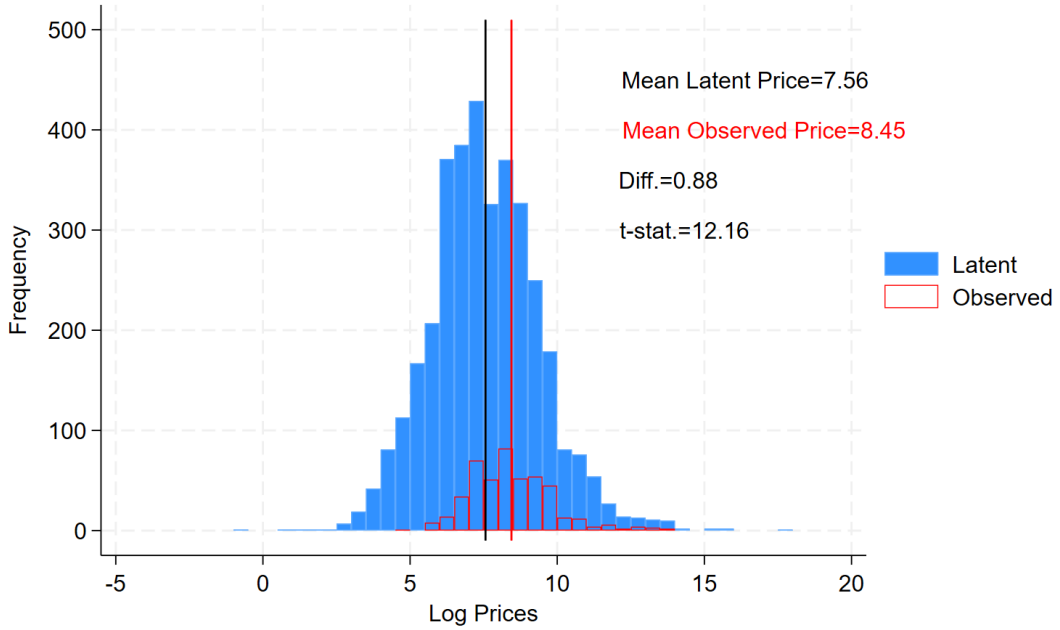


Figure 10. Distribution of latent and observed prices and returns.

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 August 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 August 2021.



(b) Mean differences between observed and latent prices.

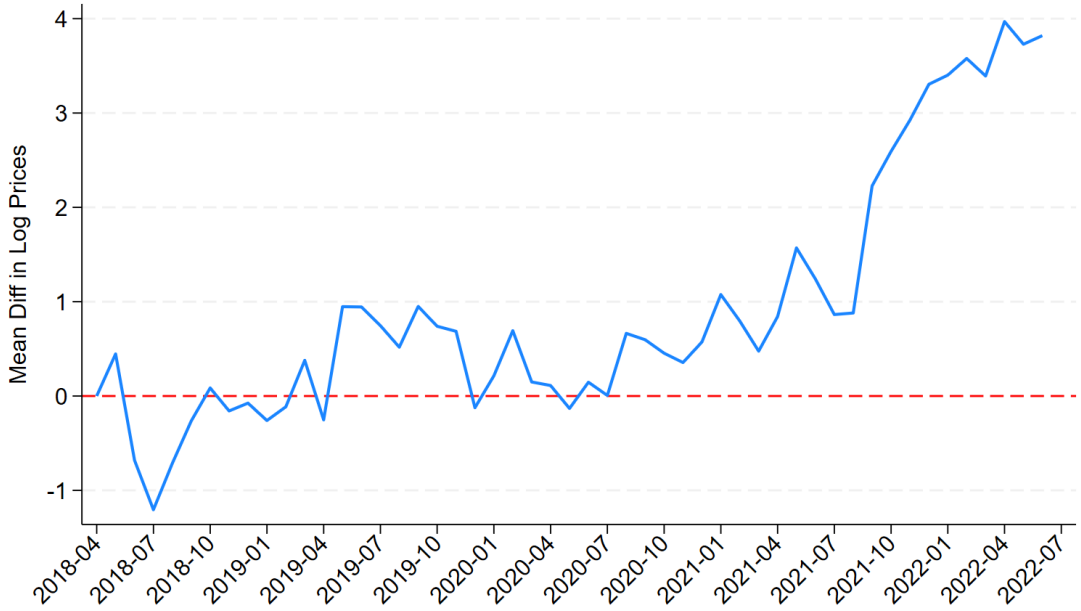


Figure 11. The effective federal funds rate and NFT price index.

This figure shows the NFT price indexes estimated by repeat-sale regression (RSR) and MCMC selection-adjusted algorithm, along with the monthly effective federal fund rate (right axis). The price indexes are normalized to 100 in April 2018.

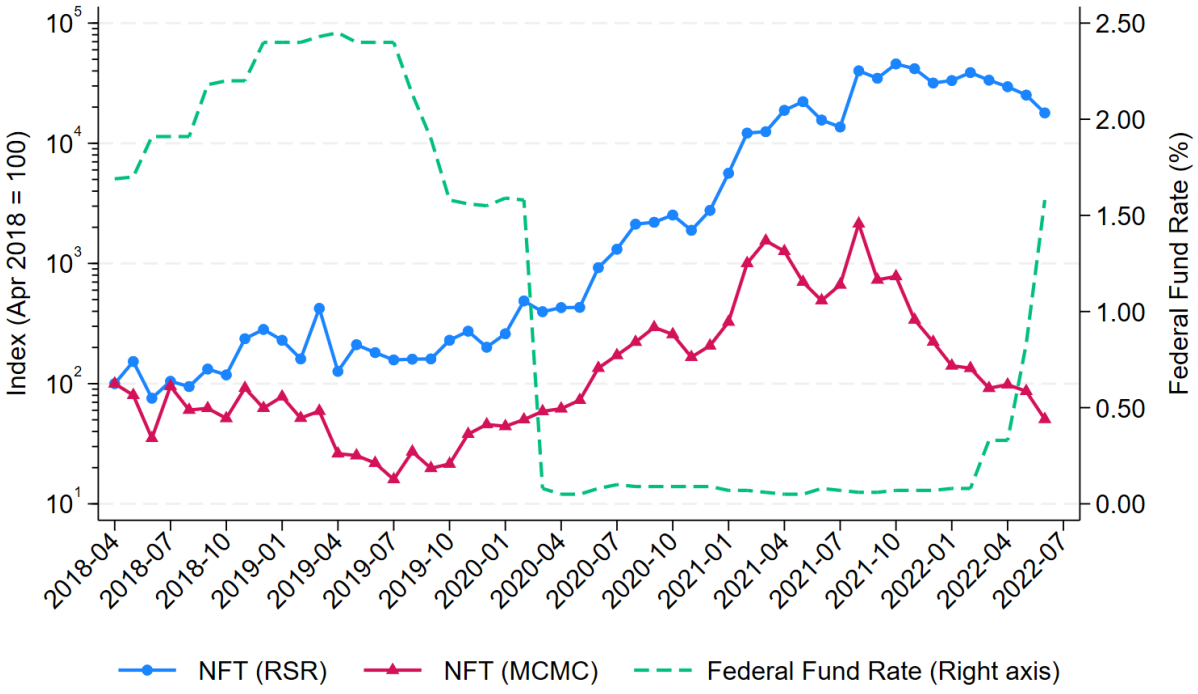


Figure 12. Selection bias as the difference between RSR index returns and MCMC-adjusted index returns.

This figure shows the time series of selection bias used for regressions in Section 5. 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.

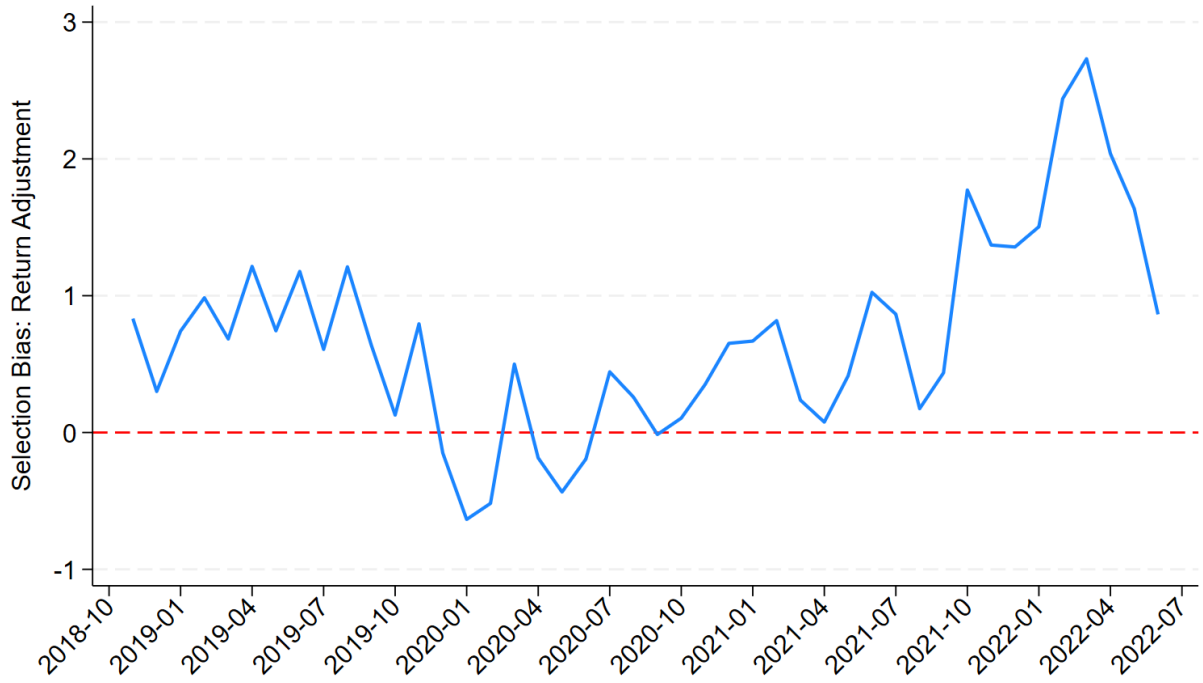
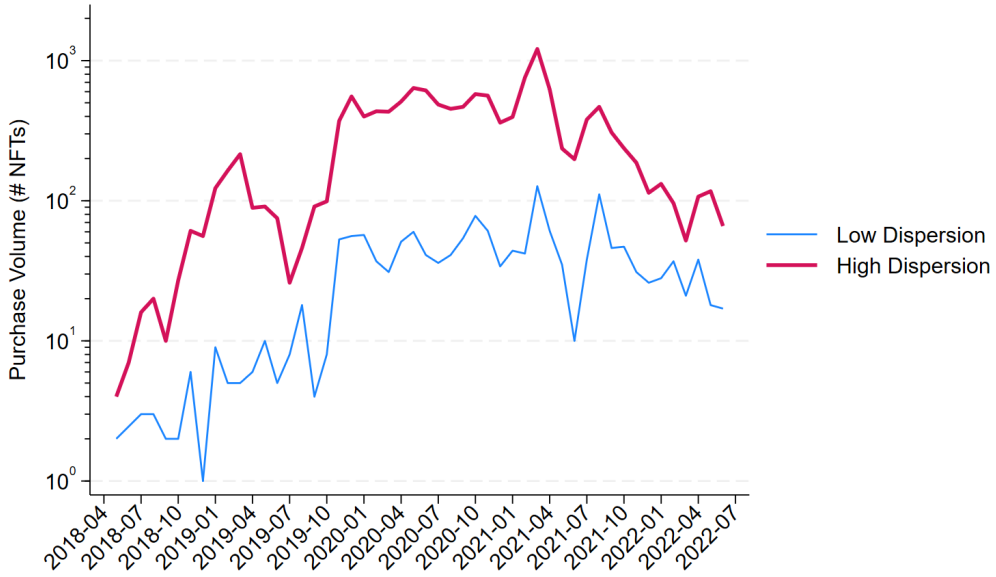


