THE UNIVERSITY OF SYDNEY BUSINESS SCHOOL

DISCIPLINE OF FINANCE

Exploring the Liquidity of NFT Collections

(Honours Thesis)

Author

Jackson Rogers

[redacted]

Supervisors

Dr Jiri Svec & Dr Angelo Aspris

A thesis submitted in partial fulfilment of the requirements for the degree of Bachelor of Commerce & Bachelor of Advanced Studies (Honours)

November 2023

STATEMENT OF ORIGINALITY

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at the University of Sydney or at any other educational institution, except where due acknowledgement is made in the thesis.

Any contribution made to the research by others, with whom I have worked at the University of Sydney or elsewhere, is explicitly acknowledged in this thesis.

I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the projects design and conception or in style, presentation and linguistic expression is acknowledged.

Signature of Author

Jackson Rogers

November 2023

ACKNOWLEDGEMENTS

First and foremost, I extend my deepest thanks to my dedicated supervisors, Dr. Jiri Svec and Dr. Angelo Aspris. Your guidance, expertise, and unwavering commitment were instrumental in shaping the direction of my research. Throughout this process, you provided me with invaluable insights, constant encouragement, and constructive feedback at every stage. Your mentorship has been truly helpful, and I am immensely grateful for your support.

I would also like to extend my appreciation to Dr. Sean Foley and his team at Macquarie University for their generous assistance with data acquisition. His willingness to share resources and knowledge greatly facilitated the data collection process, enabling me to conduct a thorough analysis for my thesis.

Lastly, I want to express my gratitude to my honour's coursework teachers and honours program director, Dr. Reuben Segara. Their dedication to delivering quality honours program and fostering a supportive learning environment has been instrumental in equipping me with the skills and knowledge required to undertake this research. Their guidance during the coursework laid the foundation for this thesis and has been invaluable.

Table of Contents

List of Tables

List of Figures

List of Appendices

1. Abstract

This paper provides the first analysis of non-fungible token (NFT) collection liquidity by applying a suite of widely used proxies that capture different dimensions of liquidity. Using transaction-level data from the OpenSea marketplace, manipulative trades are flagged and two novel methodologies for calculating liquidity are applied before performing a family of regressions to investigate its dynamics. I find that collection-specific attributes directly account for both NFT-specific liquidity idiosyncrasies and the impacts of manipulative trading. Following robustness tests, I identify that this collection-level power only exists in bull markets, similarly to real estate ZIP-code groupings. Finally, the estimated models reveal a non-linear liquidity pattern across a collection's lifetime, with successful collections dipping in liquidity before recovering quickly. This paper deepens our understanding of how liquidity operates at the collection level in NFTs, offering findings for liquidity researchers in non-fungible asset markets.

2. Introduction

Providing liquidity is one of the essential roles of financial markets. It is the market's duty to facilitate efficient exchanges by creating an environment that allows participants to trade quickly and at a low cost (O'Hara, 2003). Liquidity is frequently used as a key metric to evaluate market health by regulators and exchanges alike, with each asset class displaying unique liquidity dynamics (Le & Gregoriou, 2020; Schestag et al., 2016). Surprisingly, there has been minimal research conducted into the dynamics, nuances, and efficacy of liquidity measures within the rapidly growing market of non-fungible tokens (NFTs). This novel asset class gained global recognition in early 2020 due to its innovative technology, widespread media exposure, and annual returns magnitudes larger than any traditional asset class (Borri et al., 2022). In 2021 alone, USD 25 billion of trading volume was circulated, with notable sales like Merge, a fractionalised piece of art, being valued at $91.8 million after initial minting (Muroki, 2023).

This paper is the first study into the dynamics of NFT liquidity from a collection perspective, utilising publicly available transaction data and four widely used liquidity proxies to explore the fundamental dynamics and determinants of liquidity. All data is sourced directly from the decentralised OpenSea marketplace via their free API key and covers 7 months of trading between 1 Feb 2021 and 21 Aug 2021. Having acquired the unstructured data, numerous filters and data-cleaning processes are applied to create a sales dataset that tracks asset identifiers, collection names, buyer addresses, seller addresses, prices, times of sales and transaction IDs. Next, a duplicate sales dataset is created and flags wash-trades (manipulative trading) for removal. For both resulting datasets, the four widely used liquidity proxies are estimated following two novel methodologies: one by assuming intra-collection homogeneity, in which all NFTs within a collection are assumed identical, and the other by assuming intra-collection heterogeneity, in which individual NFT liquidity proxies are calculated and averaged by collection-day. This process yields four unique datasets with over 30,000 collection-day observations that allow for analysis of the impact of wash trading and the effect of incorporating NFT-specific liquidity information. For each dataset combination (pre/post-wash-trade exclusion, homogeneous/heterogeneous collection assumption), each liquidity proxy is regressed against various price-derived variables, with collection and date fixed effects being added and standard errors being clustered around collections. For the purposes of exploring the explanatory power of collection fixed effects, the dataset is divided into a bear and bull period and directly compared using the same group of regressions. As a further robustness test, all regressions are additionally performed using the natural log of the liquidity proxies. The results reveal multiple findings that are significant for both NFT and traditional asset markets. Firstly, the heterogeneous collection assumption provides significant NFT-specific information, making estimates more accurate at the cost of clustering low-observation count NFTs at the edges of its range. Secondly, this idiosyncratic information is captured completely by the addition of collection-specific fixed effects, irrespective of which collection assumption methodology is employed. Thirdly, the impacts of wash trading on liquidity are additionally captured by collection-specific fixed effects, which supports the findings of existing literature. Fourthly, the explanatory power of collection fixed effects seems only to be present during boom markets, drawing a significant parallel between NFT and real estate liquidity dynamics. Finally, there is a non-linear liquidity dynamic as collections age, with most unsuccessful collections showing a sharp rise in illiquidity until they fail (captured by pure transaction-cost and price-impact estimators), while a minority of long-living collections undergo a subsequent liquidity revival (captured by the long-term implicit bid-ask spread estimator).

Due to the relative youth of NFTs compared to existing asset markets, there is a scarcity of literature on the topic, and in particular regarding liquidity. At the time of submission, Wilkoff & Yildiz (2023) have published the only paper on the topic and applies a highly focussed methodology, analysing the dynamics of individual NFTs (using Amihud's illiquidity ratio) throughout the calendar week and in relation to the arrival of NFT-related news. Their niche scope provides a strong benchmark and an opportunity to explore additional liquidity proxies using two new methods of calculating collection liquidity.

As such, this paper provides a fundamental exploration regarding how to study NFT collection liquidity, what information is captured at the collection level and what drives collection liquidity. This research is particularly pressing due to the rapid growth of the market, unique features and a plethora of potential use cases. As noted by White et al. (2022b), the NFT market has experienced extreme growth with average annual returns of 70% between 2017 and 2021, although these returns are extremely volatile. The market has multiple unique attributes that attract a wide range of investors: round-the-clock trading, complete decentralisation, near-zero barriers to entry and a completely unregulated environment with total informational transparency. These features offer the NFT market as a unique market setting to which researchers can compare traditional asset markets such as equities, bonds, OTC contracts, physical art, and real estate.

Further growth of this asset class is accelerated by the development and implementation of the technology's potential uses beyond financial applications. Recently, companies such as JP Morgan and Walmart have launched and are continually developing specialised applications for NFTs within their operations, including the Onyx blockchain to enable secure intra-bank cashflow automation and supply chain tokenisation to track and identify anomalies (Blockdata, 2022). This has been noted in the literature, with various studies being optimistic about the long-term impact of NFTs in commerce, often citing widespread potential in the operation of sports, law, escrow, ticketing, digital collectibles, gaming and real estate (Chohan & Paschen, 2023; Wilson et al., 2022). Echoing these sentiments, Chalmers et al.,

(2022) find significant long-term NFT use cases for creative entrepreneurs, while urging caution in the short-term due to the dominance of speculative behaviours and boom-bust market cycles; these sentiments are mirrored by Van Haaften-Schick & Whitaker (2022), who posit that NFT's have revolutionised the artist-contract via low-cost high-complexity contracts, self-managing royalties and a novel fundraising system

The applications of NFTs reach far beyond the discipline of finance: in the literature, 93.1% of NFT literature resides in the non-finance domain and the majority of research focusses on computer science and bespoke applications to niche sectors. For example, Nobanee & Ellili (2023), note that the first journal paper published on NFTs studied the possibility of using NFTs to aid wildlife conservation (Mofokeng & Matima, 2018). Within marketing, multiple papers have been published on their potential applications for digital media (Chohan & Paschen, 2023), brand management (Colicev, 2023) and promotional purposes (Taylor, 2023). Within the information technology space, there has been heavy focus on computer architecture (Hasan & Starly, 2020), Ethereum (Arcenegui et al., 2021; Dos Santos et al., 2021), the Internet of Things (Arcenegui et al., 2021; Lee & Kwon, 2021; Manzoor et al., 2020), the metaverse (Far et al., 2022; Nadini et al., 2021) and automation (Hamledari & Fischer, 2021). This rapid growth in research and commercial implementation has resulted in one of the fastest-growing markets available, making a fundamental understanding of what liquidity looks like increasingly pertinent.

This paper offers multiple contributions to academia. From the literature review, future researchers can gain a comprehensive understanding of how NFT market liquidity differs and compares to traditional assets; there are currently multiple NFT papers dedicated to bridging the academic gap between traditional finance and NFTs, but none regarding liquidity. From this paper's methodology, academics can follow the same open-source data acquisition process and apply the same liquidity proxy equations to obtain the final datasets, which is unique amongst the majority of NFT literature that relies on third-party data services (e.g. Dune (Wilkoff & Yildiz, 2023), nonfungible.com (Anselmi & Petrella, 2023; Urom et al., 2022; Xia et al., 2022; Zhang et al., 2022) and coinmarketcap.com (Dowling, 2022; Wang et al., 2022; Yousaf & Yarovaya, 2022b); by pulling data directly from the decentralised marketplace OpenSea, the data is reputable, secure and replicable. From the results, researchers can materially benefit in two ways: firstly, by saving time by comparing their own results to the descriptive statistics, correlation coefficients and regression outputs, and secondly, by gaining insights into how liquidity is reflected at the collection level in this novel asset class.

The results show that Roll's liquidity measure has noteworthy correlations with all the other liquidity proxies, revealing a shared exposure to the same liquidity dimension. Basic regression results largely align with prior research in terms of the magnitude of relationships, especially regarding the Amihud/Roll relationship. Moreover, the inclusion of collection fixed effects effectively captures two key components of NFT-specific information: firstly, they eliminate the difference between the intra-collection homogeneous and heterogeneous assumption methodologies, capturing NFT-specific liquidity factors, and secondly, they eliminate the impacts of wash trading, capturing NFT targeted wash trading. This finding is of particular significance, as it allows for a more accurate assessment of NFT collection liquidity, and suggests that it is a critical component for future analysis of liquidity in unique assets; this has been lightly explored in real-estate with respect to ZIP-code groupings, but has potentially exists using other shared characteristics. These results underscore the importance of considering individual asset attributes and comprehensively accounting for wash trading effects, which can now be done via collection-level fixed effects. The resulting estimates reveal the widespread significance of the daily collection transaction volume, age, age squared and the Herfindahl–Hirschman index variables, with the transaction cost estimators (Lesmond, Goyenko and Roll) additionally estimating daily collection volatility to be significant. The estimated age coefficients reveal a non-linear relationship between collection age and liquidity, with the transaction cost proxies predicting that collection liquidity falls in its early stages and then begins to rise after it matures. The differences between the variable coefficient estimates for the four liquidity proxies are representative of each proxy capturing a different dimension of liquidity, with Amihud targeting the short-term price impact of transaction volume (and by extension short-term bid-ask spread), Lesmond/Goyenko estimating explicit transaction cost related liquidity and Roll approximating long-term bid-ask spreads or implicit transaction cost related liquidity.

Our findings have significant implications for future research and analysis both in the realm of NFTs and other asset classes. First, the validation of the homogeneous collection assumption suggests that future NFT studies, marketplaces, and investors can efficiently leverage collection-specific characteristics to explain differences in individual asset liquidity, potentially extending to dimensions beyond liquidity, such as returns, volatility, and inter-asset relationships. This finding also draws a new parallel with the real estate market regarding the importance of collection/ZIP-code fixed effects during boom markets, offering a new avenue for additional comparative research. Secondly, the insights regarding the impact of wash trading offer practical advantages for both future researchers and regulators, as the inclusion of collection fixed effects effectively eliminates the influence of wash trading, simplifying data sourcing and reducing computational and data processing burdens. These findings have wider implications beyond the NFT market, such as in equities, bonds, real estate, and art; researchers are able to analyse the interplay between market manipulation and liquidity without undue concern over wash trading effects. Lastly, my examination of transaction cost estimators indicates that cost-induced illiquidity, namely the implicit portion, rises and then falls rapidly as a collection ages, providing useful information for market facilitators who are considering the potential costs of facilitating unique assets that rely on age as a key value driver. This contrasts with the price-impact proxy predicting extreme liquidity at inception before quickly reverting, which is a result of its short-term focus on capturing failed collection liquidity. Overall, my results hold practical implications for researchers, regulators, and market facilitators seeking to better understand liquidity in this volatile and young asset class by offering valuable insights into the application of liquidity proxies and the importance of comprehending mechanisms like those of Amihud and Roll's estimated models in non-fungible asset markets for effective liquidity analysis and decision-making.

The remainder of the paper is structured as follows: section 3 outlines the necessary institutional knowledge of the NFT market, section 4 conducts a literature review, section 5 details the data acquisition and manipulation process, section 6 details the methodology for removing wash-trades, calculating collection liquidity proxies and regression specifications, section 7 explores the results and section 8 concludes the paper.

3. Institutional Knowledge

3.1. Liquidity and fungibility

This paper defines asset illiquidity as the discount required to immediately sell an asset rather than waiting for a buyer willing to pay the asset's fair value. For example, selling shares of a highly liquid stock immediately versus next month requires a relatively small discount, whereas selling a (highly illiquid) house immediately versus next month would require a significant discount from its fair price.

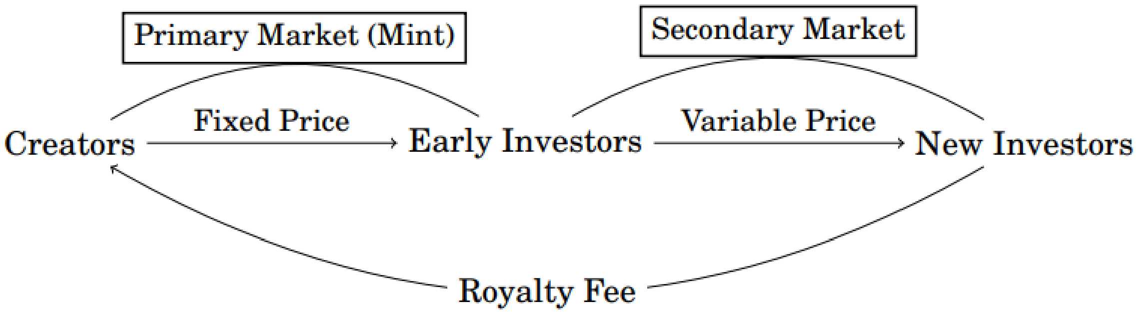
Fungibility is the uniqueness or interchangeability of an asset. For example, common shares, money and gold are the exact same and can be interchanged with each other freely. Non-fungible assets, however, are unique: houses, artwork, antiques and patents are amongst the most common. Each non-fungible asset class has a main set of variable traits (e.g., real estate has region codes, size, number of rooms and urban density) by which assets can be approximately grouped. The difference between fungible and non-fungible assets introduces complexities for estimating asset liquidity: in fungible assets, liquidity can be estimated using the trade-off between price, time and quantity of the sale, with stocks utilising a limit order book to facilitate transactions (Aidov & Lobanova, 2021). This three-way trade-off is not possible for non-fungible assets since the quantity is always one. Consequently, the analysis of non-fungible asset liquidity is reduced to a price-time trade-off, which can be found using an asset's transaction history.

3.2. Background information for NFT markets

Fundamentally, NFTs are unique units of data stored on a blockchain that can represent any conceivable asset such as art, legal contracts, real estate, intellectual property rights, or anything else that can have its ownership digitally recorded; this asset type represents the most recent non-fungible asset to gain global recognition (Rabaa'i et al., 2022). While there are some similarities with existing asset markets such as art, real estate, stocks and bonds, there are many crucial differences. Firstly, trading in the NFT market is relatively more complicated, with participants needing to create a crypto wallet and purchase the relevant cryptocurrency using fiat currency since most NFTs are not directly purchasable using fiat (there are many wallets that suit various needs such as Metamask, Phantom, Coinbase and Exodus (Hicks & Curry, 2023)). Next, participants need to find an NFT via an exchange, such as OpenSea, and pay the blockchain-specific gas fee upon purchasing the NFT. The advantage of this process is that one can remain completely anonymous, and wallets are free to create, lowering the barriers to entry. Secondly, the NFT market is almost entirely unregulated and highly susceptible to market manipulation, which is discussed in section 4.5. Thirdly, the NFT market is always open, although its times of peak activity often align with traditional markets (Umar et al., 2022). Fourthly, all NFT activity is completely transparent to all observers, allowing everyone free access to significant amounts of data, unlike stock markets that sell high-frequency data as a source of revenue (ASX, 2023; NYSE, 2023).

Although the use cases for NFTs are wide (e.g., copyright tracking, anti-piracy, supply network optimisation, etc.), this paper focusses on the art and collectibles NFT market, given their prominence in secondary market activity. **Figure 1** demonstrates the mechanism of NFT creation and reselling: initial NFT sales are facilitated directly by the creators (primary market minting), while secondary trades take place on either centralised or decentralised exchanges (secondary marketplaces), enabling individuals to list and bid on NFT auctions (Oh, 2023). Individual NFTs are either standalone or part of collections, with the majority of NFT trading volume residing in the latter. Collections vary in size from 10 NFTs to 10,000 and contain numerous traits of varying rarities, with some famous examples being Crypto Punks and BoredApesYachtClub. Collections provide an easy way to categorise and price NFTs, similar to the way location, number of bedrooms or total area can be used in real estate. However, comparing NFTs outside collections becomes increasingly difficult due to a significant diversity of price points, liquidity and demand: for example, within the OpenSea marketplace, 39% of NFTs have never been sold since minting, and 92% have been sold three times or less, despite an average market return 2.5% per week that exceeds almost all traditional asset classes (Borri et al., 2022; Lommers et al., 2022). These highly skewed resale and return distributions are also present in the physical art market and have two possible reasons: firstly, investors identify certain NFTs as buy-and-hold/long-term investments, opting not to sell within a year of purchase, and secondly, the majority of collections never gain enough popularity to be frequently traded, causing 8% of assets to constitute the majority market volume, returns and trade counts (Renneboog & Spaenjers, 2013).

Figure 1
Overview of NFT Markets by Oh (2023).



A simplified model of NFT transactions from "Market Manipulation in NFT Markets" by Oh (2023). This paper only focuses on the secondary market.