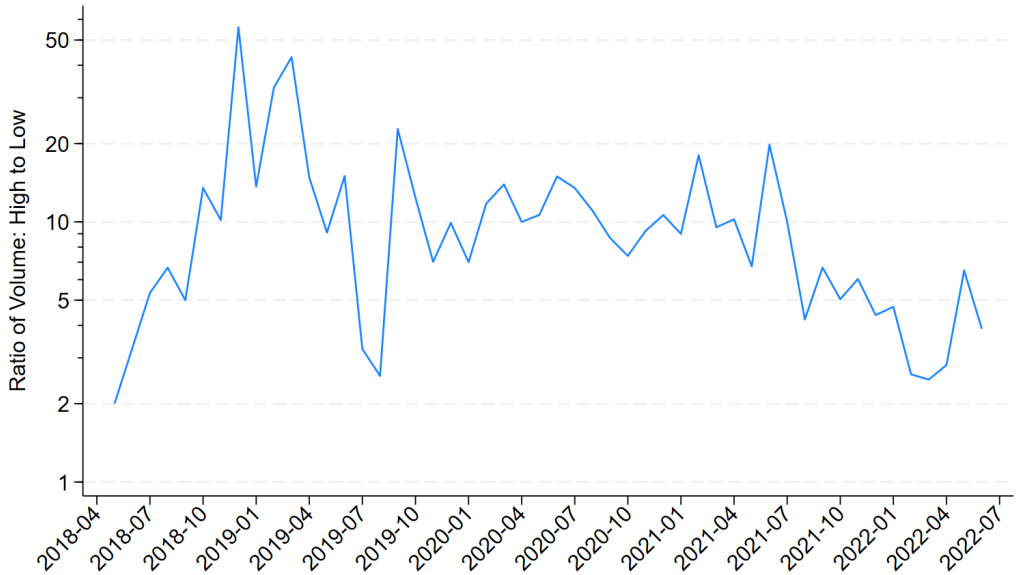
Figure 13. Purchase volume from high- and low-bid dispersion investors over time.

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 with 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: High- to low-bid dispersion investors



Tables

Table 1. Summary Statistics.

This table reports summary statistics of the SuperRare transaction record data. Panel A is about the distribution of asked, bid, and transaction (sale) prices. Prices in US dollars are converted based on the prices of ETH/USD at the end of the transaction date. The asking 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. Panel B summarizes market participants as of the end of the sample period. 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. Panel C summarizes repeat-sale pairs. Buy and sell prices have been converted to US dollars based on ETH prices 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].

Panel A. Event prices.

	P10	P25	P50	P75	P90	Mean	Std. Dev.	Max	N
Ask (ETH)	0.39	0.96	2	5	10	41.93	1,261.16	88,889	14,500
Ask (USD)	106.23	283.03	1,229.0	5,604.15	17,068.48	34,721.2	626,489.50	50,583,875	14,500
Bid (ETH)	0.07	0.2	0.55	1.75	4.24	2.4	15.78	1,630	71,165
Bid (USD)	23.41	76.45	384.39	1,919.53	6,783.39	4,581.65	55,225.49	6,591,705	71,165
Sale (ETH)	0.15	0.3	0.8	2	5.1	3.47	25.87	1,630	24,424
Sale (USD)	33.94	100.5	586.03	2,867.57	9,561.22	7,521.16	89,778.89	6,591,705	24,424

Panel B. Market participants (as of June 30, 2022).

	P10	P25	P50	P75	P90	Mean	Std. Dev.	Max	N
Artists:									
# NFTs create	1	2	6	18	40	17.45	33.18	397	2,018
# NFTs hold	0	0	1	3	6	2.91	8.36	143	2,018
Owners:									
# NFTs hold	1	1	1	2	7	4.5	17.8	509	5,510

Panel C. Repeat-sale pairs.

	Mean	Std. Dev.	Min	P25	Median	P75	Max	N
Sell price (USD)	11,981.94	155,297.84	2.61	313.96	1,090.57	3,805.40	6,591,705.47	4,281
Buy price (USD)	2,251.54	45,909.43	0.04	37.29	112.93	606.08	2,925,565.67	4,281
Log return	2.01	1.52	-1.39	0.85	1.75	2.99	6.64	4,281
Holding period (in months)	8.52	7.56	1.00	3.00	6.00	12.00	40.00	4,281
Buy date	0.50	0.18	0.00	0.40	0.52	0.64	0.97	4,281
Sell date	0.67	0.14	0.06	0.59	0.69	0.77	1.00	4,281

Table 2. Estimated coefficients of the selection equation in MCMC estimations.

This table reports the estimated coefficients of the selection equation (3) in MCMC estimations. Column 1 estimates the baseline model with quadratic terms of holding period returns (r^p) and holding period lengths (T). Columns 2 and 3 add Ethereum returns (ETH Ret) and Google search volume index (SVI) as an additional covariate, respectively. Column 4 includes both covariates. σ is the standard deviation of the residual returns in the law of motion for latent prices (Equation (2)). MCMC standard errors are reported in parentheses. Significance level: *: p<0.1, **: p<0.05, ***: p<0.01.

Specification:	(1) Baseline	(2) Add ETH Ret	(3) Add SVI	(4) Add Both
Constant	-1.781***	-1.830***	-1.853***	-1.878***
	(0.039)	(0.04)	(0.039)	(0.04)
$I(r^p > 0)$	0.326***	0.371***	0.364***	0.371***
	(0.043)	(0.043)	(0.042)	(0.043)
$I(r^p > 0) \cdot r^p$	0.543***	0.511***	0.544***	0.538***
	(0.019)	(0.021)	(0.021)	(0.021)
$I(r^p > 0) \cdot (r^p)^2$	-0.043***	-0.034***	-0.041***	-0.041***
	(0.003)	(0.004)	(0.004)	(0.004)
$I(r^p \leq 0) \cdot r^p$	0.391***	0.329***	0.424***	0.404***
	(0.047)	(0.043)	(0.044)	(0.046)
$I(r^p \leq 0) \cdot (r^p)^2$	0.040***	0.034***	0.038***	0.036***
	(0.007)	(0.007)	(0.007)	(0.007)
T	-0.103***	-0.102***	-0.106***	-0.106***
	(0.004)	(0.004)	(0.004)	(0.004)
T^2	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
ETH Ret		0.112***		0.132***
		(0.04)		(0.041)
SVI			0.006***	0.006***
			(0.001)	(0.001)
σ	0.646***	0.641***	0.655***	0.654***
	(0.008)	(0.007)	(0.008)	(0.008)

Table 3. Regressions of asking prices on potential NFT returns.

This table reports the estimations of equation (6), 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***			
	(0.021)	(0.020)	(0.021)			
$1(PotentialRet_{i,t} < 0)$	0.275***	0.232***	0.269***	0.317***	0.332***	0.339***
	(0.077)	(0.084)	(0.084)	(0.079)	(0.098)	(0.096)
$PotentialRet_{i,t}^{+}$	-0.217***	-0.454***	-0.354***	-0.063***	-0.127**	-0.124
	(0.022)	(0.044)	(0.090)	(0.016)	(0.055)	(0.110)
${PotentialRet_{i,t}^{+}}^{2}$		0.060***	0.029*		0.017	0.003
		(0.010)	(0.017)		(0.013)	(0.015)
$PotentialRet_{i,t}^{-}$	-0.267	0.015	0.013	-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			-0.026			0.000
			(0.022)			(0.024)
T^2			0.002*			0.001
			(0.001)			(0.001)
Constant	1.337***	1.424***	1.432***	1.562***	1.585***	1.583***
	(0.018)	(0.024)	(0.027)	(0.018)	(0.028)	(0.029)
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 4. 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. All specifications include past 6-month Ethereum (ETH) lagged returns as covariates. 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)	(7)	(8)
	Dependent variable: $Log(Volume_t)$							
Return Estimated from Index:	RSR	RSR	MCMC	MCMC	Up-to RSR	Up-to RSR	Up-to MCMC	Up-to MCMC
Past Log Return	0.600***		0.484***		0.379		0.275***	
	(0.214)		(0.132)		(0.226)		(0.041)	
Past Log Return^{+}		1.223**		1.686***		1.163**		0.543***
		(0.479)		(0.582)		(0.461)		(0.192)
Past Log Return^{-}		-0.411		-0.363		-1.394		-0.097
		(0.980)		(0.373)		(0.920)		(0.244)
ETH Return	0.156	0.001	0.194	-0.017	0.069	-0.228	0.232	0.397*
	(0.231)	(0.262)	(0.192)	(0.252)	(0.315)	(0.339)	(0.164)	(0.212)
Constant	5.455***	4.677***	5.872***	4.667***	5.560***	4.485***	5.839***	5.014***
	(0.239)	(0.544)	(0.143)	(0.567)	(0.201)	(0.549)	(0.148)	(0.482)
Observations	44	44	44	44	39	39	39	39
Adjusted R-squared	0.357	0.397	0.505	0.453	0.086	0.358	0.455	0.361

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable: $Log(Volume_t)$ from new buyers							
Return Estimated from Index:	RSR	RSR	MCMC	MCMC	Up-to RSR	Up-to RSR	Up-to MCMC	Up-to MCMC
Past Log Return	0.599***		0.456***		0.339		0.239***	
	(0.213)		(0.128)		(0.234)		(0.054)	
Past Log Return^{+}		1.065**		1.517***		0.956**		0.437**
		(0.425)		(0.508)		(0.393)		(0.188)
Past Log Return^{-}		-0.098		-0.288		-0.827		-0.032
		(0.826)		(0.320)		(0.803)		(0.217)
ETH Return	0.217	0.104	0.269	0.069	0.183	-0.062	0.332	0.461**
	(0.226)	(0.224)	(0.201)	(0.241)	(0.329)	(0.295)	(0.224)	(0.202)
Constant	4.170***	3.585***	4.585***	3.552***	4.299***	3.458***	4.549***	3.917***
	(0.232)	(0.463)	(0.149)	(0.485)	(0.249)	(0.476)	(0.163)	(0.436)
Observations	44	44	44	44	39	39	39	39
Adjusted R-squared	0.291	0.344	0.371	0.397	0.076	0.281	0.288	0.285

Table 5. 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)
	Dependent Variable: Log(Volume)			
Realized Return Mean	0.626**			
	(0.243)			
Realized Return Max		0.242***		
		(0.063)		
Realized Return Min			0.347	
			(0.511)	
Monthly Realized Return Mean				5.232***
				(1.186)
Constant	5.082***	4.670***	6.330***	4.184***
	(0.387)	(0.312)	(0.584)	(0.442)
Observations	44	44	44	44
Adjusted R-squared	0.226	0.218	-0.004	0.242

Table 6. Interaction between extrapolation and selection bias.

This table reports regression results for equation (8), testing the interactions between extrapolation and selection bias. In the first three columns, $LogUnbiasedReturn_{t-1}$ are the past six-month real market returns calculated with the MCMC selection-adjusted index. $SelectionBias$ is defined as the difference between the standard RSR index return and the MCMC index return using the full sample, both in the past six-month horizon. Columns 4 to 6 repeat the regressions in the first three columns using "up-to" MCMC index returns for $LogUnbiasedReturn_{t-1}$ and using $SelectionBias$ that is defined as the difference between the "up-to" RSR index return and the MCMC index return, both in the past six-month horizon. "Impact Diff." tests the difference of the interaction terms, that is, the difference of selection bias's impacts on positive and negative true market returns. All specifications include past 6-month Ethereum returns (ETH Return) as covariates. Newey-West autocorrelation adjusted standard errors are reported in parentheses. Significance level: *: p<0.1, **: p<0.05, ***: p<0.01.

Return type:	(1)	(2)	(3)	(4)	(5)	(6)
	Full-sample index returns			"Up-to" index returns		
	Dependent Variable: $Log(Volume_t)$					
$LogUnbiasedReturn_{t-1}$	0.484***	0.684***		0.275***	0.324***	
	(0.132)	(0.151)		(0.041)	(0.102)	
$LogUnbiasedReturn^{+}_{t-1}$			0.703**			0.203**
			(0.267)			(0.084)
$LogUnbiasedReturn^{-}_{t-1}$			0.665***			0.293**
			(0.170)			(0.111)
$SelectionBias_{t-1}$		-0.496**	-0.497**		-0.792***	-0.584**
		(0.213)	(0.219)		(0.253)	(0.255)
$LogUnbiasedReturn^{+}_{t-1} \times SelectionBias_{t-1}$		0.676*	0.683*		0.451***	0.311**
		(0.350)	(0.364)		(0.155)	(0.118)
$LogUnbiasedReturn^{-}_{t-1} \times SelectionBias_{t-1}$		-0.409***	-0.404***		-0.289**	-0.250**
		(0.140)	(0.128)		(0.136)	(0.114)
$ETH\ Return_{t-1}$	0.194	0.202	0.196	0.232	-0.002	0.117
	(0.192)	(0.191)	(0.201)	(0.164)	(0.179)	(0.187)
Constant	5.872***	5.872***	5.860***	5.839***	6.078***	6.130***
	(0.143)	(0.178)	(0.247)	(0.148)	(0.213)	(0.238)
Impact Diff.		1.085**	1.087**		0.740***	0.561***
		(0.460)	(0.470)		(0.259)	(0.179)
Observations	44	43	43	39	38	38
Adjusted R-squared	0.505	0.560	0.548	0.455	0.574	0.575

Table 7. 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. $SelectionBias$ is defined as the difference between the standard RSR index return and the MCMC index return, both in the past six-month horizon, and is normalized. All specifications include past 6-month Ethereum returns (ETH Return) as covariates. Newey-West autocorrelation adjusted standard errors are reported in parentheses. Significance level: *: p<0.1, **: p<0.05, ***: p<0.01.

Investor type:	(1)	(2)	(3)	(4)	(5)	(6)
	Low dispersion			High dispersion		
	Dependent Variable: $Log(Volume_t)$					
$LogUnbiasedReturn_{t-1}$	0.238	0.441**		0.601***	0.736***	
	(0.206)	(0.209)		(0.121)	(0.153)	
$LogUnbiasedReturn^{+}_{t-1}$			0.596			0.508**
			(0.379)			(0.227)
$LogUnbiasedReturn^{-}_{t-1}$			0.292			0.954***
			(0.281)			(0.147)
$SelectionBias_{t-1}$		-0.450	-0.457		-0.485**	-0.475**
		(0.394)	(0.414)		(0.230)	(0.206)
$LogUnbiasedReturn^{+}_{t-1} \times SelectionBias_{t-1}$		0.378	0.434		0.585*	0.504
		(0.471)	(0.488)		(0.323)	(0.332)
$LogUnbiasedReturn^{-}_{t-1} \times SelectionBias_{t-1}$		-0.474*	-0.431*		-0.354**	-0.418***
		(0.243)	(0.240)		(0.146)	(0.127)
$ETH\ Return_{t-1}$	0.509	0.489	0.444	0.120	0.107	0.172
	(0.322)	(0.314)	(0.318)	(0.195)	(0.183)	(0.190)
Constant	3.026***	2.903***	2.805***	5.299***	5.319***	5.462***
	(0.247)	(0.288)	(0.394)	(0.137)	(0.158)	(0.186)
Observations	44	43	43	44	43	43
Adjusted R-squared	0.291	0.337	0.322	0.621	0.656	0.654

Table 8. 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. 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. $SelectionBias$ is defined as the difference between the standard RSR index return and the MCMC index return, both in the past six-month horizon, and is normalized. All specifications include past 6-month Ethereum returns (ETH Return) as covariates. Cluster robust standard errors clustered by investors are reported in parentheses. Significance level: *: p<0.1, **: p<0.05, ***: p<0.01.

Dependent variable:	(1)	(2)	(3)	(4)	(5)	(6)
	$Log(Buy_{i,t}+1)$			$1(Buy_{i,t}>0)$		
$LogUnbiasedReturn_{t-1}$	0.210***			0.751***		
	(0.013)			(0.044)		
$LogUnbiasedReturn^{+}_{t-1}$		0.166***	0.167***		0.562***	0.598***
		(0.018)	(0.019)		(0.067)	(0.074)
$LogUnbiasedReturn^{-}_{t-1}$		0.249***	0.219***		0.949***	0.803***
		(0.019)	(0.030)		(0.073)	(0.130)
$SelectionBias_{t-1}$			-0.171***			-0.573***
			(0.017)			(0.070)
$LogUnbiasedReturn^{+}_{t-1} \times SelectionBias_{t-1}$			0.157***			0.516***
			(0.019)			(0.074)
$LogUnbiasedReturn^{-}_{t-1} \times SelectionBias_{t-1}$			-0.058***			-0.183***
			(0.012)			(0.060)
$ETH\ Return_{t-1}$	-0.269***	-0.261***	-0.247***	-0.853***	-0.821***	-0.786***
	(0.021)	(0.020)	(0.020)	(0.059)	(0.059)	(0.059)
Investor fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12,977	12,977	12,939	12,950	12,950	12,893

Table 9. Interactions between extrapolation and selection bias at the investor-level: heterogeneity over investor ability.

This table examines the interactions between extrapolation and selection bias using investor-level data 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 MCMC selection-adjusted index returns. $SelectionBias_{t-1}$ is the difference between the standard RSR index return and the MCMC index return, both in the six-month horizon, and is normalized. All specifications include past 6-month Ethereum returns (ETH Return) as covariates. 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:	$Log(Buy_{i,t}+1)$	$1(Buy_{i,t}>0)$	$Log(Buy_{i,t}+1)$	$1(Buy_{i,t}>0)$
$LogUnbiasedReturn^{+}_{t-1}$	0.163***	0.581***	0.155***	0.826***
	(0.020)	(0.081)	(0.020)	(0.101)
$LogUnbiasedReturn^{-}_{t-1}$	0.192***	0.826***	0.242***	1.293***
	(0.028)	(0.136)	(0.027)	(0.156)
$SelectionBias_{t-1}$	-0.136***	-0.556***	-0.177***	-0.874***
	(0.015)	(0.073)	(0.016)	(0.095)
$LogUnbiasedReturn^{+}_{t-1} \times SelectionBias_{t-1}$	0.144***	0.508***	0.153***	0.780***
	(0.017)	(0.079)	(0.020)	(0.103)
$LogUnbiasedReturn^{-}_{t-1} \times SelectionBias_{t-1}$	-0.042***	-0.203***	-0.092***	-0.360***
	(0.012)	(0.060)	(0.011)	(0.074)
$Ability_{i,t-1}$	0.092***	0.254***	0.132	-0.025
	(0.024)	(0.071)	(0.092)	(0.282)
$LogUnbiasedReturn^{+}_{t-1} \times Ability_{i,t-1}$	0.014	-0.020	-0.071***	-0.389***
	(0.023)	(0.074)	(0.023)	(0.110)
$LogUnbiasedReturn^{-}_{t-1} \times Ability_{i,t-1}$	-0.006	-0.105	-0.064**	-0.643***
	(0.020)	(0.081)	(0.025)	(0.169)
$SelectionBias_{t-1} \times Ability_{i,t-1}$	-0.031	0.103	-0.006	0.036
	(0.023)	(0.066)	(0.017)	(0.110)
$LogUnbiasedReturn^{+}_{t-1} \times SelectionBias_{t-1} \times Ability_{i,t-1}$	0.053**	-0.017	-0.032	-0.194
	(0.024)	(0.079)	(0.020)	(0.121)
$LogUnbiasedReturn^{-}_{t-1} \times SelectionBias_{t-1} \times Ability_{i,t-1}$	0.033**	0.165**	0.025**	0.322***
	(0.016)	(0.065)	(0.012)	(0.085)
$ETH\ Return_{t-1}$	-0.217***	-0.762***	-0.231***	-0.985***
	(0.017)	(0.061)	(0.021)	(0.070)
Investor fixed effects	Yes	Yes	Yes	Yes
Observations	12,939	12,893	11,042	10,926