4. Literature Review

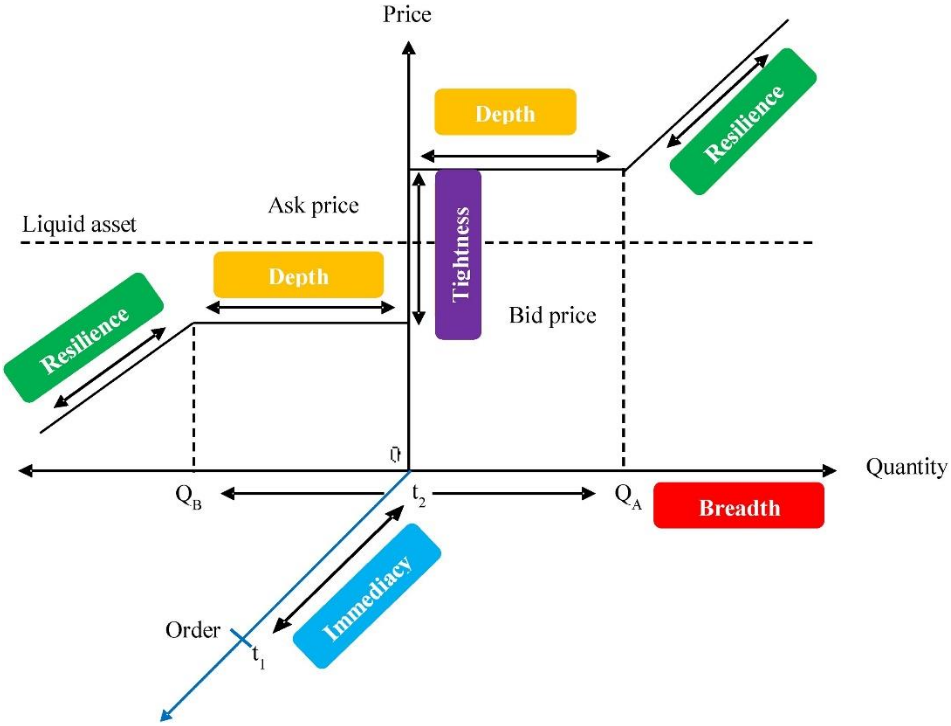
4.1. NFT literature

Early NFT literature focussed primarily on inter-asset relationships that identifies NFTs as a completely unique asset class. The relationship between NFTs and cryptocurrencies seems to be tenuous due to different analytical techniques producing divergent findings. Apostu et al. (2022) find a significant relationship between NFTs and Ethereum, but not other cryptocurrencies, using time-series analysis and Granger causality tests, whilst Dowling (2022) simultaneously finds limited evidence of volatility transmission between the two asset classes utilising spillover indices, and statistically significant co-movement between the two markets using wavelet coherence analysis. Expanding the scope, Karim et al. (2022) find NFT markets to be highly disconnected from the greater blockchain financial system and highlight significant diversification avenues. Looking at traditional asset classes, Zhang et al. (2022) find NFTs to be useful hedges against stock, bonds and fiat currency, further solidifying NFTs as independent of all major asset classes.

4.2. Quantifying liquidity in traditional assets

Compared to the relatively sparse NFT literature, asset liquidity is extensively covered within equity markets (e.g., Abdi & Ranaldo, 2016; Corwin & Schultz, 2012; Goyenko et al., 2009), bond markets (Schestag et al., 2016), currency markets (Karnaukh et al., 2015; Mancini et al., 2012) and commodity markets (e.g., Marshall et al., 2012), with a full literature survey being conducted by Marshall et al. (2018). Within these traditional markets, the previously mentioned three-way liquidity dynamic (price-quantity-time) is expanded into five dimensions and visualised in **Figure 2**.

Figure 2
Five dimensions of liquidity by Díaz & Escribano (2020).



The above illustration portrays the five dimensions characterizing liquidity: market depth, tightness, breadth, immediacy, and resilience. Q_A and Q_B denote available quantities at the prevailing bid and ask prices, representing the concept of market depth. Tightness is assessed by the bid-ask spread. Market breadth pertains to the distribution of orders across different price levels, while resilience quantifies the market's capacity to rebound from unforeseen disruptions. Lastly, t_1 and t_2 correspond to the time an order is introduced and subsequently executed, effectively quantifying immediacy, which is the speed of order execution. This interpretation draws inspiration from works by Bervas (2006) and Hibbert et al. (2009).

The five key dimensions of liquidity are as follows: market breadth (trading volume of the existing orders at different prices), market depth (number of orders around equilibrium prices), market immediacy (the speed of order execution), market resilience (ability to absorb and recover from unexpected asset shocks), and market tightness (trading costs of turning around a position) (Díaz & Escribano, 2020). To analyse these dimensions, researchers acquire high-frequency data, which are often costly, computationally cumbersome, and generally inaccessible to the public. As such, liquidity proxies, which use readily available low-frequency data, are used to approximate the various dimensions of market liquidity without the issues associated with high-frequency data. Fong et al.'s (2017) study of global equities identified low-frequency liquidity proxies that accurately capture various dimensions of liquidity with 1,000 to 10,000x the computational savings compared to traditional high-frequency measures. However, since liquidity is a multi-dimensional topic and the commonly used proxies merely approximate a handful of these dimensions, there is no one-size-fits-all liquidity proxy (Będowska-Sójka & Echaust, 2020). As such, the existing literature has developed a wide array of liquidity proxies, with some of the most popular being developed by Amihud (2002), Roll (1984), Lesmond et al. (1999), Holden (2009) and Pástor & Stambaugh (2003).

4.3. Liquidity in non-fungible assets

Given the unique nature of non-fungible asset markets with regards to liquidity, such as the lack of limit order book, significant depth or low-transaction costs, it is necessary to look more closely at liquidity research in two of the largest non-fungible asset markets, the over-the-counter (OTC) contracts and real estate markets, which are respectively valued at USD $20.7 trillion (Bank for International Settlements, 2023) and USD $3.7 trillion in 2022 (Precedence Research, 2023). At a glance, within OTC markets, Davis et al. (2023) utilise Amihud's measure and day volatility as proxies for OTC bond liquidity to analyse their relationship with the bond's tier/quality. Deuskar et al. (2011) use bid-ask spreads as a proxy for OTC option liquidity to find that illiquid options trade at higher prices due to unique market features, which contrasts with the literature consensus. Jankowitsch et al. (2011) create a custom price dispersion volatility measure for estimating OTC bond market liquidity as a way to further understand liquidity drivers. Real estate, due to its widespread appeal as a real asset and long holding periods, has a much larger literature base surrounding liquidity. A theoretical study by Krainer (2001) defines liquidity as the probability of selling at an equilibrium market price and finds that liquidity can be high while prices are high due to transaction costs, which touches on one of the key liquidity proxy categories used in later research. Ametefe et al. (2016) formalise the existing body of research and categorise a myriad of liquidity measures as transaction cost-based, volume-based, price impact-based, time-based and return-based. A full summary of their categorisations can be found in **Table 1**. With the majority of subsequent research utilising these measures or modifications of them, the literature primarily focusses on the real estate liquidity's relation to aggregate shock dynamics/market cyclicality, and the applicability of such measures to collections of residencies. Analysing the former, Kotova & Zhang (2021) and Chernobai & Hossain (2019) both find that real estate liquidity proxies perform best during boom years and are highly seasonal. Regarding the latter, with most studies grouping assets by geography, there is a general consensus that there is significant intra-grouping heterogeneity (e.g. single-family homes vs multi-family homes (Irwin & Livy, 2022), liquid vs illiquid dwellings (Cajias et al., 2020), distance from city centre (Chernobai & Hossain, 2019), etc.), as well as inter-regional heterogeneity.

For the purposes of analysing liquidity, real estate provides the best comparison point for NFTs. Both asset classes are distinctly non-fungible, and can be grouped by common features with varying degrees of heterogeneity (for example, a property's builder is analogous to an NFT's creator, special features to rarity, number of rooms to number of attributes, etc.) and allows parallels to be drawn between neighbourhoods of houses and collections of NFTs with regard to pricing, liquidity and market microstructure. As such, real estate serves as the primary benchmark for understanding NFT liquidity due to the lack of published papers on this specific topic.

Table 1
Common real estate liquidity proxies from Ametefe et al. (2016).

Category	Liquidity proxy	Liquidity dimension
Transaction costs	Absolute & relative quoted spread	(1)
	Effective & relative effective spread	(1)
Volume-based measure	Transaction volume	(4)
	Turnover ratio	(4)
	Quote size	(2)
	Number of bids	(4)
	Market depth	(2)
Price impact	Amihud’s measure	(2, 3)
	Regressed lambda	(2, 3)
	Pastor-Stambaugh liquidity factor	(2, 3)
	Lesmond Zeros (returns)	(2)
	Goyenko Zeros (volume)	(2)
	Market efficiency coefficient	(3)
Time-based measures	Holding periods	(4)
	Trading frequency	(2)
	Volumes volatility	(2, 4)
	Time on market	(5)
Return-based measures	Roll’s measure	(3)
	Run-length	(2, 3)

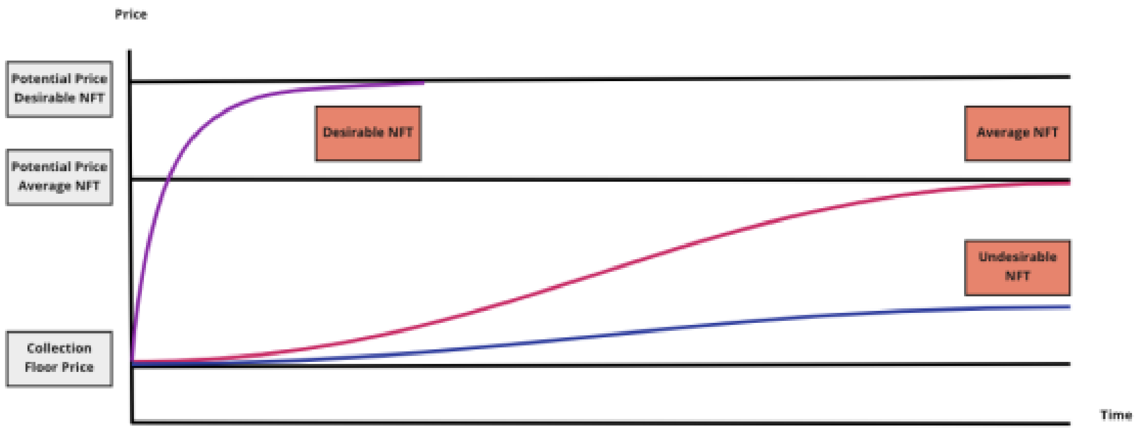
Liquidity dimensions key: (1) Tightness; (2) Depth; (3) Resilience; (4) Breadth; (5) Immediacy.

4.4. Liquidity in NFTs

Currently, the primary paper empirically analysing the determinants of NFT liquidity solely uses Amihud's (2002) illiquidity measure to calculate daily NFT illiquidity, from which they find that NFTs follow the same weekday liquidity patterns as stocks and that liquidity increases with asset age, number of trades and blockchain-related media coverage (Wilkoff & Yildiz, 2023). While the findings are robust, the paper does not discuss the appropriateness of Amihud’s measure. Furthermore, the sample size is restricted to the top quartile of NFT transactions, which could potentially bias the results due to Yousaf & Yarovaya's (2022) detection of the presence of asymmetric connectedness between volume and returns at market extremes. In contrast to their empirical approach, Lommers et al. (2022) propose a theoretical liquidity model in which NFTs are either desirable, average or undesirable and observe the following price mechanism, although it is not quantified: an NFT's expected sale price begins at the collection floor price and asymptotically approaches a theoretical potential price as time progresses (see **Figure 3**). As such, it is apparent that there is a significant gap in the understanding of which liquidity measures are appropriate to use within the NFT asset class and how to quantify illiquidity.

Figure 3

Lommers et al.'s (2022) theoretical model of NFT liquidity.



The above model proposes that NFTs within a collection can be categorised as either "desirable", "average", or "undesirable" based on their randomly generated features. For example, within the Bored Apes collection, some assets have gold backgrounds, which are seen as highly desirable compared to more generic colours. Lommers et al. (2022) propose that each class of NFT will asymptotically approach a theoretical potential price as time progresses. Note that not all NFT collections mint NFTs with unique features – some mint near-identical copies of the same asset for multi-person use.

4.5. Wash trading

The Internal Revenue Services (IRS) defines wash trading as "trades [that have an] absence of change in ownership," or more simply described as a single entity trading an asset between its accounts to give the illusion of demand for an asset at an inflated price (Oh, 2023). In attempts to generate illicit profit or exploit marketplace incentive programs, this practice inadvertently causes significant issues when studying asset liquidity (Imisiker & Tas, 2016; Serneels, 2023). Section 6.1 provides further detail and an example of such wash trading.

Within NFT markets, wash trading is abetted by the fact that market participants can create an infinite number of accounts and utilise automation to perform wash-trades far beyond what is possible in traditional markets. However, given the total transparency of the blockchain, wash-trades are significantly easier to detect to an observer.

Given the unregulated nature of the NFT market, it is critical to consider the effects of wash trading that could bias results when analysing asset volume and prices (Chalmers et al., 2022). By testing NFT market data for adherence to Benford's Law, Student's T-Tests and the Pareto-Levy law, Sifat et al. (2023) identify significant levels of wash trading in the Ethereum and Wax blockchains. Their findings support earlier market-level graphical analysis conducted by von Wachter et al. (2021), who find that at least 2% of trades within the top 52 traded NFT collections are wash-trades. To supplement this, Oh (2023) leverages the abundance of transaction-level data to directly identify individual wash-trades and traders: he proposes a logic framework, based on the IRS' guidance, to identify wash-trades utilising wallet addresses and estimates that only 0.3% of transactions are wash-trades, which are primarily clustered in popular NFT collections. Finally, Serneels (2023) proposes three novel strategies to target simple wash-traders, who use a small number of personally managed wallets, as well as sophisticated wash-traders, who automate wallet generation to leverage dozens of accounts simultaneously. Both Oh (2023) and Serneels (2023) posit that the goal of NFT wash trading is to accrue marketplace rewards that pay a fraction of trade volume, with Oh additionally finding that the returns of wash trading are negligible in the long term due to the fact that most participants can identify them. Overall, it is imperative to incorporate wash trading exclusion strategies to prevent manipulative price/volume distortions from biasing the results.

5. Data

The raw data is pulled directly from OpenSea, the largest decentralised NFT marketplace. OpenSea hosts approximately 97% of all Ethereum-based NFT transaction data which is the most popular blockchain for NFTs. The resulting dataset reports all NFT transaction data that has gone through the marketplace between 1 Feb 2021 and 21 Aug 2021. Due to the immutability of the blockchain infrastructure, the sourced data is free from tampering. The data is acquired via OpenSea's publicly available API, which limits data requests to 2 requests per second, requiring additional code to retrieve all transactions within the sample period. A snippet of the raw data for Hashmasks #6729 can be seen in **Figure 4**, with the primary data fields being described in **Table 2**.

Figure 4
Sample NFT transaction data point.


```
"asset.id": 17749891,
"asset.token_id": "6729",
"asset.num_sales": 5,
"asset.name": "Hashmasks #6729",
"asset.asset_contract.asset_contract_type": "non-fungible",
"asset.collection.name": "Hashmasks",
"created_date": "2021-02-01T23:59:12.820008",
"event_type": "successful",
"id": 77433831,
"payment_token.symbol": "ETH",
"payment_token.eth_price": 1,
"payment_token.usd_price": 4138.51,
"quantity": 1,
"seller.address": "0xfd867fa6fa9dea501dd21d84d218b9dfecc29678",
"seller.user.username": "DeFeeet",
"total_price": 1.3995e+18,
"transaction.id": 78222706,
"transaction.timestamp": "2021-02-01T23:58:37",
"winner_account.user.username": "dindon",
"winner_account.address": "0xea7d6a3873cbb644a2fa3a124b00a25c33c661b8",
"created_date_epoch": 1612223952,
"transaction.timestamp_epoch": 1612223917
```


Transaction data corresponding to the asset called "Hashmasks #6729", which sold for 1.3995 ETH at 11:59:37 PM on 1 Feb 2021 to wallet address 0xea7d6a3873cbb…

Table 2
Data field definitions and applied filters.

Data Field	Definition	Applied Filter
Asset.id	Unique asset identifier	N/A
Event_type	"Successful" indicates an asset sale	Only "successful"
Asset.asset_contract. Asset_contract_type	Describes the asset involved in the contract	Only "non-fungible"
Quantity	Number of NFTs involved in the sale.	Only trades with a quantity of 1
Payment_token.symbol	Symbol indicating the token used to pay for the NFT	Only ETH or WETH payments
Payment_token.usd_price	The price the above payment token in USD on that day	N/A
Total_price	The price of the asset in units of payment token	N/A
Seller.address	Wallet address of the seller	N/A
Winner_account.address	Wallet address of the buyer	N/A
Transaction.timestamp	Date and time of the transaction	N/A
Transaction.id	Unique transaction identifier	Remove duplicate transactions

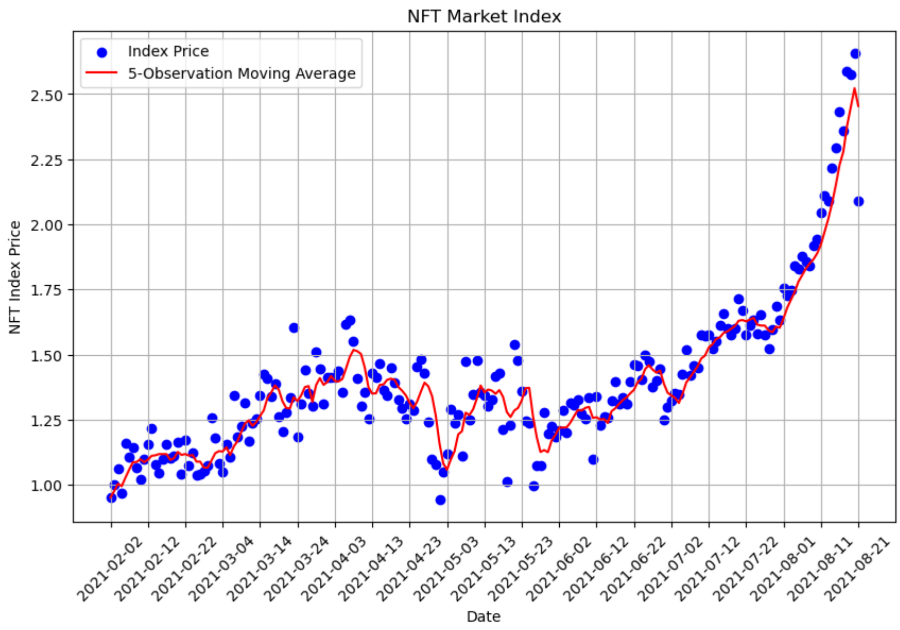
This table displays the primary data fields pulled from OpenSea, associated definitions and any filters applied. From these primary datafields, the USD sale price of the asset can be calculated by multiplying payment_token.symbol and payment_token.usd_price. The payment_token.usd_price value is updated at the start of each day.