Table 10. 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, $Log(\#buy + 1)$), 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 is 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 the 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 + 1)			-0.278***	
			(0.08)	
Avg Hold Time				-0.439***
				(0.02)
Observations	4,281	4,281	4,281	4,281
Adjusted 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***			
	(0.01)			
Missing BidDisper Dummy	0.009			
	(0.04)			
BuyLow		0.204***		
		(0.07)		
Missing BuyLow Dummy		0.079		
		(0.05)		
Log(# buy + 1)			-0.007	
			(0.009)	
Avg Hold Time				-0.067***
				(0.003)
Observations	23,610	23,610	23,610	23,610
Adjusted R-Squared	0.94	0.94	0.94	0.94

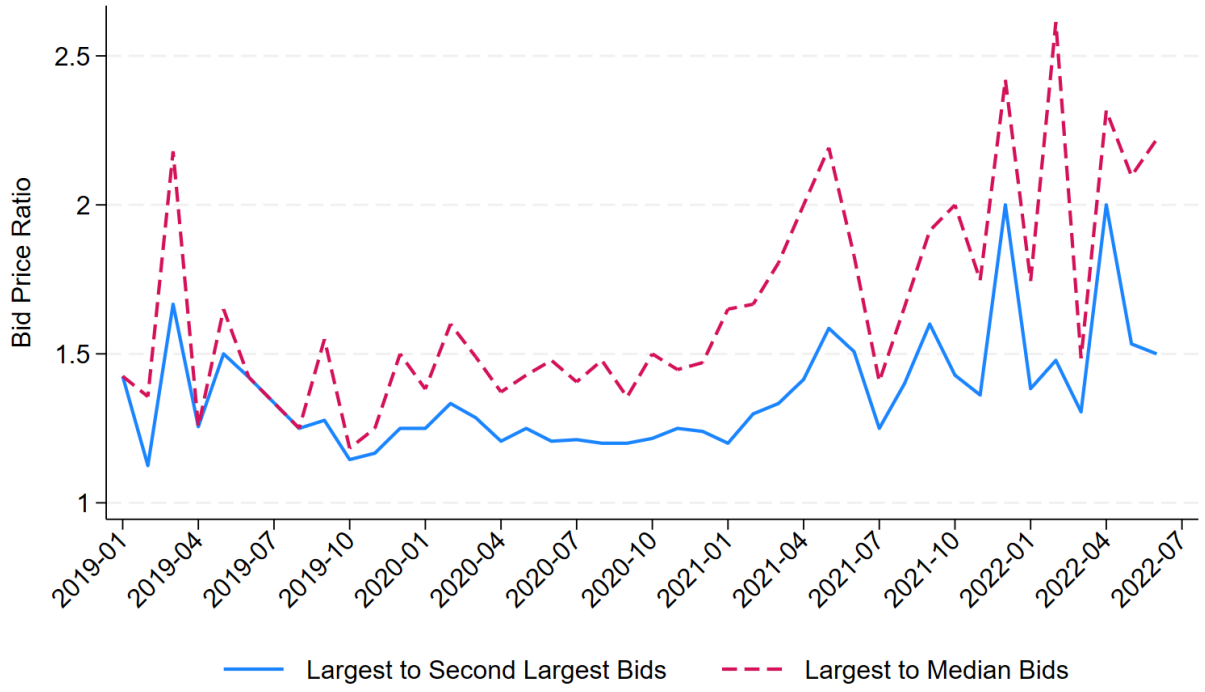
Online Appendix

"Selection-Neglect in the NFT Bubble"

Appendix A. Additional Figures and Tables

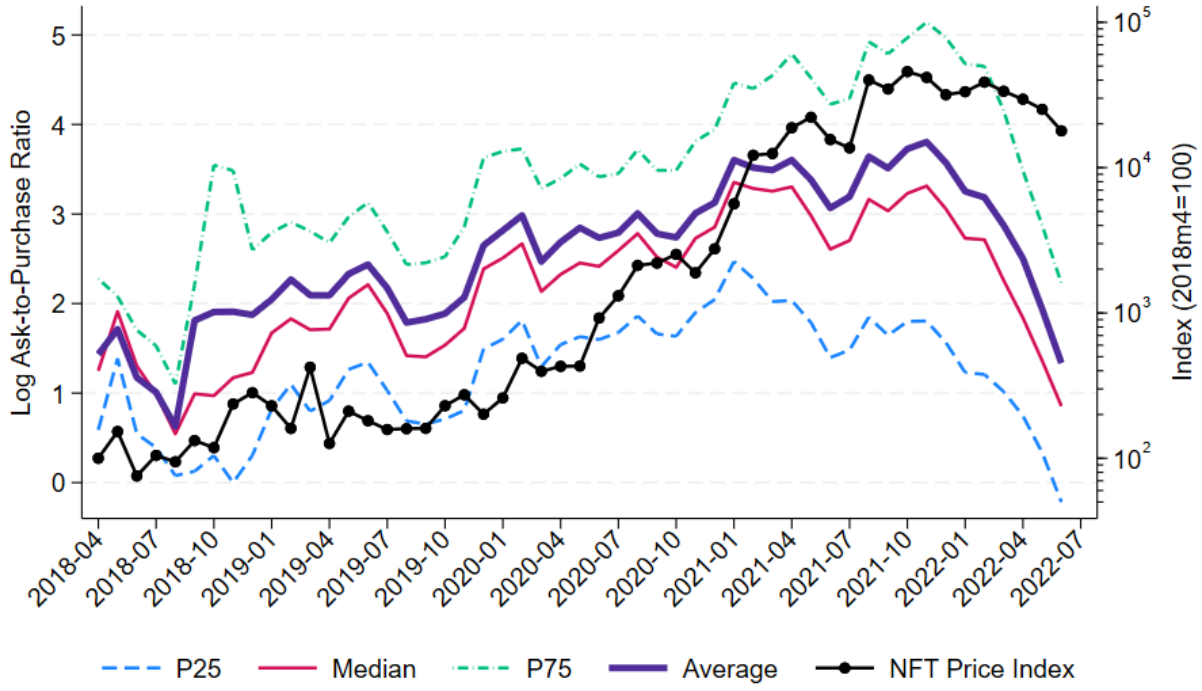
Appendix Figure A1. Dispersion in prices offered to the same NFT.

This figure presents the time series of the median dispersion of bid prices offered to the same NFT during the sample period. The blue solid line calculates the ratio of the largest (the winning bid) to the second largest offer while the red dashed line represents the ratio of the largest to the median offer. Both measures are first calculated for each NFT-month and then taken median cross-sectionally.



Appendix Figure A2. The mean, median, and quantiles of the ratio of asking price to purchase price.

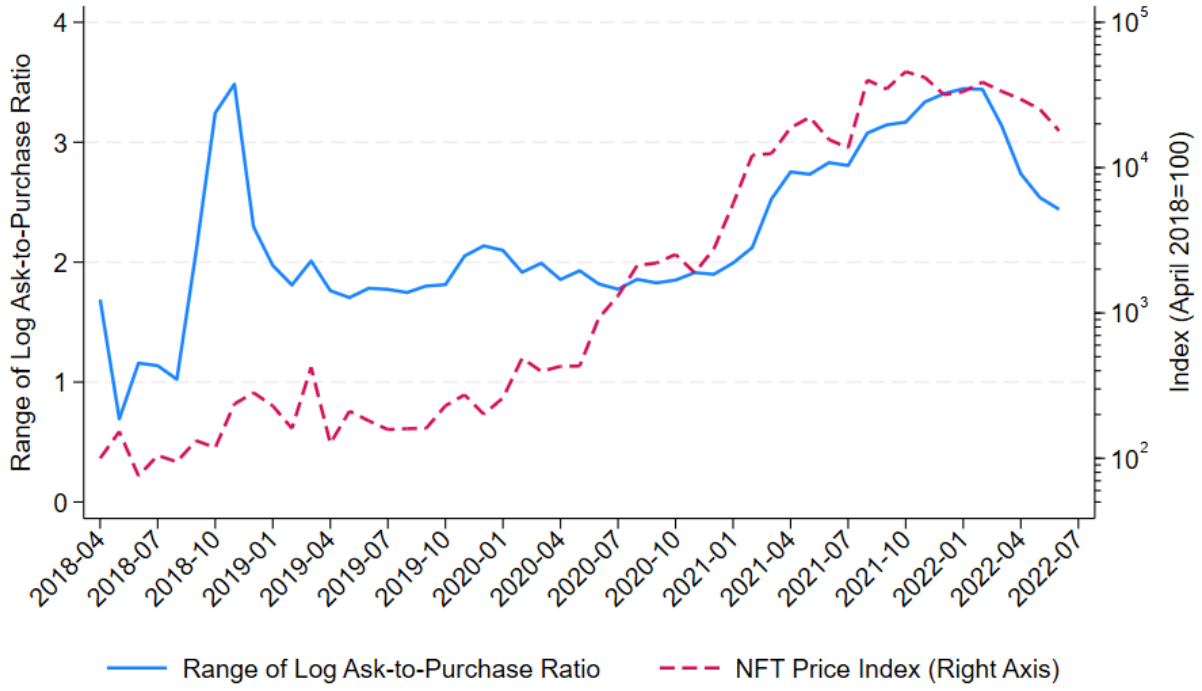
This figure shows the distribution of the ratio of the asking price (listing price) to the purchase price of NFTs. For each listed NFT in each month, we calculate the log difference between the listing price and the previous purchase price of the NFT, which is a measure of the seller's expected returns. The figure plots the 25th percentile, median, mean, and 75th percentile of the distribution on each cross-section, along with the RSR NFT price index (right axis).



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

Appendix Figure A3. The interquartile range of the ask-to-purchase price ratio.

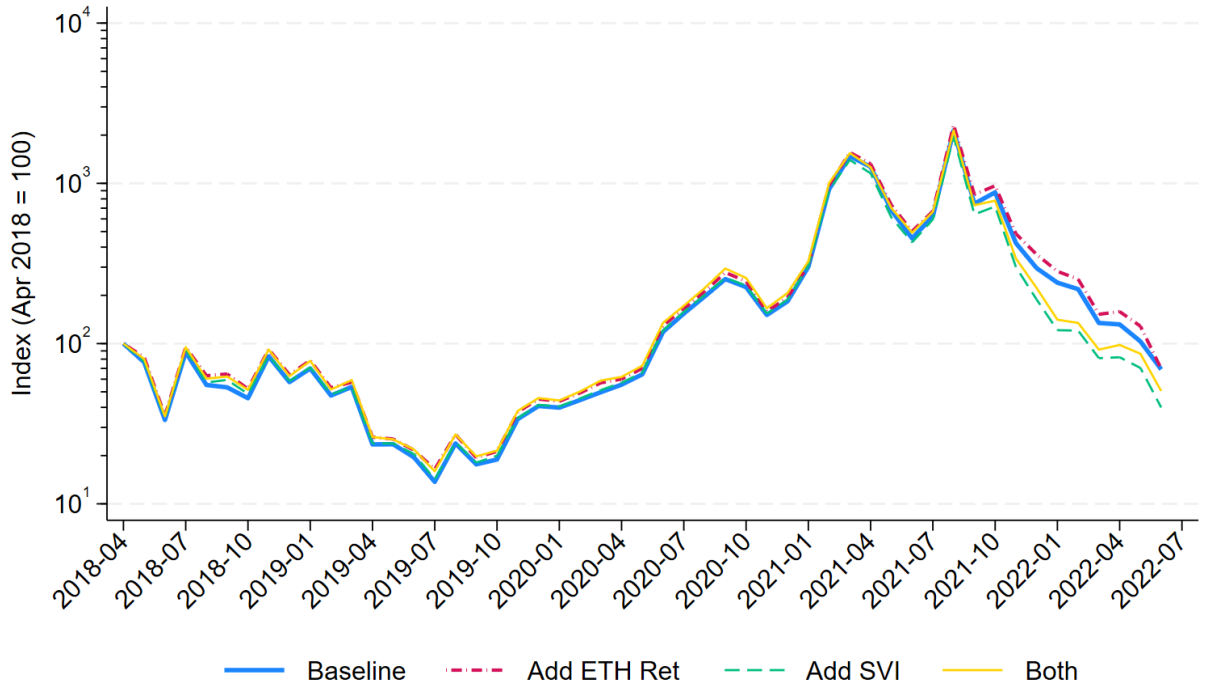
This figure shows the dispersion of the ratio of the asking price (listing price) to the purchase price of NFTs. For each listed NFT in each month, we calculate the log difference between the listing price and the previous purchase price of the NFT, which is a measure of the seller's expected returns. The figure plots the interquartile range of the distribution on each cross-section, along with the RSR NFT price index (right axis).



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

Appendix Figure A4. MCMC selection-adjusted indexes using different selection equation specifications.

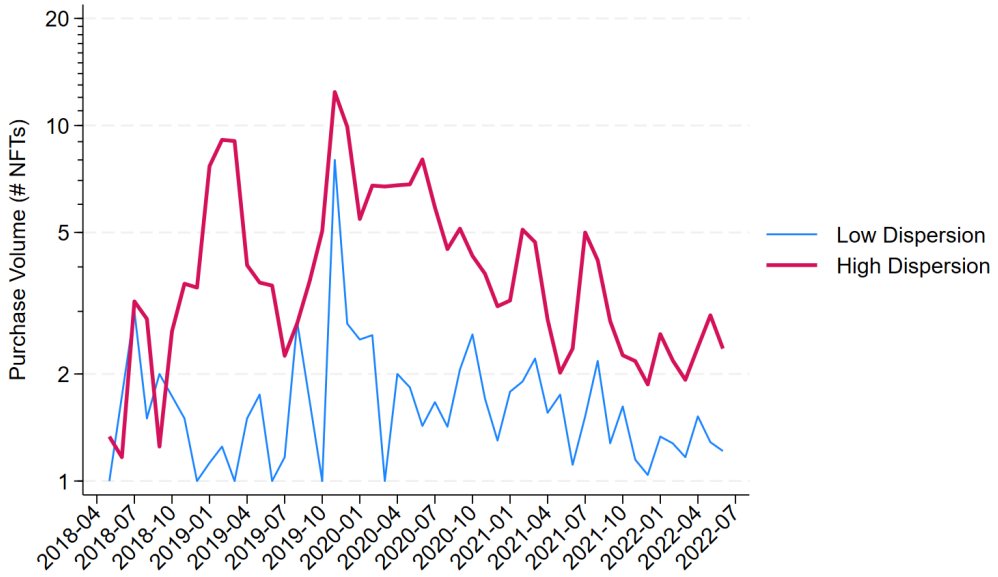
This figure shows the MCMC selection-adjusted indexes estimated with different selection equation specifications. The thick blue line is estimated by the baseline model with quadratic terms of holding period returns and holding period lengths as in Table 2 column 1. The red dot-dashed line and the green dashed line add Ethereum returns (ETH Ret) and Google search volume index (SVI) as an additional covariate, respectively, corresponding to Table 2 columns 2 and 3. The thin yellow line is estimated with the selection equation that includes both covariates as in Table 2 column 4.



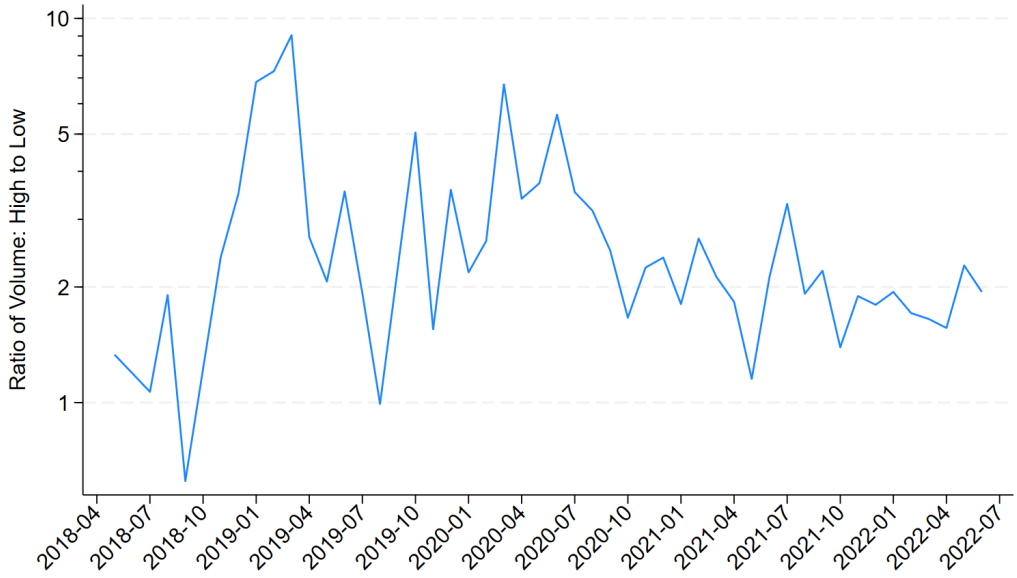
Appendix Figure A5. Purchase volume from high- and low-bid dispersion investors over time: fixed cut-off.

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 the low-dispersion group as investors with $BidDispr = 0$, i.e., only one bid in the previous three months, and the high-dispersion group $BidDispr > 0$ and plot the monthly time-series of purchase volume per investor by the two groups.

(a) Purchase volume

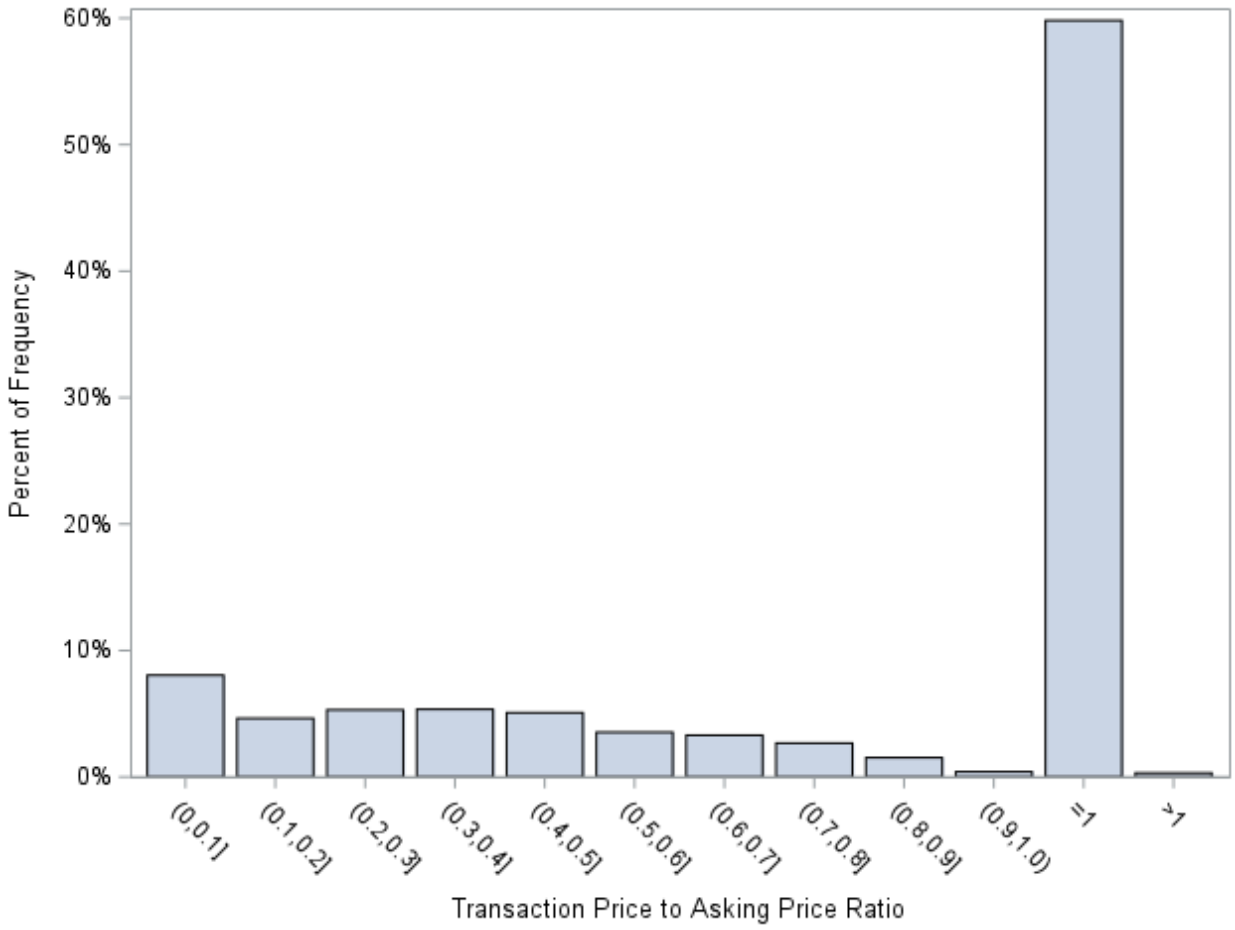


(b) Ratio of volume



Appendix Figure A6. Distribution of the ratios of transaction prices to asking prices.