The dataset retrieved from OpenSea is raw and unstructured, containing duplicates, incomplete transactions and unusable information. Following prior literature, the following filters are applied to create a cleaned dataset: only keep transactions involving NFTs, successful sale events, single NFTs at a time, payments in ETH or WETH and the removal of all transactions with missing fields or duplicate transaction IDs (Kampakis, 2022; White et al., 2022a; Wilkoff & Yildiz, 2023). In order to create a series of multiple returns, only assets with three or more recorded sales are kept.

Prior to analysing the dataset, it is important to understand the context of the NFT market with regard to returns and perceived volatility. Following Borri et al.'s (2022) methodology, who respectively create a repeat sales model following Bailey et al.'s (1963) work in real estate, an NFT market index is displayed in **Figure 5**. This index shows a clear overall bull run across the sample: although there is a month-long market pullback during April, the index rebounds and fluctuates throughout May before peaking at 2.5 by the end of the sample period. **Figure 6** by Borri et al. (2022) shows the NFT market index from 2018 to the end of 2021; it is clear that, despite the slight downturn in this paper's sample period, the NFT market is on an unprecedented bull run that began in late 2020 and ended in late 2021. At its peak, the NFT market index price was 10 times higher than its start point in 2018.

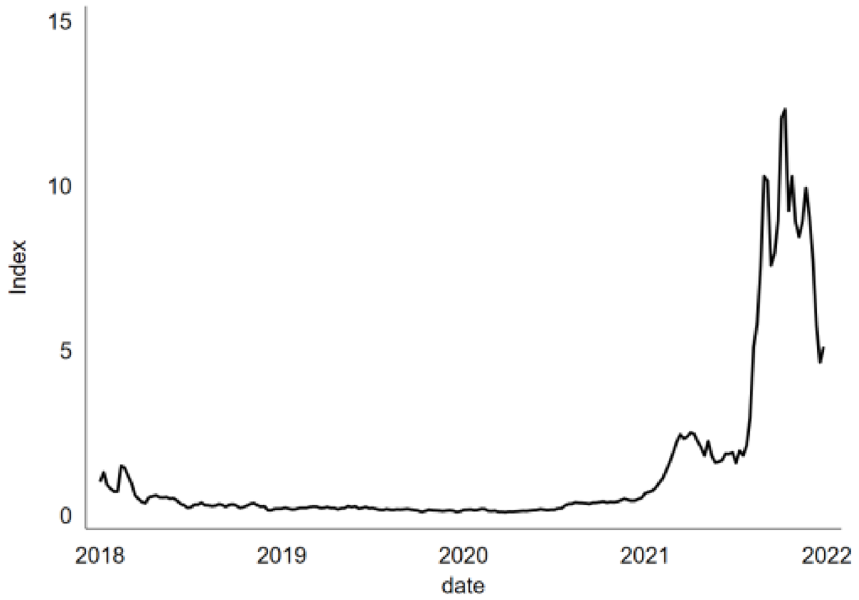
Overall, there are 1,009,177 unique transactions, 737,161 unique assets and 1,985 unique collections, with the first and last recorded trades being on the 1st of February and 21st of August 2021. The median NFT sale price is $375.14, while the mean is $1,714.12, indicating a significant positive skew. The average time between sales for the median NFT is 4.91, however, a significant number of NFTs observe average sale times of less than 2 days, indicating the presence of some highly active NFTs in the market. **Table 3** breaks down the data by individual NFT and collections.

Figure 5
NFT market index constructed via the repeat sales method of Bailey et al. (1963).



The market index is constructed using the cleaned dataset from OpenSea. The sample period spans from 1 Feb 2021, when the index price is 1, to 21 Aug 2021.

Figure 6
NFT market index from "The Economics of Non-Fungible Tokens" using the repeat sales method (Borri et al., 2022).



This figure by Borri et al. (2022) uses data from multiple exchanges (Cryptokitties, Gods Unchained, Decentraland, OpenSea and Atomic) and excludes NFTs under $1. It calculates weekly index prices from 2018 to the end of 2021.

Table 3
Descriptive statistics by NFT and collection level.

Panel A: NFT Level Statistics								
Variable	N	Mean	Std Dev	Min	25%	Median	75%	Max
Number of sales	737,162	2.37	10.93	2	2	2	2	5,928
Total volume (USD)	737,159	2,553.92	16,835.56	0.00	175.89	465.03	1,394.89	4,786,104.42
Sale price (USD)	737,159	1,714.12	12,445.22	0.00	163.88	375.14	1,040.29	3,351,120.00
Return (%)	181,759	523.75	86,068.23	-100.00	-7.55	33.06	107.14	33,806,037.69
Days to sell	183,089	12.79	22.63	0.00	1.38	4.91	13.92	199.09
Panel B: Collection Level Statistics								
Number of sales	1,986	508.15	1,650.71	1	2	9	136	18,974.00
Total volume (USD)	1,983	949,391.13	7,387,968.84	0.00	843.15	6,817.26	109,497.57	227,787,854.38
Sale price (USD)	1,983	4,050.33	26,139.98	0.00	158.99	468.94	1,682.95	603,804.71
Return (%)	893	1,004.42	21,154.51	-100.00	5.11	47.79	139.72	629,117.48
Days to sell	901	21.57	29.01	0.00	3.25	10.51	28.63	185.99

This table presents descriptive statistics for each NFT or collection's number of sales, total USD volume (sum of all trades' USD value), sale price USD (average sale price in USD), return (average return) and days to sell (average days to sell). The data set contains 1,009,177 unique transactions, 737,161 unique assets and 1,985 unique collections between 1 Feb 2021 and 21 Aug 2021.

Table 4
Collection-day descriptive statistics.

Variable	N	Mean	Std Dev	Min	25%	Median	75%	Max
Return (%)	31,388	265.01	8,818.00	-100.00	-10.00	0.00	21.06	915,012.25
Volume (USD)	31,531	56,800.00	397,000.00	0.00	541.51	2,797.38	14,981.27	23,400,000.00
Volatility (%)	31,009	2,537.65	11,498.37	0.00	0.00	66.67	570.11	99,680.55
Age (days)	31,531	61.49	55.32	1.00	15.00	41.00	101.00	200.00
HHI	31,531	1,379.24	1,371.39	267.23	553.35	839.22	1,502.01	8,979.76

This table presents descriptive statistics by collection-day observation using the following data definitions: return is the difference between the logged values of the collection c's close and open prices on day d, volume is the sum of all sale prices within c on day d, volatility is the percent difference between c's high and low prices on day d, age is the days since the c's first sale and HHI is the Herfindahl–Hirschman index, calculated as the sum of the squares of each collections percentage market share on day d. The data set contains 1,985 unique collections and 31,531 collection-day observations between 1 Feb 2021 and 21 Aug 2021.

Notably, there is a significant skew in the individual NFT data, with the majority being traded twice within the sample period and within two weeks of each other. Regarding pricing, there is a significant tail on the positive side, with a minority of assets creating the vast majority of trading volume and being priced significantly higher than the majority. While this phenomenon, more widely known as the generalised Pareto distribution or power laws, is absent from traditional financial assets such as stock indices, energy prices, exchange rates, interest rates, and precious metals (Liu, 2019), it is prominent in creative/non-fungible assets such as art (Etro & Stepanova, 2018), music (Gustar, 2020) and real estate (Blackwell, 2018).

One of the most noteworthy components of the data is the average returns and extremely short resale period; the median NFT generates a 33% return across an average of 4.91 days, while NFTs at the extremes can either become worthless or witness unprecedented growth. Like the apparent Pareto distribution, these features are common among physical art, music, and real estate markets.

Descriptive statistics by collection-day observations can be seen in **Table 4**. Notably, the median daily return for collections is 0%, with extremely fat tails on either side going to 100% loss and 915,012% return. The extreme kurtosis of the returns is reinforced by the daily volatility (calculated as the day high divided by the day low) having an extreme upside and relative flat majority. The Herfindahl–Hirschman index (HHI), which is calculated as the sum of the squares of each collection's percentage market share for each day and has a theoretical range of 0 (infinite number of collections) to 10,000 (only one collection trading). It reveals a highly dispersed market for 50% of collection-day observations (HHI<1000) and that the majority of data points occur in relatively unconcentrated markets (HHI<1500) (U.S. Department of Justice, 2018).

6. Liquidity proxies

To explore the liquidity of NFT collections, a suite of commonly used liquidity proxies (Amihud's illiquidity ratio (Amihud, 2002), Lesmond's zero measure (Lesmond et al., 1999), Goyenko's modified zero measure (Goyenko et al., 2009) and Roll's implicit measure(Roll, 1984)) is calculated using four unique datasets and analysed using a Pearson correlation matrix and family of regressions. Finally, a set of robustness tests is performed, including dividing the dataset into bull/bear periods and taking the natural log of the liquidity proxies to detect non-linear relationships. The datasets are created in two steps: firstly, wash-trades are excluded from a duplicate dataset, resulting in a pre and post-wash trading exclusion dataset. Secondly, for each dataset, two different methodologies of calculating liquidity proxies are applied, the first assuming intra-collection homogeneity and the other assuming intra-collection heterogeneity. With the resulting four datasets of liquidity proxy calculations, a simple OLS regression, a collection fixed effect regression and a collection/date fixed effect regression are performed on each of the liquidity proxies. The proxies are then normalised and logged to perform additional analysis into the drivers of liquidity in NFT collections.

6.1. Excluding wash-trades

As made apparent by a multitude of research, wash trading, regardless of its motive, has a major presence and notable impact in almost all dimensions of the blockchain financial system, impacting returns, observed trading volume, market health and perceived liquidity (Oh, 2023; Pennec et al., 2021; Serneels, 2023; Sifat et al., 2023; von Wachter et al., 2021).

To fully understand how wash trading affects the NFT collection liquidity, it is necessary to compare the performance of the proxies with and without the presence of wash-trades in the data. Possible processes for identifying individual wash-trades are outlined in **Table 5**, which collates the strategies proposed by Oh (2023) and Serneels (2023).

Table 5
Wash-trade identification strategies by Serneels (2023) and Oh (2023).

Author	Strategy Name	Criterion
Oh (2023)	Identity Trade	A sell to A
	1-1 Trade	A sell to B and B to A (within 7 days)
	Matched Order	A sell to B, B to C and C to A (within 7 days)
Serneels (2023)	Closed-Loop Token (CLT)	A sell a token and the token ever returns to A
	Closed-Loop Value (CLV)	A sell to B within 2% of the previous price
	High Transaction Volume (HTV)	Trading volume > 99% confidence range

Given the aggressiveness of HTV and the fact that CLV will incorrectly flag sideways trading (zero return trades), it is necessary to consider which flags to use when eliminating wash-trades, which requires balancing the capturing of simple vs sophisticated wash-traders. Simple wash-traders use a small number of wallets to manually trade assets and can be identified entirely by changes in ownership addresses (Oh, 2023). As such, all of Oh's strategies and Serneels' CLT strategy are able to capture this category of manipulative trading. Sophisticated wash-trades, however, utilise automation to trade throughout multiple accounts and can potentially create a new account for every NFT trade due to the market's near-zero barriers to entry, meaning that ownership tracking is not enough to detect these trades. Serneels' CLV and HTV strategies can detect these trades, although they introduce false positives (type I errors) when an asset is trading sideways (0 return trades between legitimate participants) and when an NFT generates significant investor attention (abnormal increases in transaction volume). Research by Serneels (2023) indicates that simple wash-traders make up the majority of illicit transaction volume in NFT markets and that the CLV/HTV strategies tend to generate excessive amounts of type I errors and occasional type II errors. As such, Serneels' CLT methodology, which fully captures all of Oh's strategies, is used as it completely captures the majority of all wash trading and does not generate any type I or II errors. Applying this methodology, each trade is then flagged as either 1 (suspicious) or 0 (normal) for each of the criteria. **Table 6** shows a wash-trading example of the asset "Eyes Open Bomb Squad" from the collection SuperMassive V2.

Table 6
Example of a wash traded NFT.

Asset Number	Seller	Buyer	ETH Price	USD Price	Date	CLT	CLV (USD)	CLV (ETH)	Identity Trades	1-to-1 Trades	Matched Orders
16886921	0x216b44c0…	0x1ff0caaf7…	0.065	217.82	8 Jul 2021 5:41	1	0	0	0	1	0
16886921	0x1ff0caaf7…	0x216b44c0…	0.055	184.31	8 Jul 2021 7:16	1	1	1	0	1	0
16886921	0x216b44c0…	0x1ff0caaf7…	0.055	182.61	8 Jul 2021 13:05	1	1	1	0	1	0
16886921	0x1ff0caaf7…	0x216b44c0…	0.055	182.61	8 Jul 2021 18:50	1	1	1	0	1	0
16886921	0xffaeae51f…	0x909b4ce8…	0.055	184.31	9 Jul 2021 6:42	1	1	1	0	1	0
16886921	0x909b4ce8…	0xffaeae51f…	0.055	184.87	21 Jul 2021 6:47	1	1	1	0	1	0
16886921	0xffaeae51f…	0x909b4ce8…	0.055	184.87	21 Jul 2021 7:28	1	1	1	0	1	0
16886921	0x909b4ce8…	0xffaeae51f…	0.055	184.87	21 Jul 2021 7:32	1	1	1	0	1	0
16886921	0x69c3d3df…	0x6307ce90…	0.055	184.87	21 Jul 2021 11:39	1	1	1	0	0	0

This table presents an NFT that has been wash-traded through multiple accounts. The above asset is called “Eyes Open Bomb Squad” from the collection “SuperMassive V2” and is wash-traded through four unique accounts. On 8 Jul 2021, the asset is traded 4 times between the same two parties (blue and red accounts) until it is traded to a third party outside the OpenSea marketplace. The asset then reappears on OpenSea on 9 Jul 2021 and is wash-traded between the purple and orange accounts. Note that the ETH price is the exact same for all trades after the first transaction. As such, many of its trades are flagged by CLT, CLV (USD and ETH) and 1-to-1 strategies. While the price does not change in this example, the manipulators have artificially increased the perceived demand for this asset, which is a key indicator for many traders when valuing assets and analysing liquidity.

Table 7
Wash trading descriptive statistics.

Panel A: Distribution of trades affected by wash trading

Number of flags	Number of trades flagged	Percent of trades flagged	Cumulative percent flagged
0	790,083	78.29	78.29
1	185,433	18.37	96.66
2	11,873	1.18	97.84
3	21,614	2.14	99.98
4	165	0.02	100.00
5	7	0.00	100.00

Panel B: Wash trading statistics by trades, assets, collections, and dollar value

	Trades		Assets		Collection		Dollar value	
Flag	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Identity	7	0.00%	1	0.00%	1	0.05%	$35,604.97	0.00%
One-to-one	791	0.08%	353	0.05%	154	7.57%	$3,320,990.50	0.18%
Matched	0	0.00%	0	0.00%	0	0.00%	$0.00	0.00%
CLT	24,070	2.39%	679	0.09%	244	11.99%	$17,942,152.25	0.95%
CL-USD	34,521	3.42%	7,383	1.00%	491	24.14%	$40,314,475.33	2.14%
CL-ETH	36,300	3.60%	8,253	1.12%	531	26.11%	$42,553,584.10	2.26%
HTV	179,027	17.74%	179,027	24.29%	874	42.94%	$585,609,193.67	31.11%

This table shows descriptive statistics for the wash trading flags in the dataset using the strategies defined in Table 6. In this table, Closed Loop Value (CLV) identification is applied using both USD value (CL-USD) and ETH (CL-ETH) with a margin of error of 2%. Panel A shows the flagging frequency for trades in the data. Panel B shows statistics regarding the number of trades, assets, collections and dollar amount affected by each flag. The data set contains 1,009,177 unique transactions, 737,161 unique assets and 1,985 unique collections between 1 Feb 2021 and 21 Aug 2021.

Looking at the descriptive statistics provided in **Table 7**, each identification strategy becomes increasingly aggressive and increasingly likely to include false positives. With over 10% of trades having at least one flag raised, it is critical to balance false positives with false negatives. As noted in their respective papers, both Serneels and Oh's identification strategies have their downsides, with Oh's method underreporting wash-trades (his methodology does not recognise wash-trades between over 3 accounts) and Serneels overreporting via his high transaction volume flag. Furthermore, Serneels' CLV strategies will overidentify assets that have been trading sideways (i.e. zero returns), which does not necessarily mean wash trading. As such, the optimal middle ground is to remove all trades flagged either by Serneels' CLT strategy or any of Oh's strategies. Looking at panel B of **Table 7**, it can be noted that CLT affects 2.39% of trades, 0.09% of assets and 11.99% of collections, which indicates that a small number of assets, spread across multiple collections, generates most of these wash-trades. Serneels (2023) argues that CLT has an extremely low chance of generating false positive flags, as it is highly unlikely that an NFT seller will later come back to buy that asset.

Table 8 shows the post-wash trading exclusion collection-day descriptive statistics. As noted above, 24,070 (2.39%) trades were removed due to wash trading identification, with only a 1.5% decrease in collection-day observations. This is likely due to most detected wash-trades being surrounded by legitimate trades within popular assets. The majority of collection-day statistics remain mostly unchanged, with collection open and close prices seeing a reduction in kurtosis (reduced 25% percentile value and raised the 75% percentile value), day volatility seeing a slight rise at the 50^{th} and 75^{th} percentiles and the obvious effects of day transaction volume falling and age rising. However, a noticeable change can be observed in HHI: the average HHI rises by 50, the bottom 75% of values rises and its 90^{th} percentile falls, indicating a generally more fragmented market. This is expected due to the concentration of wash trading within a few assets, which artificially increases their market shares.

Table 8
Post-wash trading exclusion descriptive statistics by collection-day.

Variable	N	Mean	Std Dev	Min	25%	Median	75%	Max
Return (%)	30,926	264.91	8,866.00	-100.00	-10.00	0.00	21.96	915,012.25
Volume (USD)	31,051	55,470.89	370,950.92	0.00	541.51	2,792.34	14,949.16	23,442,136.14
Volatility (%)	30,572	2,547.57	11,490.85	0.00	0.00	66.67	584.21	99,640.00
Age (Days)	31,051	61.85	55.47	1.00	15.00	42.00	101.00	200.00
HHI	31,051	1,428.61	1,484.21	268.83	548.56	866.02	1,590.05	8,979.76

This table presents descriptive statistics by collection-day observation after excluding wash trading and using the following data definitions: return is the difference between the logged values of the collection *c*'s close and open prices on day *d*, volume is the sum of all sale prices within *c* on day *d*, volatility is the percent difference between *c*'s high and low prices on day *d*, age is the days since the *c*'s first sale and HHI is the Herfindahl–Hirschman index, calculated as the sum of the squares of each collections percentage market share on day *d*. The data set contains 1,985 unique collections and 31,051 collection-day observations, as opposed to 31,531 observations pre-wash trading exclusion, between 1 Feb 2021 and 21 Aug 2021.