This figure shows the distribution of the ratios of transaction prices to the most recent asking (listing) prices before the transaction among all successful transactions. The ratios are grouped at 0.1 intervals below 1, with a mass point at 1, and are grouped into one group above 1.



Appendix Table A1. Interactions between extrapolation and selection bias: Individual-level regressions with "up-to" indexes.

This table tests the interactions between extrapolation and selection bias using transaction data at the investor-month levels. 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 "up-to" MCMC selection-adjusted indexes estimated with expanding windows. $SelectionBias$ is defined as the difference between the "up-to" RSR index return and the MCMC index return, both in the past six-month horizon, and is normalized. All specifications include past 6-month Ethereum returns (ETH Return) as covariates. Cluster robust standard errors clustered by investors are reported in parentheses. Significance level: *: p<0.1, **: p<0.05, ***: p<0.01.

Dependent variable:	(1)	(2)	(3)	(4)	(5)	(6)
	$Log(Buy_{i,t}+1)$			$1(Buy_{i,t}>0)$		
$LogUnbiasedReturn_{t-1}$	0.096***			0.352***		
	(0.009)			(0.029)		
$LogUnbiasedReturn^{+}_{t-1}$		0.103***	0.103***		0.341***	0.330***
		(0.011)	(0.013)		(0.033)	(0.038)
$LogUnbiasedReturn^{-}_{t-1}$		0.081***	0.166***		0.381***	0.709***
		(0.013)	(0.022)		(0.065)	(0.097)
$SelectionBias_{t-1}$			-0.050***			-0.197***
			(0.016)			(0.076)
$LogUnbiasedReturn^{+}_{t-1} \times SelectionBias_{t-1}$			0.031***			0.096**
			(0.010)			(0.039)
$LogUnbiasedReturn^{-}_{t-1} \times SelectionBias_{t-1}$			-0.072***			-0.292***
			(0.012)			(0.058)
$ETH\ Return_{t-1}$	-0.102***	-0.095***	-0.124***	-0.278***	-0.291***	-0.399***
	(0.015)	(0.015)	(0.017)	(0.046)	(0.050)	(0.061)
Investor fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12,719	12,719	12,655	12,608	12,608	12,531

Appendix Table A2. Interactions between extrapolation and selection bias at the investor level: heterogeneity over investor ability, using "up-to" indexes.

This table examines the interactions between extrapolation and selection bias using investor-level data 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) in the past. 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 "up-to" MCMC selection-adjusted indexes estimated with expanding windows. $SelectionBias$ is the difference between the "up-to" RSR index return and the MCMC index return, both in the six-month horizon, and is normalized. All specifications include past 6-month Ethereum returns (ETH Return) as covariates. 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:	$Log(Buy_{i,t}+1)$	$1(Buy_{i,t}>0)$	$Log(Buy_{i,t}+1)$	$1(Buy_{i,t}>0)$
$LogUnbiasedReturn^{+}_{t-1}$	0.075***	0.316***	0.132***	0.562***
	(0.011)	(0.042)	(0.016)	(0.067)
$LogUnbiasedReturn^{-}_{t-1}$	0.174***	0.798***	0.130***	1.080***
	(0.020)	(0.098)	(0.018)	(0.135)
$SelectionBias_{t-1}$	-0.066***	-0.195**	-0.049***	-0.300***
	(0.016)	(0.085)	(0.014)	(0.100)
$LogUnbiasedReturn^{+}_{t-1} \times SelectionBias_{t-1}$	0.044***	0.094*	0.037***	0.214***
	(0.011)	(0.048)	(0.014)	(0.063)
$LogUnbiasedReturn^{-}_{t-1} \times SelectionBias_{t-1}$	-0.082***	-0.330***	-0.058***	-0.409***
	(0.012)	(0.065)	(0.009)	(0.073)
$Ability_{i,t-1}$	0.092***	0.307***	0.142	-0.088
	(0.018)	(0.055)	(0.097)	(0.285)
$LogUnbiasedReturn^{+}_{t-1} \times Ability_{i,t-1}$	0.010	-0.043**	-0.024	-0.149**
	(0.009)	(0.022)	(0.015)	(0.063)
$LogUnbiasedReturn^{-}_{t-1} \times Ability_{i,t-1}$	-0.036**	-0.211***	-0.059***	-0.433***
	(0.015)	(0.058)	(0.017)	(0.132)
$SelectionBias_{t-1} \times Ability_{i,t-1}$	0.034**	0.047	0.028**	0.116
	(0.016)	(0.066)	(0.014)	(0.099)
$LogUnbiasedReturn^{+}_{t-1} \times SelectionBias_{t-1} \times Ability_{i,t-1}$	-0.006	0.000	-0.032**	-0.161**
	(0.008)	(0.028)	(0.014)	(0.068)
$LogUnbiasedReturn^{-}_{t-1} \times SelectionBias_{t-1} \times Ability_{i,t-1}$	0.025**	0.093**	0.024**	0.091
	(0.010)	(0.039)	(0.009)	(0.071)
$ETH\ Return_{t-1}$	-0.123***	-0.454***	-0.081***	-0.360***
	(0.017)	(0.062)	(0.016)	(0.071)
Investor fixed effects	Yes	Yes	Yes	Yes
Observations	12,655	12,531	10,897	10,677

Appendix B. Details of the MCMC algorithm

In this appendix, we provide more details about our application of the Markov Chain Monte Carlo (MCMC) algorithm proposed by Korteweg et al. (2016).

1. A Sketch of the MCMC algorithm.

Here we outline the MCMC algorithm to estimate the three-equation system specified in the main text. For additional technical information, the reader can refer to the Online Appendix of Korteweg et al. (2016).

For each NFT, $i = 1,2, \dots, N$, the natural logarithm of price, $p_i(t)$, at time $t = 1,2, \dots, T$, follows the process

$$p_i(t) = p_i(t-1) + \delta(t) + \varepsilon_i(t), \tag{B.1}$$

with the idiosyncratic risk $\varepsilon_i(t) \sim N(0, \sigma^2)$ and i.i.d. across NFTs and across months. The price is observed if and only if $w_i(t) \geq 0$, where the latent variable $w_i(t)$ is given by the selection equation,

$$w_i(t) = \sum_{k=1}^{K} g_k\big(p_i(t); Z_i(t)\big) \cdot \alpha_k + W_i(t)'\alpha_W + \eta_i(t). \tag{B.2}$$

In this equation, the vector $Z_i(t)$ contains sets of covariates from the observed data, especially the price of the prior sale $p_i(s)$ at time s. g_k is a set of nonlinear functions of the nonannualized return of the NFT since the prior sale, $r_i^p(t) = p_i(t) - p_i(s)$. The vector $W_i(t)$ can include a number of hedonic characteristics as well as macroeconomic variables other than $p_i(t)$ for identification. Here we let $W_i(t) = \left[h_i(t), \big(h_i(t)\big)^2\right]'$, where $h_i(t) = t - s$ is the number of months since the prior sale. The error term $\eta_i(t)$ is distributed i.i.d. $N(0,1)$ by normalization and is uncorrelated with $\varepsilon_i(t)$.

The parameters of interest are $\theta = [\delta(2), \dots, \delta(T), \alpha, \sigma^2]$, but we augment it with the latent prices, $\{p_i(t)\}$, and the selection variables, $\{w_i(t)\}$, for the simulation procedure. Specifically, we want to estimate the posterior distribution of the parameter set given observed data and the prior of parameters, $f(\theta, \{p_i(t)\}, \{w_i(t)\}|data)$, using a Gibbs sampler. The sampler consists of three blocks, each of which iteratively draws realizations of a subset of parameters based on their

complete conditional distributions given the remaining parameters drawn in the previous iterations. The three blocks are as follows, where superscript "−" indicates the realizations of the parameters are from the previous iteration:

Step 1. Draw latent prices by $f(\{p_i(t)\}|\{w_i^-(t)\}, \theta^-, data)$, which can be transformed to

$$f(p_i(t)|p_i(t-1), p_i(t+1), \{w_i(t)\}) \propto f\big(w_i(t)\big|p_i(t)\big) \cdot f\big(p_i(t)\big|p_i(t-1), p_i(t+1)\big), \tag{B.3}$$

The first component is specified by equation (B.1) and the second component by equation (B.2). Given the system of equations, we can use the single-state updating Metropolis-Hastings sampler to draw latent prices between two consecutive sales.

Step 2. Draw selection variables by $f(\{w_i(t)\}|\{p_i(t)\}, \theta^-, data)$. This is specified by the selection equation (B.2).

Step 3. Draw parameters by $f(\theta|\{p_i(t)\}, \{w_i(t)\}, data)$. Given simulated monthly prices and, hence, simulated monthly returns, index returns δ can be estimated by the (single-period) repeat-sale regression using the Bayesian regression approach. For the selection equation, a data-augmented probit model specified by equations (B.2) and (B.3) can be estimated with the standard Bayesian approach.

By the Clifford-Hammersley theorem, Gibbs sampler generates a Markov chain converging to the target augmented distribution of $f(\theta, \{p_i(t)\}, \{w_i(t)\}|data)$. For inference purposes, the marginal distribution $f(\theta|data)$ can then be calculated by integrating the joint distribution over $p_i(t)$ and $w_i(t)$. We conduct a total of 20,000 iterations in our estimations: the first half is dropped as the "burn-in" period to allow the sampler to converge and use the second half for inferences.

2. Specification of Initial Values.

To start the Markov chains, we need to specify the initial values of the parameters. For latent prices $\{p_i(t)\}$, since they are simulated in the very first step, no initial values are required. However, using some simple interpolation methods to initialize the latent prices can speed up the algorithm, such as filling in the unobserved prices with previous transaction prices or linearly interpolated prices. For latent selection variables $\{w_i(t)\}$, they can be simulated given simulated

latent prices and parameters of the selection equation, α's. For the first iteration, we initialize all α's to be zero.