6.2. Collection level assumptions of homogeneity

There are two primary ways to calculate a collection's liquidity proxy on a particular day, each with its own pros, cons and underlying assumptions. The first method is to treat all NFTs within a collection as identical, assuming intra-collection homogeneity; this method has the advantages of simplifying data acquisition for future research due to collection level price data being more readily available and ensuring a well-populated dataset with a larger number of trades per day when compared to tracking individual NFTs. However, assuming homogeneity within collections is a significant assumption that will hold true for some collections (e.g., some collections mint extremely similar assets with minimal discernible differences), but false for others (e.g., within the Bored Apes collection, some NFTs possess highly rare attributes, such as hats or accessories, that cause them to sell many times higher than the average NFT). Under this assumption, the unique asset identifiers are stripped and liquidity proxies are calculated treating each recorded sale price as a sale of the same asset. The second method is to calculate liquidity proxies for each individual NFT and then average them within each collection day observation, assuming intra-collection heterogeneity; this method has the advantage of being more precise and accounting for intra-collection variance. The downsides of using the heterogeneous collection assumption are that it is more computationally intensive and provides low observation counts for the majority of NFTs, which can introduce skewed proxy distributions or survivorship bias in proxies that require high observation counts. Furthermore, the proxies are winsorised at the 1 and 99 percentile levels to minimise the impact of outliers.

6.3. Illiquidity measure definitions

Despite there being a multitude of possible liquidity proxies that are thoroughly discussed in the surrounding literature of equities (Ramos & Righi, 2020), ETFs (Marshall et al., 2018), bonds (Schestag et al., 2016), OTC contracts (Jankowitsch et al., 2011) and real estate

(Ametefe et al., 2016), many have stringent data requirements/inputs and are not conducive to non-fungible asset analysis. Having considered this, it is possible to analyse liquidity across two primary dimensions in the NFT markets: market depth, which can be interpreted as price impact, and market tightness, which can be interpreted as the combination of explicit and implicit trading costs. Within market tightness, explicit costs are known upfront and include costs such as blockchain gas fees, taxes, and exchange fees, whereas implicit costs as less visible but highly significant in total transaction costs, encompassing costs such as bid-ask spreads, price slippage, and trade execution timing (Le & Gregoriou, 2020). Given the scope of this paper, only the Amihud illiquidity ratio (price impact), Lesmond zero measure (explicit transaction cost), Goyenko zero measure (explicit transaction cost) and Roll's implicit measure (implicit transaction cost) will be considered.

Notation is defined as the following: i refers to an individual NFT, s refers to a sale of i, c refers to an individual collection, d refers to the day and m refers to the month.

Amihud's (2002) illiquidity ratio, often referred to as the return to volume ratio, is a widely used measure in finance literature to assess the sensitivity of a stock's average absolute return to changes in trading volume (Amihud, 2002; Barardehi et al., 2021). It effectively encapsulates the influence of trading volume on stock price movements, with higher trading volume resulting in a lower illiquidity ratio; it also has a narrow interpretation as an implicit transaction cost estimator, with a higher illiquidity ratio indicating a wider bid-ask spread (Acharya & Pedersen, 2005). However, it has limitations, including a size bias, where larger market capitalization assets are automatically deemed less illiquid, and a failure to account for variations in trading frequency, which are increasingly relevant in contemporary markets and can impact liquidity premia (Florackis et al., 2011). Following Wilkoff & Yildiz's (2023) methodology, this paper uses the daily Amihud measure. Assuming homogeneous collections, the daily collection Amihud measure for collection c is calculated as the absolute return from day d's open to close price, divided by the total trading dollar volume of collection c on day d in USD, as seen in Equation 1. The heterogeneous collection assumption Amihud measure is found by calculating the daily Amihud measure for each constituent NFT i of collection c by replacing the c with i in Equation 1; collection c's subsequent Amihud measure is the average of each NFT i's Amihud measure. If a collection or individual NFT experiences only one sale on a day, its return is 0, making the Amihud measure 0. Since Amihud's measure is a measure of asset illiquidity, a higher Amihud value implies a fall in asset liquidity, specifically indicating a rise in the price impact of a single additional dollar of transaction volume.

$$Amihud_{c,d} = \frac{|Total\ Return_{c,d}|}{Trading\ Volume_{c,d}} \tag{1}$$

A similarly popular measure was developed by Roll (1984) to implicitly measure the effective bid-ask spread using serial covariance, which suggests that illiquid assets should have strong auto-correlation patterns. Ametefe et al. (2015) identify a key benefit of Roll's measure as only requiring daily prices to be calculated, with some drawbacks being the assumption of stationarity (which is verified in **Appendix A**), market informational efficiency (which is a key feature of decentralised marketplaces) and a lack of meaningful interpretation when the sample serial covariance is positive (highly prominent in emerging markets and real estate). To remedy the final drawback, Goyenko et al. (2009) created a modified Roll measure which is used in this paper and seen in Equation 2. As with the Amihud measure, the homogeneous Roll proxy is calculated using Equation 2, while the heterogeneous Roll proxy is calculated by finding each constituent NFT's Roll measure and averaging them within each day observation. A higher Roll measure indicates a fall in asset liquidity, specifically indicating a rise in the implicit bid-ask spread and thus implicit trading costs.

$$Roll_c = \begin{cases} 2 * \sqrt{-cov(\Delta P_{c,s}, \Delta P_{c,s-1})}, when\ cov(\Delta P_{c,s}, \Delta P_{c,s-1}) < 0 \\ 0 \qquad , when\ cov(\Delta P_{c,s}, \Delta P_{c,s-1}) \geq 0 \end{cases} \quad (2)$$

Where $cov(\Delta P_{i,s}, \Delta P_{i,s-1})$ is the sample covariance between the change in price of trade *s* ($\Delta P_{i,s}$) and the previous trade ($\Delta P_{i,s-1}$).

The Zeros measure proposed by Lesmond et al. (1999), and later modified by Goyenko et al. (2009), is widely popular in the financial liquidity literature. It has been found to be highly correlated with other traditional liquidity measures in emerging markets and accurately estimate explicit transaction costs (Bekaert et al., 2007; Lee, 2011; Lesmond, 2005). Lesmond's measure (see Equation 3) assumes that zero return days should occur when the expected return does not exceed the transaction cost, whereas Goyenko's modification (see Equation 4) argues that highly illiquid assets will have a higher proportion of zero trading volume days more consistently than zero return days (Ametefe et al., 2015). The homogeneous and heterogeneous collection assumption proxies are calculated using the same methodology as Amihud and Roll: the homogeneous collection assumption proxy applies Equations 3 and 4, while the heterogeneous collection replaces *c* with *i* and then averages them within each collection day observation. Following Goyenko et al.'s (2009) methodology, the zero proxies are measured at the monthly frequency. A higher Lesmond or Goyenko measure indicates a fall in asset liquidity, specifically indicating a rise in the explicit transaction costs.

$$Lesmond_{c,m} = \frac{NR_{c,m}}{T_m} \quad (3)$$

$$Goyenko_{c,m} = \frac{NV_{c,m}}{T_m} \quad (4)$$

Where *NR* is the number of zero-return days for collection *c* in month *m*, *NV* is the number of zero trading volume days for collection *c* in month *m* and *T* is the number of trading days in month m, which is always the number of days in the month since NFT markets are always open.

6.4. Comparison of liquidity proxy methodologies

Having calculated the liquidity proxies using both the homogeneous and heterogeneous collection assumption methodologies in both the pre and post-wash trading exclusion datasets, the liquidity proxies' descriptive statistics are found in **Table 9**. Differences between the panels of **Table 9** can help reveal both the information captured by the homogeneous/ heterogeneous collection assumptions and the impact of wash trading on collection liquidity. Comparing panels A to C and B to D, one of the most noticeable differences between the homogeneous and heterogeneous assumptions is the abundance of zeros: in the former, only 25% of Amihud and Roll values are 0, whereas the latter has 75% of values being 0. This difference is expected due to the abundance of individual NFTs that either sell once in a day, generating a 0 value for Amihud, or experience a consistent rise in sale prices, causing the return autocorrelation measure to be positive and thus Roll's measure to be 0. This issue does not appear in the homogeneous assumption datasets, which implies that it contains both ample daily transaction volume to generate a majority of non-zero Amihud measures and a noticeable amount of price pullbacks to generate a majority of negative return autocorrelation and thus non-zero Roll measures. Furthermore, the homogeneous assumption appears to generate a significantly smoother distribution of liquidity values that spans the full range of possible values for each proxy, as opposed to the heterogeneous dataset that witnesses significant clustering near the upper/lower bounds of each proxy, which is primarily noticeable in the Lesmond/Goyenko measures. Overall, these differences indicate that the homogeneous assumption provides a framework more conducive to studying liquidity, as a wider range of proxy values are attainable, there are fewer outliers, and the observable sale prices more closely represent the trading activity that is capturable by traditional proxies.

There are also noteworthy differences between the descriptive statistics before and after excluding wash trading. Within the homogeneous assumption datasets, the exclusion of wash-trades sees a rise in all non-zero values of Amihud and Roll measures, implying a fall in liquidity across both the price-impact and transaction cost dimensions. However, Lesmond reports a very slight decrease in its values around the median of its distribution, most likely indicating a shift in the shape of the distribution and containing minimal information on liquidity. Across the heterogeneous assumption proxies, the max and mean values for both Amihud and Roll fall, while Lesmond and Goyenko experience a slight rise in their means; these changes are minimal and likely reflect a reduction in the number of trades as opposed to a fundamental shift in liquidity. Overall, the removal of wash-trades does make a change in calculated liquidity, although, it is difficult to separate the impact of wash trading from the reduction in sample size and increased abundance of null values without further research.

Looking at the distribution of each liquidity proxy, there are noticeable differences. Notably, when using the heterogeneous collection assumption, which contains more idiosyncratic information than the homogeneity assumption, Goyenko is the only variable that provides a reasonable distribution of values: Amihud and Roll both are 0.00 for 75% of the sample while Lesmond is 1,000,000 for 75% of the sample, whereas Goyenko has a more dispersed array of measures (see **Table 9**). It is reasonably expected that the homogeneous collection assumption will yield higher-quality regression estimates due to reduced clustering near the edges of the distribution.

Table 9
Liquidity proxy descriptive statistics by collection-day observation.

Variable	N	Mean	Std Dev	Min	25%	Median	75%	Max
Panel A: Homogeneous collections - including wash-trades								
Amihud	31,531	824.64	3,578.69	0.00	0.00	6.68	133.71	28,551.61
Lesmond	31,531	591,869.61	301,528.95	0.00	366,666.67	633,333.33	866,666.67	1,000,000.00
Goyenko	31,531	449,429.82	286,202.88	0.00	225,806.45	428,571.43	677,419.35	967,741.94
Roll	31,531	2,215,582.11	6,864,991.70	0.00	0.00	345,552.00	1,216,518.95	53,380,403.77
Panel B: Heterogeneous collections - including wash-trades								
Amihud	31,531	7.00	36.09	0.00	0.00	0.00	0.00	288.99
Lesmond	31,531	999,124.12	3,914.02	967,741.94	1,000,000.00	1,000,000.00	1,000,000.00	1,000,000.00
Goyenko	31,531	948,848.75	20,278.47	886,628.08	941,666.67	956,989.25	966,666.67	967,741.94
Roll	31,531	11,350.03	57,994.47	0.00	0.00	0.00	0.00	467,212.70
Panel C: Homogeneous collections - excluding wash-trades								
Amihud	31,051	825.63	3,554.23	0.00	0.00	7.11	137.05	28,183.59
Lesmond	31,051	588,485.53	301,148.91	0.00	354,838.71	612,903.23	866,666.67	1,000,000.00
Goyenko	31,051	450,024.73	285,820.69	0.00	225,806.45	428,571.43	677,419.35	967,741.94
Roll	31,051	2,239,816.70	6,915,626.42	0.00	0.00	353,553.39	1,237,051.74	53,593,134.48
Panel D: Heterogeneous collections - excluding wash-trades								
Amihud	31,051	6.56	33.79	0.00	0.00	0.00	0.00	270.41
Lesmond	31,051	999,529.47	1,641.88	989,247.31	1,000,000.00	1,000,000.00	1,000,000.00	1,000,000.00
Goyenko	31,051	954,733.48	15,247.75	900,000.00	948,888.89	958,870.97	967,741.94	967,741.94
Roll	31,051	10,502.52	54,799.72	0.00	0.00	0.00	0.00	445,540.19

This table shows descriptive statistics for collection-day level liquidity proxies using both the homogeneous (calculating proxies at the collection level assuming all NFTs are identical) and heterogeneous (calculating proxies at the NFT level and averaging within the collection-day observation) collection assumptions as well as before and after excluding wash trading. All proxy values are multiplied by 1,000,000 to improve readability.

6.5. Regression specification

Knowing that each of the four proxies captures slightly different dimensions of liquidity, performing regressions with varying fixed effects helps to decompose the drivers of liquidity. The first regression is a simple ordinary least square (OLS) regression using widely available and commonly used price controls. Subsequent regressions add collection fixed effects, date fixed effects and cluster standard errors around collections. Finally, all regressions are repeated with logged proxy values as a robustness test to uncover non-linear relationships.

The base regression model uses the daily collection closing price ($Close_{c,d}$), daily transaction volume in USD ($Volume_{c,d}$), daily price volatility ($Volatility_{c,d}$), collection age ($Age_{c,d}$), age squared ($Age^2_{c,d}$), and the day's HHI score (HHI_d). Unlike the existing literature, this paper focusses on variables that are attainable solely via sale price data, both to understand if price variables are significant in predicting liquidity proxies and to provide a benchmark for future studies. The daily close price is added as a control for the relative sale price of the collection – if positive and significant, it indicates that more expensive collections are associated with more illiquidity. Adding the daily transaction volume and volatility respectively accounts for market interest and uncertainty in a particular collection. The age variables are added to detect the presence of a non-linear relationship between the age of a collection and its liquidity, which is anticipated to be influenced by the majority of collections dying relatively soon after launch, while a minority experiences most of the returns. The day's HHI score is also added to determine the impact of market saturation on liquidity; intuitively, a higher HHI is expected to reduce illiquidity since there is less market attention and demand for other collections.

For collection c on day d, $Close_{c,d}$ is the last recorded NFT sale price within c, $Volume_{c,d}$ is the sum of daily transaction volume in USD ($Volume_{c,d} = \sum_{s}^{N} Price_{s,d}$), $Volatility_{c,d}$ is the percent difference between the day's highest sale price and the lowest sale price, $Age_{c,d}$ is the number of days since the first recorded sale of the collection, $Age^2_{c,d}$ is the square of $Age_{c,d}$, and HHI_d is calculated as the sum of the square of each collections market share by $Volume_{c,d}$ ($HHI_d = \sum_{c,d}^{N} \left(\frac{Volume_{c,d}}{\sum_{c,d}^{N} Volume_{c,d}} * 100\right)^2$).

The baseline OLS regression is performed as specified in Equation 5.

$$Liquidity\ Proxy_{c,d} = \alpha_{c,d} + \beta_1\left(Close_{c,d,}\right) + \beta_2\left(Volume_{c,d,}\right) + \beta_3\left(Volatility_{c,d}\right) + \beta_4\left(Age_{c,d,}\right) + \beta_5\left(Age^2_{c,d,}\right) + \beta_6(HHI_d) + \epsilon_{c,d} \tag{5}$$

Subsequently, collection fixed effects are added to isolate the impact of the price variables by controlling for collection-specific factors such as the creator/artist, collection history, past returns, and public sentiment for that collection. These are coupled with standard errors clustered around collections to address intra-collection correlations and account for potential heteroskedasticity. Date fixed effects are then added to account for long-term trends and market-wide liquidity events. Finally, all the above regressions are repeated using logged liquidity proxies, which requires the original proxies to be normalised by adding one to each of them.

7. Results

7.1. Basic correlations

By analysing the Pearson correlation coefficients between the liquidity proxies found in **Table 10**, it is possible to investigate how the different dimensions of liquidity relate to each other and to basic price variables. Roll's measure, approximating implicit transaction costs (specifically bid-ask spread), has the strongest correlations with the other proxies, having a correlation coefficient of 0.14 with Lesmond, 0.13 with Goyenko and 0.17 with Amihud. The relationship with Lesmond and Goyenko is expected due to all three measures capturing the different aspects of transaction cost-associated illiquidity; however, these coefficients are not as strong as the Lesmond-Goyenko relationship (0.92), which likely is caused by the similarities between Lesmond and Goyenko's methodology and the fact they both specifically estimate explicit transaction costs. The difference in magnitude between the Roll-Lesmond/Goyenko and the Lesmond-Goyenko correlation coefficients supports the expectation of the literature that Roll and Lesmond/Goyenko do not capture the same aspects of transaction cost liquidity. Roll's closer relationship to Amihud (0.17) is less intuitive since Amihud is primarily a price impact estimator, yet they have been found to be strongly related in the research of Acharya & Pedersen (2005). In the same paper, they posit that Amihud's illiquidity measure could have an alternative, yet more narrow, interpretation as a measure for implicit transaction costs, which would explain the 0.17 correlation between itself and Roll's measure due to the inherent relationship between price impact and the implied bid-ask spread. As such, this relationship can be interpreted as both Roll and Amihud capturing the bid-ask spread component of the implicit transaction costs, which is reinforced by Amihud's correlation of 0 with both Lesmond and Goyenko, indicating that Amihud's liquidity dimension is completely distinct from Lesmond/Goyenko.

The inter-proxy relationships identified above are mostly supported by the existing literature, except for the Roll-Lesmond and Roll-Goyenko correlations. When analysing the relationships between liquidity proxies and liquidity benchmarks, Goyenko et al. (2009) note that the monthly Amihud measure is not significantly different from Roll: their respective average cross-sectional correlation coefficients to an effective spread benchmark are 0.56 and 0.57. Increasing the observation frequency to daily, as done by Langedijk et al. (2018) when analysing bond liquidity, Amihud's correlation to these benchmarks falls between 0.1 and 0.2. Given the reduction in correlation when moving from monthly to daily frequency, the Amihud/Roll correlation of 0.17 appears to be in line with the magnitude proposed by analgous literature. Additional studies reaffirm that Amihud captures bid-ask spread related illiquidity in multiple settings such as equity markets (Fong et al., 2017), emerging markets (Będowska-Sójka, 2018) and government bonds (Su & Tokmakcioglu, 2021). Regarding the zeros measure, Goyenko et al. (2009) also note that the Lesmond and Goyenko zero measures are insignificant from each other when using pure-time series analysis (respectively achieving correlation coefficients of 0.96 and 0.95 to the realised spread benchmark), which aligns with my findings that Lesmond and Goyenko and strongly related. Where the results of **Table 10** differ from the literature is the relationship between Roll and the zero measures, which I find to be as strong as the Amihud/Roll relationship. Contrary to this, Goyenko et al. (2009) identify them to be significantly different at the monthly frequency, yet this deviation disappears at the annual level and has not been readily studied at lower frequencies.