3. Specifications of the Bayesian Regressions in Step 3.

Based on the law of motion equation (B.1), parameters $\delta = [\delta(2), \dots, \delta(T)]'$ and σ^2 can be obtained by regressing one-period returns on a set of dummy variables indicating the time of the corresponding returns given simulated latent prices in Step 1. Specifically, the regression equation is

$$p_i(t) - p_i(t-1) = \sum_{\tau=2}^{T} I(t=\tau) \cdot \delta(\tau) + \varepsilon_i(t), \tag{B.4}$$

for all i and all $t = 2, \dots, T$. Hence, in matrix form, the dependent variable Y_p stacks all the one-period log returns, $p_i(t) - p_i(t-1)$, and the corresponding row of the independent variable X_p is all zeros except for a one in column $t-1$. Let N be the size of Y_p.

The standard conjugate normal-inverse gamma prior for δ and σ^2 are,

$$\sigma^2 \sim IG(a_0, b_0), \qquad \delta|\sigma^2 \sim N(\mu_0, \sigma^2 \Sigma_0^{-1}).$$

The posterior distributions are

$$\sigma^2 \big| X_p, Y_p \sim IG(a, b), \;\; \delta \big| \sigma^2, X_p, Y_p \sim N(\mu, \sigma^2 \Sigma^{-1}),$$

where

$$a = a_0 + N, \quad b = b_0 + e'e + (\mu - \mu_0)' \Sigma_0 (\mu - \mu_0), \quad e = Y_p - X_p \mu,$$

$$\Sigma = \Sigma_0 + X_p' X_p, \quad \mu = \Sigma^{-1} \big(\Sigma_0 \mu_0 + X_p' Y p \big).$$

We use prior $a_0 = 2.1$ and $b_0 = 1$, corresponding to a prior $\sigma^2 \approx 0.91$ per month and a variance of 8.26, which is almost an uninformative prior. The prior of δ are $\mu_0 = 0$, a null prior, and Σ_0 the identity matrix, indicating that the prior standard deviation of δ is the same as the idiosyncratic risks. In addition, to tune the algorithm in the "up-to" MCMC estimations, we use a relatively stronger prior for δ by setting Σ_0^{-1} to be the identity matrix multiplied by 0.2 to account for the fact that we have much fewer transactions to obtain robust convergence if the weaker prior is still used in the early periods. The tighter prior implies that the prior distribution of δ, the monthly log index return, is within about ±4 with 99% probability – still quite diffused a prior in our setting.

We also confirm that the choice of Σ_0 has negligible impact on the posterior of δ as our sample period for estimation extends to about 24 months when sufficient transactions are included.

The parameters $\alpha = [\alpha_0, \alpha_1, \dots, \alpha_K, \alpha_w]'$ are estimated by regressing the selection variables $w_i(t)$ on the covariates $\left[1, g_1\left(p_i(t); Z_i(t)\right), \dots\ g_k\left(p_i(t); Z_i(t)\right), W_i(t)'\right]$ with known variance equal to 1, which are denoted by Y_w and X_w in matrix form. The prior distribution of α is

$$\alpha \sim N(\theta_0, \Omega_0^{-1}),$$

and the posterior is

$$\alpha | Y_w, X_w \sim N(\theta, \Omega^{-1}),$$

where

$$\Omega = \Omega_0 + X_w' X_w, \quad \theta = \Omega^{-1}(\Omega_0 \theta_0 + X_w' Y_w).$$

We use a diffuse prior with $\theta_0 = 0$ and $\Omega_0 = I_K \cdot 1/10{,}000$.

4. Diagnostics of the Convergence.

We implement several tests for the convergence of our MCMC algorithm following Korteweg and Sorensen (2010) and Gelman et al. (2013). We first present the trace plots of parameters during the whole simulation process in Appendix Figure B1. The shaded areas indicate the second half of the draws for inference. The red horizontal lines mark the parameter means calculated with the second half of draws. As shown in the figure, our Markov chains actually converge quite rapidly in the first few thousand draws. The paths to convergence are not even noticeable for some parameters. Dropping the first 10,000 draws as the "burn-in" period is conservative.

Another test of the performance of our MCMC algorithm is checking the autocorrelation functions of simulated parameters. The autocorrelation function of draws should decline to zero quickly for a good sampler so that there are sufficient uncorrelated draws for inferences. Appendix Figure B2 reports the autocorrelation functions up to the lag of 100 iterations. The figure suggests that the autocorrelation functions drop to a low level at a reasonable rate.

We next implement two formal tests of convergence proposed by Geweke (1992) and Gelman and Rubin (1992). The Geweke (1992) approach is a test on a single Markov chain. It compares the means of simulated parameters in the first 10% and the last 50% of draws after the

burn-in period. It is similar to a standard two-sample *t*-test except that the standard deviations of the samples are adjusted for autocorrelation. If the chain has converged, the means of the two parts should display no significant difference. Appendix Table B1 reports the test results. Mean 1 and S.D. 1 are the mean and autocorrelation-adjusted standard deviation of the first 10% draws after the burn-in period and Mean 2 and S.D. 2 are for the last 50%. The Z-scores of the differences of means are calculated and the corresponding *p*-values are reported. All Z-scores are small and all *p*-values are greater than 0.1, suggesting that the first and last portions of the chain have similar means – an indication of convergence.

Finally, we conduct maybe the most popular convergence diagnostic using multiple simulated chains first proposed by Gelman and Rubin (1992). The test examines how well several simulated Markov chains starting from different initial values converge to the same posterior distribution. The intuition is that, if the algorithm converges to a stable distribution, the impacts of initial values should fade away as the simulation continues. Therefore, chains with different initial values that are sufficiently dispersed should look similar ("mix well") after a sufficiently long burn-in period.

We run the MCMC algorithm for three chains. Following Gelman et al. (2013), we generate three sets of well-dispersed initial values from a rough estimation of the model using an extended Kalman filter, which linearizes the system and can be estimated with the standard Forward Filtering Backwards Sampling procedure. We also set different random seeds for the simulations. Each chain has 20,000 draws, of which the first half is dropped as the "burn-in" period, so only the second half of 10,000 draws is used for diagnostic. Following Gelman et al. (2013), we further split each chain of 10,000 draws by half so that we are able to detect any "within-chain" performance of mixing. Therefore, we have a total of six chains, each of which consists of 5,000 draws, for the multi-chain test of convergence.

The last two columns of Appendix Table B1 reports the results from the Gelman-Rubin test. The "N_{eff}" reports the effective chain length after accounting for autocorrelation. It is a formal statistics for the autocorrelation test in Appendix Figure B2. According to Gelman et al. (2013), a few hundreds of effective draws are sufficient for inferences, as is the case in our simulations. The last column reports the "R-hat" statistics, which measures the degree of chain mixing as the ratio of the "between-chain" variance to average "within-chain" variances. A smaller "R-hat" implies

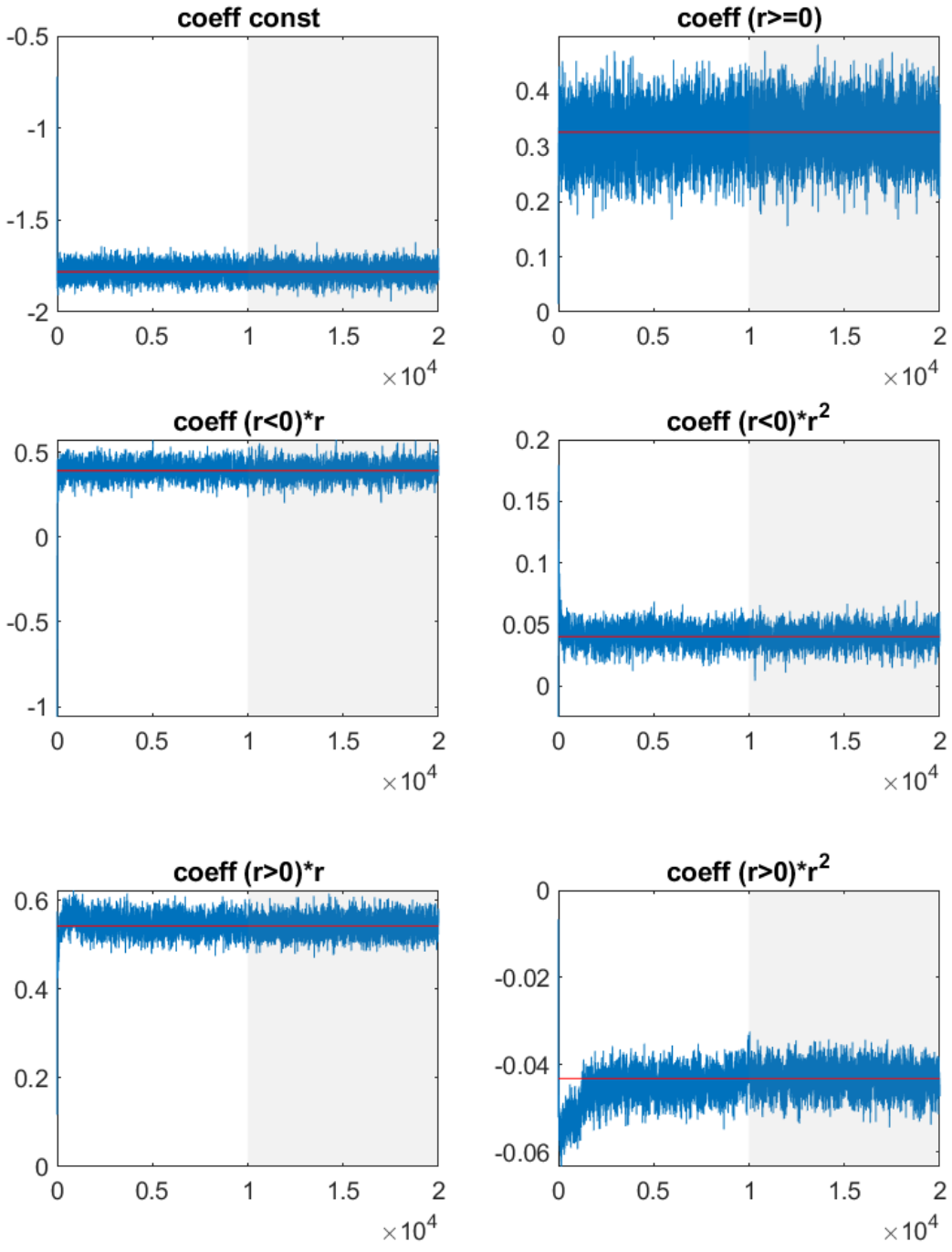
better mixing. Usually, an "R-hat" below 1.1 is recommended. Here, we follow Vehtari et al (2021) and report the maximum of the rank normalized split-R-hat and the folded-split-R-hat. As shown in the table, all R-hat are well below 1.1. We therefore conclude that our algorithm performs well and converge to a stable posterior distribution, which gives reliable inferences.