Roll's measure has the strongest relationship to the observable variables out of any proxy: a coefficient of 0.17 with age, which is unexpected since Roll solely uses price returns and is thus time agnostic. The next strongest proxy-variable relationship is Lesmond with daily transaction volume at 0.11; this is expected from Lesmond's measure since many zero-return days are also zero-transaction volume days, however, Goyenko only holds a 0.07 relationship with the same variable. The remainder of proxy-variable relationships are mostly insignificant (0.07 or lower), which is a positive since it minimises the risk of multicollinearity and allows for more robust regressions.

There are no differences between homogeneous and heterogeneous collection assumption correlations to two decimal places, which suggests that assuming heterogeneous collections reveals no additional information regarding inter-proxy relationships as opposed to assuming homogeneous collections. This finding is tenuous, however, as the correlation matrix only analyses variables on a one-to-one basis and needs to be confirmed in subsequent analysis. Furthermore, the exclusion of wash-trades also has minimal effect on correlation coefficients, with the largest changes in correlations being 0.01, and suggests that all liquidity proxies, and their relationships to the price variables, are equally affected by the exclusion of wash-trades.

Table 10
Pearson correlation matrix.

Panel A: Pre-Wash-trade Exclusion

	Variables	1	2	3	4	5	6	7	8	9	10
1	Amihud	1	0.00	0.00	**0.17**	-0.02	-0.03	0.00	0.07	0.08	0.00
2	Lesmond	0.00	1	**0.92**	**-0.14**	0.04	**-0.11**	-0.02	0.02	0.03	0.00
3	Goyenko	0.00	**0.92**	1	**-0.13**	0.04	-0.07	-0.01	-0.01	0.03	-0.06
4	Roll	**0.17**	**-0.14**	**-0.13**	1	-0.02	-0.02	0.00	**0.17**	**0.18**	0.01
5	Close price	-0.02	0.04	0.04	-0.02	1	**0.18**	0.00	0.03	0.03	0.00
6	Day volume	-0.03	**-0.11**	-0.07	-0.02	**0.18**	1	0.00	-0.03	-0.02	0.02
7	Day volatility	0.00	-0.02	-0.01	0.00	0.00	0.00	1	0.00	0.00	0.01
8	Age	0.07	0.02	-0.01	**0.17**	0.03	-0.03	0.00	1	**0.96**	**-0.11**
9	Age^2	0.08	0.03	0.03	**0.18**	0.03	-0.02	0.00	**0.96**	1	**-0.13**
10	HHI	0.00	0.00	-0.06	0.01	0.00	0.02	0.01	**-0.11**	**-0.13**	1

Panel B: Post-Wash-trade Exclusion

	Variables	1	2	3	4	5	6	7	8	9	10
1	Amihud	1	0.00	0.00	**0.17**	-0.03	-0.03	0.00	0.07	0.07	0.00
2	Lesmond	0.00	1	**0.93**	**-0.13**	0.04	**-0.11**	-0.02	0.02	0.04	0.01
3	Goyenko	0.00	**0.93**	1	**-0.12**	0.04	-0.07	-0.01	-0.01	0.03	-0.06
4	Roll	**0.17**	**-0.13**	**-0.12**	1	-0.02	-0.02	0.00	**0.17**	**0.18**	0.01
5	Close price	-0.03	0.04	0.04	-0.02	1	**0.19**	0.00	0.03	0.03	0.00
6	Day volume	-0.03	**-0.11**	-0.07	-0.02	**0.19**	1	0.00	-0.03	-0.03	0.02
7	Day volatility	0.00	-0.02	-0.01	0.00	0.00	0.00	1	0.00	0.00	0.01
8	Age	0.07	0.02	-0.01	**0.17**	0.03	-0.03	0.00	1	**0.96**	**-0.10**
9	Age^2	0.07	0.04	0.03	**0.18**	0.03	-0.03	0.00	**0.96**	1	**-0.13**
10	HHI	0.00	0.01	-0.06	0.01	0.00	0.02	0.01	**-0.10**	**-0.13**	1

This table displays the Peason correlation matrix of the liquidity proxies and variables used in the analysis. The correlations are based on panel data observations and the values below and above the diagonal respectively use the homogeneous and heterogeneous collection assumption datasets. Coefficients with magnitudes greater than 0.1 are boldened. The data set contains 1,985 unique collections between 1 Feb 2021 and 21 Aug 2021.

7.2. Regression results

As noted above, correlation matrices only allow a one-to-one analysis of the relationship between liquidity proxies. Further analysis requires the use of OLS regressions to decompose the influencing factors for each liquidity proxy and effectively analyse what drives the various dimensions of liquidity in NFT collections.

The simple OLS regression results outlined in **Table 11** reaffirm the strong relationships identified in **Table 10**, while simultaneously adding complexity. The primary similarities between the correlation matrix and the simple OLS results are Amihud and Roll maintaining a close relationship with age/age squared and Lesmond/Goyenko maintaining their relationship with daily transaction volume. A significant change across all proxies, however, is the widespread significance of HHI despite its correlation coefficient with Amihud, Lesmond, Goyenko and Roll being respectively 0.00, 0.01, -0.06 and 0.01. This discrepancy between the correlation matrix and OLS regressions is likely due to the presence of multiple factors simultaneously contributing to collection liquidity that cannot be detected by the Pearson correlation calculation. A common feature across all proxies, irrespective of assuming collection homogeneity/heterogeneity or including wash trading, is that the estimated day volatility coefficient is 0 to two decimal places, even after scaling the proxies up by 1,000,000 and the day volatility variable down by 1,000. Due to the fact that it is statistically significant in all datasets for Lesmond, Goyenko and Roll, it can be inferred that day volatility only plays a significant role in estimating liquidity when it is extremely large in magnitude: as noted in **Table 8**, 75% of day volatility values are 584.21% or less, yet the average value is 2,547.57%, revealing a highly skewed distribution.

Comparing the regression estimates from the homogeneous and heterogeneous collection assumptions, there are notable differences in how each proxy reacts to the incorporation of NFT-specific information. Amihud has both its day volume and HHI coefficients change signs, and its close price coefficient becomes significantly closer to zero. These are coupled with a newfound significance in the linear age variable and a significant reduction in the estimated intercept (617.65 to 13.87). These changes likely occur due to the drastic change in the distribution of values demonstrated in **Table 9**; notably, the homogeneous assumption places the intercept below the mean value while the heterogeneous assumption has its intercept roughly two times larger than the mean value. Additionally, it is notable that for all the proxies except Goyenko, the heterogeneous collection assumption generates a higher R-squared value, implying a higher-performing model, which is consistent before and after excluding wash trading. This finding suggests that incorporating NFT-specific liquidity information is preferable when using price level data, which confirms expectations and follows economic intuition.

Analysing the differences between the pre- and post-wash trading exclusion results, it is clear that the impact of wash trading on liquidity arises at the NFT level. When assuming homogeneous collections, there is no change in the R-squared score after excluding wash-trades. When assuming heterogeneous collections, the exclusion of wash-trades increases the R-squared values for all proxies. This result is likely due to the increase in perceived liquidity of individual undeserving NFTs, which aligns with research by Oh (2023) and Serneels (2023) who both find that wash trading is clustered by collection, not within collections nor by asset characteristics. As such, NFTs randomly selected for wash trading, which would otherwise have very low liquidity proxy values, experience abnormal trade volumes and thus higher proxy values – this is clearly seen in the difference between the Goyenko intercept estimate, which is 419,741.98 prior to wash trading exclusion and 951,871.88 post-wash trading exclusion. Goyenko, which counts the number of zero transaction volume days, fully captures the exclusion of wash trading at an individual level: the lack of change in Amihud's measure, which also uses transaction volume as a direct input, is expected due to the majority of wash trading being 0 return trades, thus making that asset's Amihud measure 0 for the day. It is noteworthy that when assuming homogeneous collections, there is minimal change in the estimates before and after excluding wash trading; this result bolsters the finding that the heterogeneous assumption captures more information than the homogeneous assumption. In summary, the impact of wash trading is clearly an NFT-specific issue that is captured most effectively by the Goyenko measure.

Table 11
NFT collection liquidity using simple OLS regression.

Panel A: Including wash-trades								
	Homogeneous collections				Heterogeneous collections			
Parameter	Amihud	Lesmond	Goyenko	Roll	Amihud	Lesmond	Goyenko	Roll
Intercept	617.65***	612,386.14***	518,002.71***	1,147,765.15***	13.87***	998,783.41***	419,741.98***	6,919.81***
	(45.23)	(3,670.67)	(3,429.33)	(77,510.10)	(0.55)	(52.68)	(115,425.42)	(647.03)
Close price	-2.37***	539.97***	448.98***	-5,754.26***	-0.03***	3.62***	5,286.74***	-12.4
	(0.84)	(187.91)	(155.09)	(2,217.51)	(0.01)	(0.72)	(1,717.10)	(16.69)
Day volume	-0.25***	-96.42***	-65.89***	-98.12*	0.01***	-1.46***	-81.36	31.23***
	(0.04)	(16.88)	(10.75)	(58.70)	(0.01)	(0.19)	(189.19)	(6.00)
Day volatility	0.00	-0.00***	-0.00***	0.00	0.00	0.00***	0.00***	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age	-0.35	-1,304.02***	-2,521.27***	134.99	-0.17***	11.08***	-41,566.63***	173.19***
	(1.34)	(108.02)	(102.77)	(2,874.10)	(0.01)	(1.33)	(3,963.83)	(22.82)
Age^2	0.03***	8.19***	14.42***	121.13***	0.01***	-0.05***	201.24***	-0.61***
	(0.01)	(0.60)	(0.57)	(19.52)	(0.00)	(0.01)	(22.48)	(0.15)
HHI	0.03**	4.26***	-9.20***	160.20***	-0.02***	0.07***	-74.03*	-2.78***
	(0.01)	(1.30)	(1.25)	(28.05)	(0.00)	(0.02)	(44.02)	(0.18)
Collection Fixed Effects	No	No	No	No	No	No	No	No
Date Fixed Effects	No	No	No	No	No	No	No	No
Adjusted R^2	0.01	0.02	0.03	0.03	0.01	0.02	0.004	0.05
Observations	30,780	30,780	30,780	30,780	30,780	30,780	30,780	30,780

Table 11 continued

Panel B: Excluding wash-trades

	Homogeneous collections				Heterogeneous collections			
Parameter	Amihud	Lesmond	Goyenko	Roll	Amihud	Lesmond	Goyenko	Roll
Intercept	642.28***	607,308.52***	516,598.38***	1,176,593.33***	13.55***	999,354.79***	951,871.88***	6,077.08***
	(44.55)	(3,688.54)	(3,443.00)	(77,667.99)	(0.53)	(23.04)	(193.36)	(621.58)
Close price	-2.29***	574.06***	456.10***	-5,775.78***	-0.03***	2.51***	-1.43	-19.079
	(0.82)	(193.46)	(155.54)	(2,221.15)	(0.01)	(0.51)	(2.71)	(15.74)
Day volume	-0.27***	-109.32***	-71.49***	-101.01	0.01***	-1.27***	-0.94***	33.80***
	(0.05)	(17.26)	(11.58)	(66.03)	(0.00)	(0.18)	(0.20)	(5.89)
Day volatility	0.00	-0.00***	-0.00***	0.00	0.00	0.00***	0.00**	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age	-0.7	-1,303.42***	-2,512.18***	319.06	-0.17***	5.17***	54.28***	176.37***
	(1.34)	(108.73)	(103.41)	(2,907.76)	(0.01)	(0.63)	(5.69)	(21.62)
Age^2	0.03***	8.37***	14.37***	119.05***	0.01***	-0.03***	-0.14***	-0.66***
	(0.01)	(0.61)	(0.57)	(19.69)	(0.00)	(0.00)	(0.03)	(0.14)
HHI	0.02*	4.67***	-7.53***	146.69***	-0.01***	0.07***	0.39***	-2.62***
	(0.01)	(1.23)	(1.18)	(26.15)	(0.00)	(0.01)	(0.06)	(0.17)
Collection Fixed Effects	No	No	No	No	No	No	No	No
Date Fixed Effects	No	No	No	No	No	No	No	No
Adjusted R^2	0.01	0.02	0.03	0.03	0.02	0.08	0.01	0.06
Observations	30,350	30,350	30,350	30,350	30,350	30,350	30,350	30,350

This table displays the estimates of Equation (5), in which the dependent variable is one of the four liquidity proxies' measure for collection *i* on day *d*. Panel A uses a dataset including wash-trades, while Panel B uses a dataset excluding wash-trades. Within each of them, the homogeneous collection columns calculate the liquidity proxies at the collection level by assuming all NFTs within a collection are identical, whereas the heterogeneous collection columns calculate individual NFTs' liquidity proxies and averages them by collection-day. All proxy values are multiplied by 1,000,000 and close price, day volume and day volatility are divided by 1,000 to improve readability. Panel A's sample consists of 30,780 collection-day observations and Panel B's sample consists of 30,350 observations between 1 Feb 2021 and 21 Aug 2021. Standard errors are shown in parentheses. ***, ** and * represent statistical significance at the 1%, 5% and 10% levels respectively.

Table 12
NFT collection liquidity regressions using collection fixed effects.

Parameter	Amihud	Lesmond	Goyenko	Roll
Intercept	13.55***	999,354.79***	951,871.88***	6,077.08***
	(0.53)	(23.04)	(193.36)	(621.58)
Close price	-0.03***	2.51***	-1.43	-19.08
	(0.01)	(0.51)	(2.71)	(15.74)
Day volume	0.01***	-1.27***	-0.94***	33.80***
	(0.00)	(0.18)	(0.20)	(5.89)
Day volatility	0.00	0.00***	0.00**	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)
Age	-0.17***	5.17***	54.28***	176.37***
	(0.01)	(0.63)	(5.69)	(21.62)
Age^2	0.01***	-0.03***	-0.14***	-0.66***
	(0.00)	(0.00)	(0.03)	(0.14)
HHI	0.00***	0.07***	0.39***	-2.62***
	(0.00)	(0.01)	(0.06)	(0.17)
Collection Fixed Effects	Yes	Yes	Yes	Yes
Date Fixed Effects	No	No	No	No
Adjusted R^2	0.02	0.08	0.01	0.06
Observations	30,780	30,780	30,780	30,780

This table displays the estimates of Equation (5), including collection level fixed effects, in which the dependent variable is one of the four liquidity proxies' measure for collection *i* on day *d*. This table shows the estimates for the homogeneous assumption, heterogeneous assumption, prewash trading exclusion and post-wash trading exclusion datasets, which are identical. All proxy values are multiplied by 1,000,000 and close price, day volume and day volatility are divided by 1,000 to improve readability. The sample consists of 30,780 collection-day observations between 1 Feb 2021 and 21 Aug 2021. Robust standard errors are clustered to the collection level to account for within-collection cross-sectional correlations in regression residuals and are shown in parentheses. ***, ** and * represent statistical significance at the 1%, 5% and 10% levels respectively.

The results following the addition of collection fixed effects and collection-based clustered standard errors further support the notion that the heterogeneous collection assumption captures more information than the homogeneous collection assumption. Adding these regression features causes the model estimates for both homogeneous and heterogeneous collection assumptions, both before and after excluding wash trading, to become identical to two decimal places: the estimates are shown in **Table 12**. Notably, the estimates from this model are the same as the heterogeneous collection assumption estimates after excluding wash trading (see the heterogeneous collections section of panel B of **Table 11**). Statistically, this means that collection-specific features, irrespective of the proxy calculation methodology employed, absorb intra-collection idiosyncrasies regarding both NFT-specific features and wash trading effects.

An implication of this finding is that the primary benefit of assuming heterogeneous collections, the informational benefit of capturing NFT-specific liquidity attributes, is now fully accounted for using collection fixed effects. Although further research is required to determine exactly which aspects of collection fixed effects are able to capture the NFT level information (e.g., collection publicity, creator, available features, track record), this finding allows researchers analysing liquidity in NFT collections to assume that NFTs are homogeneous within collections with respect to liquidity. It is possible that this assumption is valid for models attempting to explain other aspects of NFT collection such as returns, volatility and inter-asset relationships; each time a new dimension of NFT collections is analysed, this assumption will need to be validated, as it may not hold outside of liquidity studies.

Another major finding from **Table 12** is that the addition of collection fixed effects removes all differences between the models generated using pre- and post-wash trading data. This evidence can be interpreted in two main ways: A) wash trading has minimal influence on liquidity or B) individual NFT wash trading is fully accounted for by collection fixed effects. Option A is unlikely, since Goyenko, which is more sensitive to wash trading than the other proxies, has the same intercept and coefficient estimates using both pre and post-wash trading exclusion datasets alongside fixed effects. This option is also discouraged by the results in **Table 11**, which show distinct differences between models generated using the two datasets. Option B is more probable: as noted in **Table 7**, only 2.39% of trades were removed due to wash trading, which affected 679 (0.09%) assets and 244 (11.99%) listed collections, suggesting a wide distribution of wash trading and an average of three wash-traded assets per affected collection. Since the collection fixed effects fully capture the effects of wash trading, it can be assumed that wash trading is targeted at the collection level and randomly targeted at the NFT level, which is supported by the existing literature (Oh, 2023).

The primary implications of this wash trading finding pertain to further studies in NFT liquidity. While there are multiple existing papers that analyse NFT wash trading, this is the first to consider its effect on liquidity, which Imisiker & Tas (2016) found to be significantly impacted in stock markets. With 11.99% of collections being affected by wash trading, wash trading has a noticeable effect on the OLS model estimates of liquidity and their accuracy. However, knowing that including collection fixed effects and clustering standard errors by collection causes these differences to disappear, future research into NFT liquidity at the collection level does not need to be concerned with NFT-specific wash trading. This can introduce significant computational and data processing savings, simplifying the research approach. Notably, since individual NFTs are no longer required to be tracked to identify wash trading, the initial data sourcing from marketplaces can be simplified to no longer include asset identifiers, seller addresses and buyer addresses, instead solely requiring a history of collection sale prices, which are widely available without custom API pulls from marketplaces, as required for this paper.

The addition of date fixed effects generates identical results to **Table 12** to two decimal places, which implies that collection fixed effects are able to effectively capture the variance generated by market-wide liquidity shocks within the sample period. This dynamic is also present in a real estate study by Chernobai & Hossain (2019), which finds that there are significantly higher rates of inter-ZIP code liquidity dispersion during bull markets, in which ZIP-code fixed effects have increased explanatory power. Investigating this possible similarity between the real estate and NFT markets, and serving as a robustness test, the dataset is divided into a bear period (24 March 2021 to 23 May 2021) and a bull period (23 May 2021 to 21 August 2021), which is visually apparent in **Figure 5**. The same family of regressions is performed on each of the two datasets, analysing both homogeneous and heterogeneous collections assumptions as well as pre and post-wash trading exclusion. As anticipated by the real estate literature, during the extreme bull period the sample, the exact same dynamic is present as previously mentioned: the addition of collection fixed effects unifies the homogeneous/heterogeneous assumptions and pre/post-wash trading exclusion estimates, and the addition of date fixed effects does not change the regressions. The results of the bull period OLS regressions can be seen in **Appendix B**, while the outputs with fixed effects can be seen in **Table 13**. This dynamic is missing, however, from the bear period of the sample: during this period, the addition of collection fixed effects does little to unify the regression estimates, and adding date fixed effects further disperses the model estimates (see **Appendix C** and **Appendix D**). This market-dependent relationship reveals that, with respect to inter-group liquidity dispersion, real estate ZIP codes and NFT collections experience the same dynamics.

As an additional robustness test, to uncover if non-linear relationships exist with collection liquidity, the proxies are normalised and logged. Applying the same family of regressions, the collection fixed effects results can be seen in **Table 14**. The simple OLS and date fixed effects results can be found in **Appendix E** and **Appendix F**. As expected, the logged models exhibit the same behaviour as the normal proxies: the simple OLS gives varying results, adding collection fixed effects unifies homogeneous/heterogeneous assumptions and pre/post-wash trading, and adding date fixed effects has no impact on the models.

Table 13
Bull market NFT collection liquidity regressions using collection fixed effects.

Parameter	Amihud	Lesmond	Goyenko	Roll
Intercept	1.10*	999,916.90***	935,226.48***	136.14
	(0.60)	(39.73)	(432.68)	(681.26)
Close price	-0.01	0.80	12.57	-15.88
	(0.00)	(0.55)	(13.27)	(14.57)
Day volume	0.00	-0.433***	0.11	-0.47
	(0.00)	(0.16)	(0.16)	(0.51)
Day volatility	2.19***	3.333	35.83	-49.62
	(0.64)	(6.26)	(34.64)	(58.36)
Age	-0.07	4.48	-16.20	90.58
	(0.06)	(5.37)	(40.97)	(76.64)
Age^2	0.00	-0.01	0.197	-0.59
	(0.00)	(0.05)	(0.32)	(0.75)
HHI	0.00	0.02*	0.13	-0.31
	(0.00)	(0.01)	(0.10)	(0.23)
Collection Fixed Effects	Yes	Yes	Yes	Yes
Date Fixed Effects	No	No	No	No
Adjusted R^2	0.02	0.2	0.17	0.06
Observations	5,369	5,369	5,369	5,369

This table displays the estimates of Equation (5), including collection level fixed effects, in which the dependent variable is one of the four liquidity proxies' measures for collection *i* on day *d*. This table shows the estimates for the homogeneous assumption, heterogeneous assumption, prewash trading exclusion and post-wash trading exclusion datasets, which are identical. All proxy values are multiplied by 1,000,000 and close price, day volume and day volatility are divided by 1,000 to improve readability. The sample consists of 5,369 collection-day observations between the bull market period of 23 May 2021 and 21 Aug 2021. Robust standard errors are clustered to the collection level to account for within-collection cross-sectional correlations in regression residuals and are shown in parentheses. ***, ** and * represent statistical significance at the 1%, 5% and 10% levels respectively.

Table 14
Logged NFT collection liquidity regressions using collection fixed effects

Parameter	Ln(Amihud)	Ln(Lesmond)	Ln(Goyenko)	Ln(Roll)
Intercept	13.55***	692,824.16***	668,760.28***	5,599.30***
	(0.53)	(11.54)	(99.59)	(531.19)
Close price	-0.03***	1.26***	-0.77	-17.45
	(0.01)	(0.26)	(1.39)	(13.06)
Day volume	0.01***	-0.63***	-0.47***	28.90***
	(0.00)	(0.09)	(0.10)	(5.02)
Day volatility	0.00	0.00***	0.00**	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)
Age	-0.17***	2.59***	27.74***	149.61***
	(0.01)	(0.31)	(2.93)	(18.25)
Age^2	0.01***	-0.01***	-0.07***	-0.56***
	(0.00)	(0.00)	(0.02)	(0.12)
HHI	-0.01***	0.04***	0.20***	-2.29***
	(0.00)	(0.00)	(0.03)	(0.14)
Collection Fixed Effects	Yes	Yes	Yes	Yes
Date Fixed Effects	No	No	No	No
Adjusted R^2	0.02	0.08	0.01	0.06
Observations	30,350	30,350	30,350	30,350

This table displays the estimates of Equation (5), including collection level fixed effects, in which the dependent variable is the natural log of one of the four liquidity proxies' measure for collection *i* on day *d*. This table shows the estimates for the homogeneous assumption, heterogeneous assumption, prewash trading exclusion and post-wash trading exclusion datasets, which are identical. All proxy values are normalised by adding 1 to each of them prior to taking the natural log. They are then multiplied by 1,000,000 and close price, day volume and day volatility are divided by 1,000 to improve readability. The sample consists of 30,530 collection-day observations between 1 Feb 2021 and 21 Aug 2021. Robust standard errors are clustered to the collection level to account for within-collection cross-sectional correlations in regression residuals and are shown in parentheses. ***, ** and * represent statistical significance at the 1%, 5% and 10% levels respectively.

7.3. Economic interpretation

The pure transaction cost estimators, Lesmond and Goyenko, generate significant findings for the relationships between transaction cost associated illiquidity and the tested variables. Both proxies attempt to measure when expected returns do not exceed transaction costs and observe a positive age coefficient, negative age square coefficient and positive HHI coefficient. The positive linear age component and negative age squared component reveal that transaction cost-induced illiquidity rises (reduction in liquidity) in the early stages of a collection's lifetime before reversing (potentially due to the top 8% of collections surviving extended periods of time and experiencing abnormally high returns that always exceed transaction costs). This age dynamic is also present in Roll's estimate, which indicates that the implicit transaction cost-induced illiquidity also rises before falling and that total transaction costs are very likely to follow this pattern. Furthermore, the positive HHI coefficient indicates that illiquidity and transaction costs rise as the market approaches monopoly; this is likely due to concentrated markets being a symptom of reduced buyer demand across the market. As a result of this reduced demand, there are lower prospective returns for sellers, who then either choose not to sell, thus further increasing the HHI score and increasing both Lesmond and Goyenko, or selling for a near 0% return, which only increases Lesmond: this feedback loop helps explain the difference between the magnitude of their HHI coefficients. Their negative daily transaction volume coefficients are expected since they are the criteria by which both measures count liquidity: Lesmond increases if there are fewer non-zero trading days and Goyenko increases if there are fewer actively traded days. Regarding the other variables, the differences in their overall estimate magnitudes are likely caused by the reduced explanatory power of the Goyenko model, which attempts to explain day-to-day changes between the variables more dynamically. Regarding the implications for alternative asset markets, such as in art and real estate, market facilitators can use these findings to answer questions regarding fixed fee transaction costs (i.e., transaction costs that do not change depending on the value of the asset). This type of transaction fees is dominant in the NFT market, as the primary cost for generic NFT transactions is the associated gas fee, which changes day to day based on electricity costs, computational supply and network demand, but does not change depending on the size/value of the NFT (Laurent et al., 2022). Knowing that the primary component of the explicit transaction costs, the gas fee, is relatively fixed irrespective of asset value, it can be inferred that the age-liquidity relationship is predominantly driven by the implicit cost. As such, non-fungible asset market facilitators can infer their assets age, it is primarily implicit costs driving either rises or reductions in liquidity.

Considering the cross-over of the Amihud and Roll measure with regard to capturing the tightness dimension of liquidity (bid-ask spread) and their individual nuances, these two proxies generate compelling findings regarding the time frame of liquidity within the NFT market's structure. Analysing the age coefficient estimates, Amihud's measure suggests that illiquidity dips (rise in liquidity) very early in collection life (less than a month in this sample), but soon reverses. Meanwhile, Roll's model suggests that illiquidity rises (reduction in liquidity) for a significant amount of time and reverses much later (approximately four months in this sample period). Even after considering robust standard errors, the Amihud-age inflection point is significantly earlier, as seen by Roll's significantly larger linear age variable relative to the squared component. In the context of a non-fungible asset market, characterised by the auction style of market making, the disparity in estimates between the Amihud and Roll models is primarily attributed to the fact that they emphasise different timeframes of liquidity. Amihud's proxy centres on the price impact of transaction volume, capturing short-term illiquidity influenced by immediate trading activity; the total impact of these short-term effects seems to increase as the collection ages. In this market context, it can reflect the heightened sensitivity to auction dynamics and current bid-ask spreads. In contrast, Roll's measure focusses on the implied effective bid-ask spread through return autocorrelation and captures long-term structural liquidity factors, such as the consistency of auction results and bid-ask spreads over time; these long-term illiquidity factors seem to fall in prevalence once a collection reaches a certain age. A possible implication of this finding is that Amihud's measure, by virtue of focussing on the short-term, is able to capture liquidity dynamics for both successful and unsuccessful collections, whereas Roll mainly captures the long-term liquidity effects present in successful collections that have significant market presence and momentum; this dynamic would explain why Amihud predicts a short-term reduction in liquidity since short-term failing projects will experience reduced demand and thus reduced liquidity.

8. Conclusion

This paper conducts the first comprehensive examination of collection-level liquidity in the rapidly growing NFT market. I calculate a suite of widely used liquidity proxies (Amihud, Lesmond, Goyenko and Roll's measures) under assumptions of both collection homogeneity and heterogeneity in both pre and post-wash trading datasets. I then analyse these proxies using Pearson correlation matrices, OLS regressions and varying fixed effects.

Results identify Roll's measure (approximating implicit bid-ask spread) as the preferred liquidity measure due to its wide dispersion of values, coverage of both tightness and depth dimensions and exposure to long-term liquidity market factors. I find that adding collection fixed effects eliminates all differences between models estimated using the homogeneous/heterogeneous assumption and pre/post-wash trading exclusion datasets; these estimates are the same as the heterogeneous post-wash-trading-exclusion results, indicating that collection-specific characteristics effectively capture NFT-specific information. Robustness tests reveal that this explanatory power is only present in bull markets, behaving similarly to real estate ZIP-code groups. Finally, I find that collections observe a non-linear liquidity lifecycle, with successful NFTs observing a fall in liquidity before a sharp uptake while others fail to resell.

Due to its youth, there are multiple avenues for future research within NFT liquidity literature. A liquidity horserace of multiple NFT liquidity proxies against high-frequency benchmark, mimicking Goyenko et al. (2009), would provide significant insights as to the empirical efficacy of proxies that underpin the all liquidity research. Secondly, an investigation into the parallels between real estate and NFT markets regarding collection dynamics would offer promising findings for both asset classes. Thirdly, the presence of cyclical liquidity dynamics suggests that this exists in many other aspects of NFT markets such as asset pricing, market microstructures and trading behaviour.

9. References

Abdi, F., & Ranaldo, A. (2016). A Simple Estimation of Bid-Ask Spreads from Daily Close, High, and Low Prices. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2725981

Acharya, V., & Pedersen, L. (2005). Asset pricing with liquidity risk. *Journal of Financial Economics*, *77*(2), 375–410. https://doi.org/10.1016/j.jfineco.2004.06.007

Aidov, A., & Lobanova, O. (2021). The Relation between Intraday Limit Order Book Depth and Spread. *International Journal of Financial Studies*, *9*(4), 60. https://doi.org/10.3390/ijfs9040060

Ametefe, F., Devaney, S., & Marcato, G. (2015). *Estimating Liquidity in Real Estate Markets—Literature Review* (IPF Research Programme 2011-2015). Investment Property Forum. https://www.ipf.org.uk/static/uploaded/f79ab58d-a84e-4449-83cf842d8da4d020.pdf

Ametefe, F., Devaney, S., & Marcato, G. (2016). Liquidity: A Review of Dimensions, Causes, Measures, and Empirical Applications in Real Estate Markets. *Journal of Real Estate Literature*, *24*(1), 3–30.

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, *5*(1), 31–56. https://doi.org/10.1016/S1386-4181(01)00024-6

Anselmi, G., & Petrella, G. (2023). Non-fungible token artworks: More crypto than art? *Finance Research Letters*, *51*, 103473. https://doi.org/10.1016/j.frl.2022.103473

Apostu, S. A., Panait, M., Vasa, L., Mihaescu, C., & Dobrowolski, Z. (2022). NFTs and Cryptocurrencies—The Metamorphosis of the Economy under the Sign of Blockchain: A Time Series Approach. *Mathematics*, *10*(17), 3218. https://doi.org/10.3390/math10173218

Arcenegui, J., Arjona, R., Román, R., & Baturone, I. (2021). Secure Combination of IoT and Blockchain by Physically Binding IoT Devices to Smart Non-Fungible Tokens Using PUFs. *Sensors*, *21*(9), 3119. https://doi.org/10.3390/s21093119

ASX. (2023, July 10). *ASX Fees and subscriptions*. https://www.asx.com.au/connectivity-and-data/information-services/benchmarks/benchmark-data/fees-and-subscriptions

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

Bank for International Settlements. (2023, May 17). Statistical Release—OTC derivatives at end-December 2022. *OTC Derivatives Statistics at End-December 2022*. https://stats.bis.org/statx/srs/tseries/OTC_DERIV/H.D.A.A.5J.A.5J.A.TO1.TO1.A.A.3.C?t=d5.1&c=&p=20222&i=1.8

Barardehi, Y. H., Bernhardt, D., Ruchti, T. G., & Weidenmier, M. (2021). The Night and Day of Amihud's (2002) Liquidity Measure. *The Review of Asset Pricing Studies*, *11*(2), 269–308. https://doi.org/10.1093/rapstu/raaa022

Będowska-Sójka, B. (2018). The coherence of liquidity measures. The evidence from the emerging market. *Finance Research Letters*, *27*, 118–123. https://doi.org/10.1016/j.frl.2018.02.014

Będowska-Sójka, B., & Echaust, K. (2020). What is the best proxy for liquidity in the presence of extreme illiquidity? *Emerging Markets Review*, *43*, 100695. https://doi.org/10.1016/j.ememar.2020.100695

Bekaert, G., Harvey, C. R., & Lundblad, C. (2007). Liquidity and Expected Returns: Lessons from Emerging Markets. *Review of Financial Studies*, *20*(6), 1783–1831. https://doi.org/10.1093/rfs/hhm030

Bervas, A. (2006). Market liquidity and its incorporation into risk management. *Financial Stability Review*, *8*(5), 63–79.

Blackwell, C. (2018). Power Laws in Real Estate Prices? Some Evidence. *The Quarterly Review of Economics and Finance*, *69*, 90–98. https://doi.org/10.1016/j.qref.2018.03.014

Blockdata. (2022). The Top 100 Public Companies Using Blockchain in 2022. *Blockdata Research*. https://www.blockdata.tech/blog/general/the-top-100-public-companies-using-blockchain-in-2022

Borri, N., Liu, Y., & Tsyvinski, A. (2022). The Economics of Non-Fungible Tokens. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4052045

Cajias, M., Freudenreich, P., & Freudenreich, A. (2020). Exploring the determinants of real estate liquidity from an alternative perspective: Censored quantile regression in real estate research. *Journal of Business Economics*, *90*(7), 1057–1086. https://doi.org/10.1007/s11573-020-00988-w

Chalmers, D., Fisch, C., Matthews, R., Quinn, W., & Recker, J. (2022). Beyond the bubble: Will NFTs and digital proof of ownership empower creative industry entrepreneurs? *Journal of Business Venturing Insights*, *17*, e00309. https://doi.org/10.1016/j.jbvi.2022.e00309

Chernobai, E., & Hossain, T. (2019). Liquidity Imbalance in the Residential Real Estate Market. *Journal of Housing Research*, *28*(2), 180–207. https://doi.org/10.1080/10527001.2020.1776513

Chohan, R., & Paschen, J. (2023). NFT marketing: How marketers can use nonfungible tokens in their campaigns. *Business Horizons*, *66*(1), 43–50. https://doi.org/10.1016/j.bushor.2021.12.004

Colicev, A. (2023). How can non-fungible tokens bring value to brands. *International Journal of Research in Marketing*, *40*(1), 30–37. https://doi.org/10.1016/j.ijresmar.2022.07.003

Corwin, S., & Schultz, P. (2012). A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices. *The Journal of Finance*, *67*(2), 719–759.

Davis, R., Griffith, T., Van Ness, B., & Van Ness, R. (2023). Modern OTC market structure and liquidity: The tale of three tiers. *Journal of Financial Markets*, *64*, 100815. https://doi.org/10.1016/j.finmar.2023.100815

Deuskar, P., Gupta, A., & Subrahmanyam, M. G. (2011). Liquidity effect in OTC options markets: Premium or discount? *Journal of Financial Markets*, *14*(1), 127–160. https://doi.org/10.1016/j.finmar.2010.08.003

Díaz, A., & Escribano, A. (2020). Measuring the multi-faceted dimension of liquidity in financial markets: A literature review. *Research in International Business and Finance*, *51*, 101079. https://doi.org/10.1016/j.ribaf.2019.101079

Dos Santos, R. B., Torrisi, N. M., & Pantoni, R. P. (2021). Third Party Certification of Agri-Food Supply Chain Using Smart Contracts and Blockchain Tokens. *Sensors*, *21*(16), 5307. https://doi.org/10.3390/s21165307

Dowling, M. (2022). Is non-fungible token pricing driven by cryptocurrencies? *Finance Research Letters*, *44*, 102097. https://doi.org/10.1016/j.frl.2021.102097

Etro, F., & Stepanova, E. (2018). Power-laws in art. *Physica A: Statistical Mechanics and Its Applications*, *506*, 217–220. https://doi.org/10.1016/j.physa.2018.04.057

Far, S. B., Bamakan, S. M. H., Qu, Q., & Jiang, Q. (2022). A Review of Non-fungible Tokens Applications in the Real-world and Metaverse. *Procedia Computer Science*, *214*, 755–762. https://doi.org/10.1016/j.procs.2022.11.238

Florackis, C., Gregoriou, A., & Kostakis, A. (2011). Trading frequency and asset pricing on the London Stock Exchange: Evidence from a new price impact ratio. *Journal of Banking & Finance*, *35*(12), 3335–3350. https://doi.org/10.1016/j.jbankfin.2011.05.014

Fong, K. Y. L., Holden, C. W., & Trzcinka, C. A. (2017). What Are the Best Liquidity Proxies for Global Research?*. *Review of Finance*, *21*(4), 1355–1401. https://doi.org/10.1093/rof/rfx003

Goyenko, R. Y., Holden, C. W., & Trzcinka, C. A. (2009). Do liquidity measures measure liquidity?☆. *Journal of Financial Economics*, *92*(2), 153–181. https://doi.org/10.1016/j.jfineco.2008.06.002

Gustar, A. (2020). Fame, Obscurity and Power Laws in Music History. *Empirical Musicology Review*, *14*(3–4), 186. https://doi.org/10.18061/emr.v14i3-4.7003

Hamledari, H., & Fischer, M. (2021). Construction payment automation using blockchain-enabled smart contracts and robotic reality capture technologies. *Automation in Construction*, *132*, 103926. https://doi.org/10.1016/j.autcon.2021.103926

Hasan, M., & Starly, B. (2020). Decentralized cloud manufacturing-as-a-service (CMaaS) platform architecture with configurable digital assets. *Journal of Manufacturing Systems*, *56*, 157–174. https://doi.org/10.1016/j.jmsy.2020.05.017

Hibbert, J., Kirchner, A., Kretzschmar, G., & McNeil, A. (2009). *Liquidity Premium: Literature review of theoretical and empirical evidence* (1.0; Barrie & Hibbert Research Report).