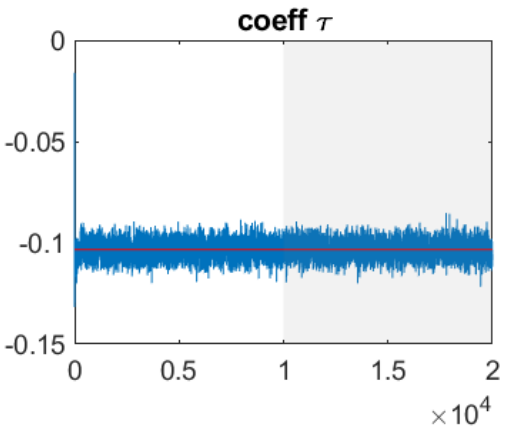
Additional Reference:

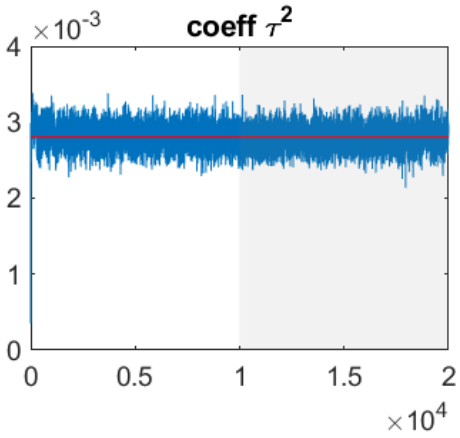
Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P. C. (2020). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC. arXiv:1903.08008.

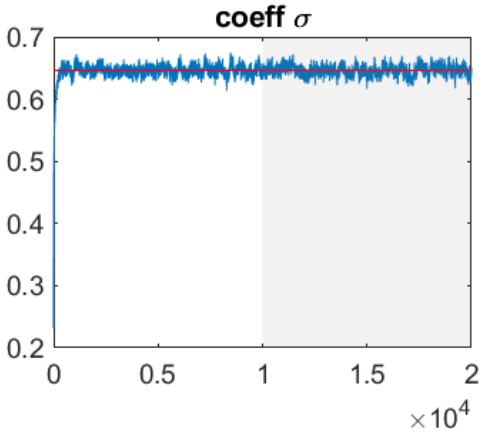
Appendix Figure B1. Trace plots of coefficients.

This figure presents the trace plots of the parameters. The shaded areas indicate the second half of the draws for inference. The horizontal lines mark the parameter means calculated with the second half of draws.



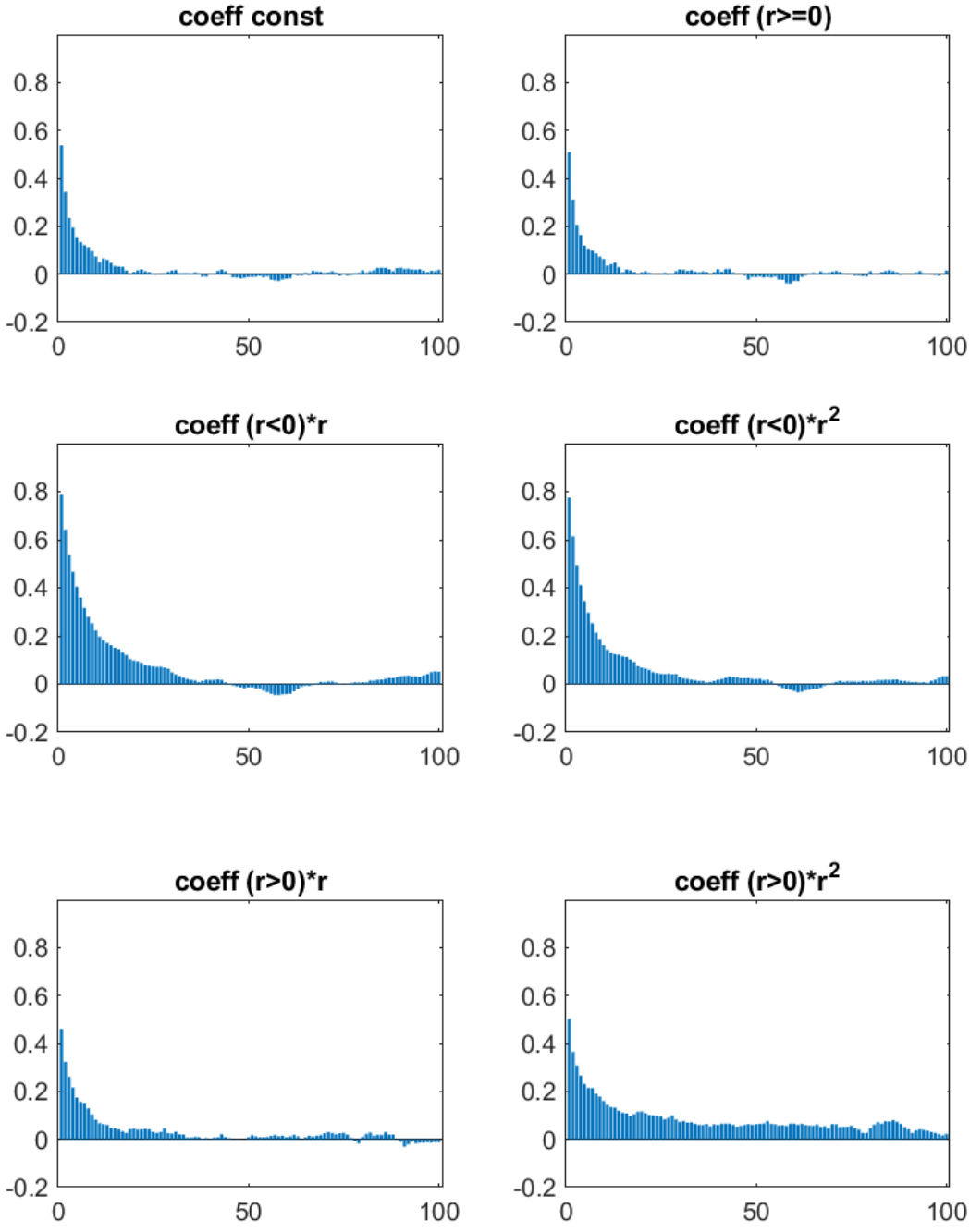

coeff τ
0
-0.05
-0.1
-0.15
0
0.5
1
1.5
2
$\times 10^4$

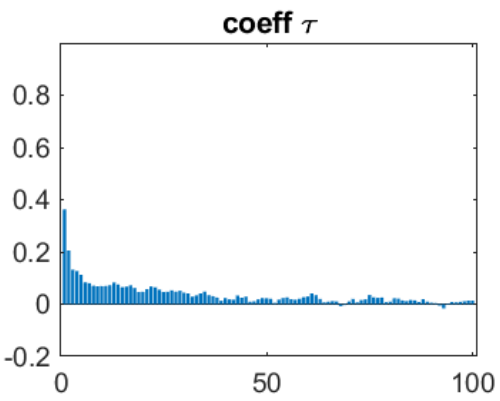

coeff τ^2
$\times 10^{-3}$
4
3
2
1
0
0
0.5
1
1.5
2
$\times 10^4$

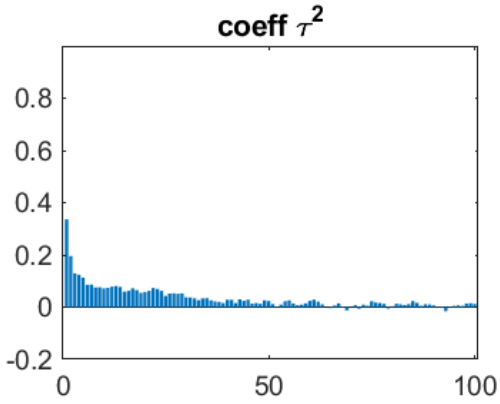

coeff σ
0.7
0.6
0.5
0.4
0.3
0.2
0
0.5
1
1.5
2
$\times 10^4$

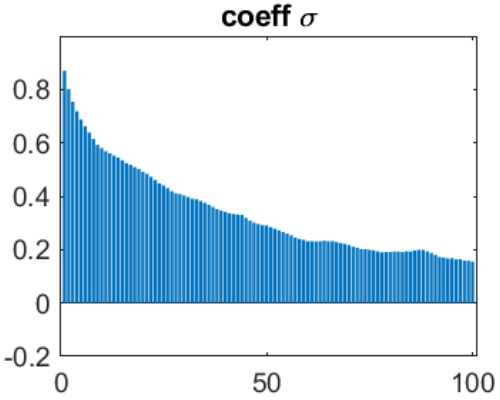
Appendix Figure B2. Autocorrelation of Draws.

This figure shows the autocorrelation functions of parameter draws with the lags of 1 to 100.




coeff τ
0.8
0.6
0.4
0.2
0
-0.2
0
50
100


coeff τ^2
0.8
0.6
0.4
0.2
0
-0.2
0
50
100


coeff σ
0.8
0.6
0.4
0.2
0
-0.2
0
50
100

Appendix Table B1. Tests of convergence by Geweke (1992) and Gelman and Rubin (1992).

This table reports the test results of convergence diagnostic by Geweke (1992) and Gelman and Rubin (1992). The Geweke's diagnostic compares the means of simulated parameters in the first 10% (Mean 1) and the last 50% (Mean 2) of draws after the burn-in period. A Z-score of the difference is calculated, and the corresponding *p*-value is reported. The Gelman-Rubin diagnostic examines how well multiple chains estimated with different initial values converge to the same distribution. The effective sample size (N_{eff}) and the "R-hat" statistics are reported.

Parameter	Mean 1	S.D. 1	Mean 2	S.D. 2	Z-score	*p*-value	N_{eff}	R-hat
Constant	-1.777	0.349	-1.780	0.292	-0.86	0.390	242.6	1.0147
1(r>0)	0.322	0.333	0.325	0.345	0.58	0.561	200.1	1.0180
r*1(r<0)	0.399	0.717	0.393	0.646	-0.66	0.511	490.1	1.0072
r^2*1(r<0)	0.040	0.106	0.041	0.074	0.56	0.573	1576.4	1.0022
r*1(r>0)	0.542	0.136	0.544	0.157	1.40	0.162	1096.0	1.0032
r^2*1(r>0)	-0.043	0.035	-0.043	0.031	-0.87	0.382	191.7	1.0188
t	-0.103	0.021	-0.103	0.046	-0.89	0.374	7517.6	1.0005
t^2	0.003	0.001	0.003	0.001	1.55	0.122	2145.3	1.0016
sigma	0.647	0.085	0.646	0.223	-0.64	0.523	904.4	1.0081