Hicks, C., & Curry, B. (2023, August 30). *Understanding Crypto Wallets*. https://www.forbes.com/advisor/investing/cryptocurrency/crypto-wallets/

Holden, C. W. (2009). New Low-Frequency Spread Measures. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1410758

Imisiker, S., & Tas, B. K. O. (2016). *Wash Trades as a Stock Market Manipulation Tool* (SSRN Scholarly Paper 2476874). https://doi.org/10.2139/ssrn.2476874

Irwin, N. B., & Livy, M. R. (2022). Price and Liquidity Dynamics for Single and Multi-Family Homes during Housing Market Shocks. *The Journal of Real Estate Finance and Economics*, *65*(1), 22–47. https://doi.org/10.1007/s11146-021-09846-0

Jankowitsch, R., Nashikkar, A., & Subrahmanyam, M. G. (2011). Price dispersion in OTC markets: A new measure of liquidity. *Journal of Banking & Finance*, *35*(2), 343–357. https://doi.org/10.1016/j.jbankfin.2010.08.016

Kampakis, S. (2022). Non-fungible Tokens as an Alternative Investment – Evidence from CryptoPunks. *The Journal of The British Blockchain Association*, *5*(1), 1–12. https://doi.org/10.31585/jbba-5-1-(2)2022

Karim, S., Lucey, B. M., Naeem, M. A., & Uddin, G. S. (2022). Examining the interrelatedness of NFTs, DeFi tokens and cryptocurrencies. *Finance Research Letters*, *47*, 102696. https://doi.org/10.1016/j.frl.2022.102696

Karnaukh, N., Ranaldo, A., & Söderlind, P. (2015). Understanding FX Liquidity. *Review of Financial Studies*, *28*(11), 3073–3108. https://doi.org/10.1093/rfs/hhv029

Kotova, N., & Zhang, L. (2021). *Liquidity in Residential Real Estate Markets*. American Economic Association. https://www.aeaweb.org/conference/2021/preliminary/paper/88hHHsbA

Krainer, J. (2001). A Theory of Liquidity in Residential Real Estate Markets. *Journal of Urban Economics*, *49*(1), 32–53. https://doi.org/10.1006/juec.2000.2180

Langedijk, S., Monokroussos, G., & Papanagiotou, E. (2018). Benchmarking liquidity proxies: The case of EU sovereign bonds. *International Review of Economics & Finance*, *56*, 321–329. https://doi.org/10.1016/j.iref.2017.11.002

Laurent, A., Brotcorne, L., & Fortz, B. (2022). Transaction fees optimization in the Ethereum blockchain. *Blockchain: Research and Applications*, *3*(3), 100074. https://doi.org/10.1016/j.bcra.2022.100074

Le, H., & Gregoriou, A. (2020). HOW DO YOU CAPTURE LIQUIDITY? A REVIEW OF THE LITERATURE ON LOW-FREQUENCY STOCK LIQUIDITY. *Journal of Economic Surveys*, *34*(5), 1170–1186. https://doi.org/10.1111/joes.12385

Lee, J., & Kwon, T. (2021). Distributed Watchdogs Based on Blockchain for Securing Industrial Internet of Things. *Sensors*, *21*(13), 4393. https://doi.org/10.3390/s21134393

Lee, K.-H. (2011). The world price of liquidity risk. *Journal of Financial Economics*, *99*(1), 136–161. https://doi.org/10.1016/j.jfineco.2010.08.003

Lesmond, D. (2005). Liquidity of emerging markets. *Journal of Financial Economics*, *77*(2), 411–452. https://doi.org/10.1016/j.jfineco.2004.01.005

Lesmond, D. A., Ogden, J. P., & Trzcinka, C. A. (1999). A New Estimate of Transaction Costs. *Review of Financial Studies*, *12*(5), 1113–1141. https://doi.org/10.1093/rfs/12.5.1113

Liu, W.-H. (2019). An empirical re-examination of extreme tail behavior: Testing the assumptions of the power laws and the generalized Pareto distribution on the financial series. *Applied Economics*, *51*(30), 3310–3324. https://doi.org/10.1080/00036846.2019.1574968

Lommers, K., Kim, J., & Baioumy, M. (2022). Market Making in NFTs. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4226987

Mancini, L., Ranaldo, A., & Wrampelmeyer, J. (2012). Liquidity in the Foreign Exchange Market: Measurement, Commonality, and Risk Premiums. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1447869

Manzoor, A., Samarin, M., Mason, D., & Ylianttila, M. (2020). Scavenger Hunt: Utilization of Blockchain and IoT for a Location-Based Game. *IEEE Access*, *8*, 204863–204879. https://doi.org/10.1109/ACCESS.2020.3037182

Marshall, B., Nuttawat, V., & Nguyen, N. (2012). Commodity Liquidity Measurement and Transaction Costs. *The Review of Financial Studies*, *25*(2), 599–638.

Marshall, B. R., Nguyen, N. H., & Visaltanachoti, N. (2018). Do liquidity proxies measure liquidity accurately in ETFs? *Journal of International Financial Markets, Institutions and Money*, *55*, 94–111. https://doi.org/10.1016/j.intfin.2018.02.011

Mofokeng, N. E. M., & Matima, T. K. (2018). Future tourism trends: Utilizing non-fungible tokens to aid wildlife conservation. *African Journal of Hospitality, Tourism and Leisure*, *7*(4). https://www.ajhtl.com/uploads/7/1/6/3/7163688/article_21_vol_7_4__2018.pdf

Muroki, D. (2023). Discover the Most Expensive NFT in the World. *History Computer*. https://history-computer.com/discover-the-most-expensive-nft-in-the-world/#:~:text=Merge%20is%20an%20NFT%20collection,expensive%20NFT%20in%20the%20world.

Nadini, M., Alessandretti, L., Di Giacinto, F., Martino, M., Aiello, L. M., & Baronchelli, A. (2021). Mapping the NFT revolution: Market trends, trade networks, and visual features. *Scientific Reports*, *11*(1), 20902. https://doi.org/10.1038/s41598-021-00053-8

Nobanee, H., & Ellili, N. O. D. (2023). Non-fungible tokens (NFTs): A bibliometric and systematic review, current streams, developments, and directions for future research. *International Review of Economics & Finance*, *84*, 460–473. https://doi.org/10.1016/j.iref.2022.11.014

NYSE. (2023). *New York Stock Exchange Price List 2023*. https://www.nyse.com/publicdocs/nyse/markets/nyse/NYSE_Price_List.pdf

Oh, S. (2023). *Market Manipulation in NFT Markets* (SSRN Scholarly Paper 4397409). https://doi.org/10.2139/ssrn.4397409

O’Hara, M. (2003). Presidential Address: Liquidity and Price Discovery. *The Journal of Finance*, *58*(4), 1335–1354.

Pástor, Ľ., & Stambaugh, R. F. (2003). Liquidity Risk and Expected Stock Returns. *Journal of Political Economy*, *111*(3), 642–685. https://doi.org/10.1086/374184

Precedence Research. (2023). *Real Estate Market (By Property: Residential, Commercial, Industrial, Land, Others; By Type: Sales, Rental, Lease)—Global Industry Analysis, Size, Share, Growth, Trends, Regional Outlook, and Forecast 2023-2032* (Global Industry Analysis 3236; ICT). Precedence Research. https://www.precedenceresearch.com/real-estate-market

Rabaa’i, A., Zhu, X., & Jayaraman, J. D. (2022). *Understanding Non-Fungible Tokens (NFTs): Overview, Opportunities, and Challenges*. The 49th Annual Northeast Business & Economics Association Conference, Portsmouth, New Hampshire, USA.

Ramos, H. P., & Righi, M. B. (2020). Liquidity, implied volatility and tail risk: A comparison of liquidity measures. *International Review of Financial Analysis*, *69*, 101463. https://doi.org/10.1016/j.irfa.2020.101463

Renneboog, L., & Spaenjers, C. (2013). Buying beauty: On prices and returns in the art market. *Management Science*, *59*(1), 36–54.

Roll, R. (1984). A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market. *The Journal of Finance*, *39*(4), 1127–1139. https://doi.org/10.1111/j.1540-6261.1984.tb03897.x

Schestag, R., Schuster, P., & Uhrig-Homburg, M. (2016). Measuring Liquidity in Bond Markets. *Review of Financial Studies*, *29*(5), 1170–1219. https://doi.org/10.1093/rfs/hhv132

Serneels, S. (2023). Detecting wash trading for nonfungible tokens. *Finance Research Letters*, *52*, 103374. https://doi.org/10.1016/j.frl.2022.103374

Sifat, I., Tariq, S. A., & Van Donselaar, D. (2023). *Suspicious Trading in Nonfungible Tokens (NFTs)*. https://doi.org/10.13140/RG.2.2.17058.91843

Su, E., & Tokmakcioglu, K. (2021). A comparison of bid-ask spread proxies and determinants of bond bid-ask spread. *Borsa Istanbul Review*, *21*(3), 227–238. https://doi.org/10.1016/j.bir.2020.10.005

Taylor, C. R. (2023). Non-fungible tokens (NFTs) as promotional devices: Research needs and future projections. *International Journal of Advertising*, *42*(5), 799–800. https://doi.org/10.1080/02650487.2023.2211856

Umar, Z., Abrar, A., Zaremba, A., Teplova, T., & Vo, X. V. (2022). The Return and Volatility Connectedness of NFT Segments and Media Coverage: Fresh Evidence Based on News About the COVID-19 Pandemic. *Finance Research Letters*, *49*, 103031. https://doi.org/10.1016/j.frl.2022.103031

Urom, C., Ndubuisi, G., & Guesmi, K. (2022). Dynamic dependence and predictability between volume and return of Non-Fungible Tokens (NFTs): The roles of market factors and geopolitical risks. *Finance Research Letters*, *50*, 103188. https://doi.org/10.1016/j.frl.2022.103188

U.S. Department of Justice, A. D. (2018, July 31). *Herfindahl-Hirschman Index*. https://www.justice.gov/atr/herfindahl-hirschman-index#:~:text=The%20term%20%E2%80%9CHHI%E2%80%9D%20means%20the,then%20summing%20the%20resulting%20numbers

Van Haaften-Schick, L., & Whitaker, A. (2022). From the Artist's Contract to the blockchain ledger: New forms of artists' funding using equity and resale royalties. *Journal of Cultural Economics*, *46*(2), 287–315. https://doi.org/10.1007/s10824-022-09445-8 von Wachter, V., Jensen, J. R., Regner, F., & Ross, O. (2021). NFT Wash Trading: Quantifying Suspicious Behaviour in NFT markets. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4037143

Wang, Y., Horky, F., Baals, L. J., Lucey, B. M., & Vigne, S. A. (2022). Bubbles all the way down? Detecting and date-stamping bubble behaviours in NFT and DeFi markets. *Journal of Chinese Economic and Business Studies*, *20*(4), 415–436. https://doi.org/10.1080/14765284.2022.2138161

White, B., Mahanti, A., & Passi, K. (2022a). Characterizing the OpenSea NFT Marketplace. *Companion Proceedings of the Web Conference 2022*, 488–496. https://doi.org/10.1145/3487553.3524629

White, J. T., Wilkoff, S., & Yildiz, S. (2022b). The Role of the Media in Speculative Markets: Evidence from Non-Fungible Tokens (NFTs). *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4074154

Wilkoff, S., & Yildiz, S. (2023). The behavior and determinants of illiquidity in the non-fungible tokens (NFTs) market. *Global Finance Journal*, *55*, 100782. https://doi.org/10.1016/j.gfj.2022.100782

Wilson, K. B., Karg, A., & Ghaderi, H. (2022). Prospecting non-fungible tokens in the digital economy: Stakeholders and ecosystem, risk and opportunity. *Business Horizons*, *65*(5), 657–670. https://doi.org/10.1016/j.bushor.2021.10.007

Xia, Y., Li, J., & Fu, Y. (2022). Are non-fungible tokens (NFTs) different asset classes? Evidence from quantile connectedness approach. *Finance Research Letters*, *49*, 103156. https://doi.org/10.1016/j.frl.2022.103156

Yousaf, I., & Yarovaya, L. (2022a). Static and dynamic connectedness between NFTs, Defi and other assets: Portfolio implication. *Global Finance Journal*, *53*, 100719. https://doi.org/10.1016/j.gfj.2022.100719

Yousaf, I., & Yarovaya, L. (2022b). The relationship between trading volume, volatility and returns of Non-Fungible Tokens: Evidence from a quantile approach. *Finance Research Letters*, *50*, 103175. https://doi.org/10.1016/j.frl.2022.103175

Zhang, Z., Sun, Q., & Ma, Y. (2022). The hedge and safe haven properties of non-fungible tokens (NFTs): Evidence from the nonlinear autoregressive distributed lag (NARDL) model. *Finance Research Letters*, *50*, 103315. https://doi.org/10.1016/j.frl.2022.103315

10. Appendix

Appendix A
Augmented Dickey-Fuller unit root tests

		Including wash-trades				Excluding wash-trades			
		Homogeneous collection		Heterogeneous collection		Homogeneous collection		Heterogeneous collection	
Proxy	ADF Test	Rho	Pr < Rho	Rho	Pr < Rho	Rho	Pr < Rho	Rho	Pr < Rho
Amihud	Zero Mean	-30,677.00	<.0001	-30,071.00	<.0001	-30,457.00	<.0001	-30,843.00	<.0001
	Single Mean	-30,735.00	<.0001	-30,072.00	<.0001	-30,514.00	<.0001	-30,851.00	<.0001
	Trend	-30,735.00	<.0001	-30,073.00	<.0001	-30,514.00	<.0001	-30,851.00	<.0001
Lesmond	Zero Mean	-236.84	<.0001	-894.84	<.0001	-236.78	<.0001	**-0.26**	**0.6246**
	Single Mean	-1,147.30	<.0001	-1,524.10	<.0001	-1,138.90	<.0001	-18,071.00	<.0001
	Trend	-1,147.30	<.0001	-1,529.10	<.0001	-1,139.00	<.0001	-18,073.00	<.0001
Goyenko	Zero Mean	-443.48	<.0001	-8,961.40	<.0001	-434.83	<.0001	**-8.32**	**0.0461**
	Single Mean	-1,534.70	<.0001	-9,026.70	<.0001	-1,510.50	<.0001	-24,440.00	<.0001
	Trend	-1,534.70	<.0001	-9,108.00	<.0001	-1,510.50	<.0001	-24,441.00	<.0001
Roll	Zero Mean	-1,360.70	<.0001	-26,524.00	<.0001	-1,348.90	<.0001	-27,055.00	<.0001
	Single Mean	-1,391.90	<.0001	-26,529.00	<.0001	-1,380.10	<.0001	-27,127.00	<.0001
	Trend	-1,392.70	<.0001	-26,531.00	<.0001	-1,380.80	<.0001	-27,130.00	<.0001

This table shows the results of the augmented Dickey-Fuller unit root tests for the Amihud, Lesmond, Goyenko and Roll liquidity proxies. Each of them is tested against the zero mean, single mean and trend tests to prove stationarity. Each proxy is tested in under both pre and post-wash trading exclusion and under both the homogeneous collection assumption, in which liquidity proxies are calculated at the collection level by assuming all NFTs within a collection are identical, and the heterogeneous collection columns, in which individual NFTs' liquidity proxies are calculated and then averaged within each collection-day observation. Results that are not significant to the 1% level are boldened.

Appendix B
Bull market NFT collection liquidity OLS regressions.

Panel A: Including wash-trades								
	Homogenous collections				Heterogenous collections			
Parameter	Amihud	Lesmond	Goyenko	Roll	Amihud	Lesmond	Goyenko	Roll
Intercept	706.18***	573,015.53***	502,139.18***	1,235,536.08***	18.20***	998,679.63***	888,007.99***	8,995.76***
	(55.91)	(4,140.38)	(3,869.81)	(9,3162.49)	(0.72)	(61.09)	(101,970.70)	(793.43)
Close price	-2.20	315.09	248.81	-6,997.64	-0.03**	4.03***	3,782.41	-29.08
	(1.40)	(203.19)	(166.25)	(4,469.12)	(0.01)	(1.21)	(2,341.93)	(28.94)
Day volume	-0.33***	-93.45***	-57.67***	22.49	0.00***	-1.57***	264.34***	51.20***
	(0.05)	(14.95)	(10.09)	(108.95)	(0.00)	(0.23)	(50.61)	(6.34)
Day volatility	0.00***	0.00***	-0.00***	-0.00***	0.00	0.00***	0.00***	-0.00*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age	-4.42***	-513.77***	-1,908.53***	-20,405.23***	-0.23***	12.23***	-42,259.76***	227.50***
	(1.55)	(126.25)	(119.62)	(3,285.10)	(0.01)	(1.66)	(4,105.56)	(29.82)
Age^2	0.04***	5.03***	11.51***	235.55***	0.01***	-0.04***	186.21***	-0.99***
	(0.00)	(0.68)	(0.64)	(21.78)	(0.00)	(0.01)	(22.94)	(0.18)
HHI	0.10***	-0.23	-17.05***	366.93***	-0.01***	0.00	-99.65	-4.37***
	(0.03)	(2.28)	(2.21)	(54.72)	(0.00)	(0.03)	(76.49)	(0.35)
Collection Fixed Effects	No	No	No	No	No	No	No	No
Date Fixed Effects	No	No	No	No	No	No	No	No
Adjusted R^2	0.01	0.02	0.03	0.04	0.02	0.02	0.01	0.08
Observations	21,565	21,565	21,565	21,565	21,565	21,565	21,565	21,565

Appendix B continued

Panel B: Excluding wash-trades

	Homogenous collections				Heterogenous collections			
Parameter	Amihud	Lesmond	Goyenko	Roll	Amihud	Lesmond	Goyenko	Roll
Intercept	725.67***	569,501.37***	503,852.57***	1,242,168.93***	18.02***	999,188.22***	950,374.58***	8,695.49***
	(56.05)	(4,172.73)	(3,885.33)	(94,695.96)	(0.70)	(29.21)	(229.24)	(760.33)
Close price	-2.16	323.19	247.76	-7,014.08	-0.03**	3.11***	-9.04**	-27.59
	(1.38)	(205.85)	(165.08)	(4,450.35)	(0.01)	(0.90)	(4.14)	(27.69)
Day volume	-0.33***	-94.63***	-58.09***	23.48	0.01***	-1.36***	-0.37	49.25***
	(0.05)	(15.25)	(10.12)	(109.51)	(0.00)	(0.19)	(0.23)	(6.12)
Day volatility	0.00***	0.00***	-0.00***	-0.00	0.00	0.00**	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age	-4.77***	-533.68***	-1,898.51***	-20,413.83***	-0.23***	7.47***	64.07***	228.91***
	(1.55)	(126.91)	(120.02)	(3,328.18)	(0.01)	(0.79)	(6.84)	(28.57)
Age^2	0.05***	5.33***	11.08***	234.73***	0.00***	-0.03***	-0.12***	-1.04***
	(0.01)	(0.69)	(0.64)	(21.99)	(0.00)	(0.00)	(0.03)	(0.17)
HHI	0.09***	-1.42	-17.29***	370.40***	-0.00***	0.06***	0.02	-4.58***
	(0.03)	(2.28)	(2.21)	(55.38)	(0.00)	(0.01)	(0.11)	(0.33)
Collection Fixed Effects	No	No	No	No	No	No	No	No
Date Fixed Effects	No	No	No	No	No	No	No	No
Adjusted R^2	0.01	0.02	0.03	0.04	0.021	0.078	0.027	0.079
Observations	21,251	21,251	21,251	21,251	21,251	21,251	21,251	21,251

This table displays the estimates of Equation (5), in which the dependent variable is one of the four liquidity proxies' measure for collection *i* on day *d*. Panel A uses a dataset including wash-trades, while Panel B uses a dataset excluding wash-trades. Within each of them, the homogeneous collection columns calculate the liquidity proxies at the collection level by assuming all NFTs within a collection are identical, whereas the heterogeneous collection columns calculate individual NFTs' liquidity proxies and averages them by collection-day. All proxy values are multiplied by 1,000,000 and close price, day volume and day volatility are divided by 1,000 to improve readability. Panel A's sample consists of 21,565 collection-day observations and Panel B's sample consists of 21,151 observations between 23 May 2021 and 21 Aug 2021. Standard errors are shown in parentheses. ***, ** and * represent statistical significance at the 1%, 5% and 10% levels respectively.

Appendix C
Bear market NFT collection liquidity OLS regressions.

Panel A: Including wash-trades								
	Homogenous collections				Heterogenous collections			
Parameter	Amihud	Lesmond	Goyenko	Roll	Amihud	Lesmond	Goyenko	Roll
Intercept	135.22	706,686.96***	548,172.34***	743,543.91***	2.05***	998,832.63***	216,329.30	310.80
	(89.11)	(10,432.71)	(10,575.47)	(164,044.00)	(193.95)	(479,170.20)	(871.59)	(871.59)
Close price	-2.53***	767.25***	819.47***	-6,949.55***	-0.00***	3.17***	16,652.09***	-10.81***
	(0.36)	(96.05)	(101.12)	(1,668.21)	(0.62)	(2,095.33)	(4.19)	(4.19)
Day volume	-0.05**	-0.05**	-61.77**	-55.39***	0.01	-1.09***	107.29	0.73
	(0.02)	(26.54)	(20.20)	(55.86)	(0.37)	(87.36)	(0.46)	(0.46)
Day volatility	62.31	62.31	-14,852.55**	-11,903.59**	373,335.11**	2.71***	8.68	51,777.13*
	(57.02)	(5,979.05)	(4,655.82)	(145,165.80)	(5.35)	(31,292.34)	(54.29)	(54.29)
Age	1.54	-3,139.57***	-4,031.66***	22,315.32**	-0.05	20.08***	-142,790.14***	99.23*
	(4.55)	(487.15)	(477.38)	(9643.97)	(6.57)	(21642.35)	(55.60)	(55.60)
Age^2	0.09*	17.52***	26.22***	179.26	-0.17*	1418.49***	-0.22	-0.22
	(0.05)	(4.86)	(4.66)	(113.60)	(0.05)	(204.11)	(0.62)	(0.62)
HHI	0.03	9.97***	6.47**	-101.15*	-0.03	-157.26	-0.08	-0.08
	(0.02)	(2.63)	(2.60)	(52.01)	(0.03)	(126.64)	(0.32)	(0.32)
Collection Fixed Effects	No	No	No	No	No	No	No	No
Date Fixed Effects	No	No	No	No	No	No	No	No
Adjusted R^2	0.017	0.02	0.03	0.03	0.05	0.04	0.03	0.01
Observations	5,456	5,456	5,456	5,456	5,456	5,456	5,456	5,456

Appendix C continued

Panel B: Excluding wash-trades								
	Homogenous collections				Heterogenous collections			
Parameter	Amihud	Lesmond	Goyenko	Roll	Amihud	Lesmond	Goyenko	Roll
Intercept	167.53*	699,438.13***	548,000.16***	748,653.87***	2.15***	999,783.06***	955,491.32***	553.59
	(88.36)	(10,579.41)	(10,726.35)	(163,184.20)	(0.48)	(41.29)	(606.45)	(662.94)
Close price	-2.50***	828.90***	848.11***	-7,101.06***	-0.08***	1.30***	2.71	-9.67**
	(0.37)	(110.77)	(109.63)	(1,678.39)	(0.00)	(0.36)	(6.88)	(4.14)
Day volume	-0.06*	-95.91**	-73.34**	88.80	0.00	-0.75***	-1.27***	1.31
	(0.04)	(47.25)	(36.16)	(95.79)	(0.00)	(0.24)	(0.46)	(0.85)
Day volatility	54.67	-14,794.25***	-11,991.28***	339,964.15***	2.32***	-5.66**	62.42	-68.84
	(47.99)	(5,536.80)	(4,351.84)	(116,751.30)	(0.65)	(2.32)	(40.69)	(46.29)
Age	0.88	-2,951.21***	-3,977.03***	21,803.10**	-0.05*	2.13	38.73	72.40
	(4.60)	(493.54)	(4,84.86)	(9,817.99)	(0.02)	(1.94)	(25.55)	(51.09)
Age^2	0.09*	16.71***	25.77***	184.65	0.00**	-0.03*	-0.17	0.10
	(0.05)	(4.91)	(4.71)	(115.22)	(0.00)	(0.01)	(0.24)	(0.57)
HHI	0.01	7.91***	5.20**	-77.18*	-0.00*	0.01**	0.11	-0.30
	(0.02)	(2.15)	(2.11)	(42.66)	(0.00)	(0.01)	(0.10)	(0.20)
Collection Fixed Effects	No	No	No	No	No	No	No	No
Date Fixed Effects	No	No	No	No	No	No	No	No
Adjusted R^2	0.02	0.033	0.034	0.044	0.049	0.052	0.003	0.005
Observations	5,369	5,369	5,369	5,369	5,369	5,369	5,369	5,369

This table displays the estimates of Equation (5), in which the dependent variable is one of the four liquidity proxies' measure for collection *i* on day *d*. Panel A uses a dataset including wash-trades, while Panel B uses a dataset excluding wash-trades. Within each of them, the homogeneous collection columns calculate the liquidity proxies at the collection level by assuming all NFTs within a collection are identical, whereas the heterogeneous collection columns calculate individual NFTs' liquidity proxies and averages them by collection-day. All proxy values are multiplied by 1,000,000 and close price, day volume and day volatility are divided by 1,000 to improve readability. Panel A's sample consists of 5,456 collection-day observations and Panel B's sample consists of 5,369 observations between 24 Mar 2021 and 22 May 2021. Standard errors are shown in parentheses. ***, ** and * represent statistical significance at the 1%, 5% and 10% levels respectively.

Appendix D
Bear market NFT collection liquidity regressions using collection and date fixed effects.

Panel A: Including wash-trades								
	Homogeneous collections				Heterogeneous collections			
Parameter	Amihud	Lesmond	Goyenko	Roll	Amihud	Lesmond	Goyenko	Roll
Intercept	320.74	996,514.83***	913,757.23***	3,428,996	6.08	999,620.60***	7293924	-16,870
	(2,740.12)	(55,075.97)	(69,374.98)	(3,454,857.00)	(5.51)	(518.78)	(6,318,449.00)	(34,952.67)
Close price	2.29*	112.86	44.3	400.03	0.00	0.37	8253.38	-13.46
	(1.19)	(120.59)	(93.40)	(1,280.61)	(0.01)	(0.66)	(6,176.46)	(17.15)
Day volume	-0.01	-3.85	-4.65	-20.25	0.00	-0.26**	80.96	-0.16
	(0.02)	(5.33)	(4.92)	(95.69)	(0.00)	(0.12)	(63.61)	(0.28)
Day volatility	28.31	407.1	98.85	3,367.04	2.69***	2.12	-18081.8	-20.15
	(40.77)	(445.56)	(546.48)	(28,243.17)	(0.49)	(4.06)	(34,676.50)	(57.84)
Age	-108.98	842.72	4128.02	-322,187	-0.72	90.97	-967241	2,641
	(335.46)	(5,975.05)	(7,957.25)	(337,414.00)	(0.58)	(55.22)	(799,053.40)	(4,237.86)
Age^2	0.05	15.18	30.76***	123.42	0.00	-0.04	-15.03	-0.9
	(0.06)	(9.50)	(9.00)	(244.34)	(0.00)	(0.06)	(251.42)	(0.83)
HHI	0.23	-1.86	-6.17	-185.46	0.00	-0.15	816.41**	-2.34
	(0.26)	(4.15)	(4.99)	(162.43)	(0.00)	(0.11)	(388.05)	(1.93)
Collection Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.13	0.86	0.84	0.83	0.06	0.79	0.73	0.13
Observations	5,456	5,456	5,456	5,456	5,456	5,456	5,456	5,456

Appendix D continued

Panel B: Excluding wash-trades

	Homogeneous collections				Heterogeneous collections			
Parameter	Amihud	Lesmond	Goyenko	Roll	Amihud	Lesmond	Goyenko	Roll
Intercept	232.02	999,499.78***	921,898.23***	3,393,714	6.31	999,604.56***	959,197.89***	-15,124.00
	(2,727.17)	(54,244.52)	(68,902.79)	(3,469,906.00)	(5.48)	(454.68)	(11,504.26)	(34,349.67)
Close price	2.25*	116.26	43.01	429.97	-0.01	0.70	11.50	-16.63
	(1.17)	(121.72)	(94.12)	(1,301.30)	(0.01)	(0.55)	(13.07)	(16.65)
Day volume	-0.03	3.67	2.37	-49.58	0.00	-0.43***	0.06	-0.48
	(0.02)	(6.48)	(6.38)	(164.57)	(0.00)	(0.15)	(0.15)	(0.55)
Day volatility	26.77	678.7	292.29	4,762.77	2.18***	5.47	37.63	-73.11
	(36.14)	(573.27)	(625.86)	(23,495.36)	(0.62)	(6.47)	(41.00)	(62.65)
Age	-97.35	394.11	3,052.62	-316,471	-0.73	81.24*	-2,696.88**	2,442
	(333.55)	(5,869.77)	(7,894.36)	(338,217.00)	(0.57)	(48.00)	(1,329.67)	(4,153.74)
Age^2	0.06	12.85	29.00***	125.59	0.00	-0.01	0.19	-0.43
	(0.06)	(9.55)	(9.10)	(246.05)	(0.00)	(0.04)	(0.34)	(0.73)
HHI	0.23	-1.43	-5.59	-191.07	0.00	-0.11	0.18	-2.34
	(0.26)	(4.10)	(4.95)	(164.65)	(0.00)	(0.08)	(0.96)	(1.90)
Collection Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.13	0.86	0.84	0.82	0.02	0.20	0.17	0.06
Observations	5,369	5,369	5,369	5,369	5,369	5,369	5,369	5,369

This table displays the estimates of Equation (5), including collection level fixed effects and clustering standard errors by collection, in which the dependent variable is one of the four liquidity proxies' measure for collection *i* on day *d*. Panel A uses a dataset including wash-trades, while Panel B uses a dataset excluding wash-trades. Within each of them, the homogeneous collection columns calculate the liquidity proxies at the collection level by assuming all NFTs within a collection are identical, whereas the heterogeneous collection columns calculate individual NFTs' liquidity proxies and averages them by collection-day. All proxy values are multiplied by 1,000,000 and close price, day volume and day volatility are divided by 1,000 to improve readability. The sample consists of 5,369 collection-day observations between the bear market period of 24 Mar 2021 and 23 May 2021. Robust standard errors are clustered to the collection level to account for within-collection cross-sectional correlations in regression residuals and are shown in parentheses. ***, ** and * represent statistical significance at the 1%, 5% and 10% levels respectively.

Appendix E

Logged NFT collection liquidity using simple OLS regression.

Panel A: Including wash-trades								
	Homogeneous collections				Heterogeneous collections			
Parameter	Ln(Amihud)	Ln(Lesmond)	Ln(Goyenko)	Ln(Roll)	Ln(Amihud)	Ln(Lesmond)	Ln(Goyenko)	Ln(Roll)
Intercept	613.44***	465,536.10***	405,593.32***	422,263.00***	13.87***	692,536.16***	654,429.06***	6,344.32***
	-(44.70)	-(2,436.34)	-(2,409.17)	-(9,516.20)	(0.55)	(26.53)	(2,051.27)	(549.39)
Close price	-2.35***	346.08***	312.28***	-1,024.03**	-0.03***	1.82***	103.33***	-11.79
	-(0.83)	-(124.78)	-(110.45)	-(450.99)	(0.01)	(0.36)	(27.10)	(13.70)
Day volume	-0.25***	-66.15***	-47.45***	-17.07	0.01***	-0.73***	-38.46***	26.51***
	-(0.04)	-(11.71)	-(7.81)	-(10.91)	(0.00)	(0.10)	(10.27)	(5.08)
Day volatility	0.00	-0.00***	-0.00***	0.00	0.00	0.00***	0.00***	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age	-0.34	-1,030.86***	-1,931.79***	2,504.27***	-0.17***	5.57***	-224.83***	145.81***
	-(1.32)	-(71.57)	-(72.16)	-(315.74)	(0.01)	(0.67)	(68.73)	(19.14)
Age^2	0.03***	6.31***	11.00***	-1.15	0.01***	-0.02***	1.39***	-0.52***
	-(0.01)	-(0.40)	-(0.40)	-(1.98)	(0.00)	(0.00)	(0.37)	(0.12)
HHI	0.03**	1.39	-8.08***	14.02***	-0.01***	0.04***	3.08***	-2.42***
	-(0.01)	-(0.87)	-(0.89)	-(3.36)	(0.00)	(0.01)	(0.52)	(0.16)
Collection Fixed Effects	No	No	No	No	No	No	No	No
Date Fixed Effects	No	No	No	No	No	No	No	No
Adjusted R^2	0.01	0.02	0.03	0.03	0.01	0.02	0.01	0.05
Observations	30,780	30,780	30,780	30,780	30,780	30,780	30,780	30,780

Appendix E continued

Panel B: Excluding Wash-trades								
	Homogeneous collections				Heterogeneous collections			
Parameter	Ln(Amihud)	Ln(Lesmond)	Ln(Goyenko)	Ln(Roll)	Ln(Amihud)	Ln(Lesmond)	Ln(Goyenko)	Ln(Roll)
Intercept	637.85***	462,237.23***	40,4471.09***	424,268.14***	13.55***	692,824.16***	668,760.28***	5,599.30***
	(44.03)	(2,451.13)	(2,417.90)	(9,514.43)	(0.53)	(11.54)	(99.59)	(531.19)
Close price	-2.27***	369.50***	317.45***	-1,032.16**	-0.03***	1.26***	-0.77	-17.45
	(0.82)	(128.50)	(110.77)	(452.87)	(0.01)	(0.26)	(1.39)	(13.06)
Day volume	-0.27***	-75.22***	-51.52***	-16.68	0.01***	-0.63***	-0.47***	28.90***
	(0.05)	(11.98)	(8.44)	(12.33)	(0.00)	(0.09)	(0.10)	(5.02)
Day volatility	0.00	-0.00***	-0.00***	0.00	0.00	0.00***	0.00**	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age	-0.68	-1,030.25***	-1,925.43***	2,592.34***	-0.17***	2.59***	27.74***	149.61***
	(1.32)	(72.14)	(72.56)	(318.67)	(0.01)	(0.31)	(2.93)	(18.25)
Age^2	0.03***	6.42***	10.97***	-1.78	0.01***	-0.01***	-0.07***	-0.56***
	(0.01)	(0.40)	(0.40)	(1.99)	(0.00)	(0.00)	(0.02)	(0.12)
HHI	0.02*	1.76**	-6.72***	13.65***	-0.01***	0.04***	0.20***	-2.29***
	(0.01)	(0.82)	(0.84)	(3.13)	(0.00)	(0.00)	(0.03)	(0.14)
Collection Fixed Effects	No	No	No	No	No	No	No	No
Date Fixed Effects	No	No	No	No	No	No	No	No
Adjusted R^2	0.01	0.03	0.03	0.03	0.02	0.08	0.01	0.06
Observations	30,350	30,350	30,350	30,350	30,350	30,350	30,350	30,350

This table displays the estimates of Equation (5), in which the dependent variable is the natural log of one of the four liquidity proxies' measure for collection *i* on day *d*. Panel A uses a dataset including wash-trades, while Panel B uses a dataset excluding wash-trades. Within each of them, the homogeneous collection columns calculate the liquidity proxies at the collection level by assuming all NFTs within a collection are identical, whereas the heterogeneous collection columns calculate individual NFTs' liquidity proxies and averages them by collection-day. All proxy values are normalised by adding 1 to each of them prior to taking the natural log. They are then multiplied by 1,000,000 and close price, day volume and day volatility are divided by 1,000 to improve readability. Panel A's sample consists of 30,780 collection-day observations and Panel B's sample consists of 30,350 observations between 1 Feb 2021 and 21 Aug 2021. Standard errors are shown in parentheses. ***, ** and * represent statistical significance at the 1%, 5% and 10% levels respectively.

Appendix F

Logged Liquidity Proxies NFT collection liquidity regressions using collection and date fixed effects

Parameter	Ln(Amihud)	Ln(Lesmond)	Ln(Goyenko)	Ln(Roll)
Intercept	13.55***	692,824.16***	668,760.28***	5,599.30***
	(0.53)	(11.54)	(99.59)	(531.19)
Close price	-0.03***	1.26***	-0.77	-17.45
	(0.01)	(0.26)	(1.39)	(13.06)
Day volume	0.01***	-0.63***	-0.47***	28.90***
	(0.00)	(0.09)	(0.10)	(5.02)
Day volatility	0.00	0.00***	0.00**	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)
Age	-0.17***	2.59***	27.74***	149.61***
	(0.01)	(0.31)	(2.93)	(18.25)
Age^2	0.01***	-0.01***	-0.07***	-0.56***
	(0.00)	(0.00)	(0.02)	(0.12)
HHI	-0.01***	0.04***	0.20***	-2.29***
	(0.00)	(0.00)	(0.03)	(0.14)
Collection Fixed Effects	Yes	Yes	Yes	Yes
Date Fixed Effects	Yes	Yes	Yes	Yes
Adjusted R^2	0.02	0.08	0.01	0.06
Observations	30,350	30,350	30,350	30,350

This table displays the estimates of Equation (5), including both collection level and date fixed effects, in which the dependent variable is the natural log of one of the four liquidity proxies' measure for collection *i* on day *d*. This table shows the estimates for the homogeneous assumption, heterogeneous assumption, prewash trading exclusion and post-wash trading exclusion datasets, which are identical. All proxy values are normalised by adding 1 to each of them prior to taking the natural log. They are then multiplied by 1,000,000 and close price, day volume and day volatility are divided by 1,000 to improve readability. The sample consists of 30,530 collection-day observations between 1 Feb 2021 and 21 Aug 2021. Robust standard errors are clustered to the collection level to account for within-collection cross-sectional correlations in regression residuals and are shown in parentheses. ***, ** and * represent statistical significance at the 1%, 5% and 10% levels respectively.