DAI DIGITAL ART INDEX : A ROBUST PRICE INDEX FOR HETEROGENEOUS DIGITAL ASSETS

Min-Bin Lin, Bingling Wang, Fabian Y.R.P. Bocart, Christian M. Hafner, Wolfgang K. Härdle

LIDAM Discussion Paper ISBA 2022 / 36



UCLouvain

ISBA | LIDAM Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/isba/publication.html

代 DAI Digital Art Index

A robust price index for heterogeneous digital assets

Min-Bin Lin[1] Bingling Wang[2] Fabian Y.R.P. Bocart[3]
Christian M. Hafner[4] Wolfgang Karl Härdle[5]

November 18, 2022

Abstract

The big bang of non-fungible tokens (NFTs) has caused the birth of a brand new era for digital art. NFTs, driven by blockchain and smart contracts, provide both artists and art collectors an unprecedented marketplace which is equipped with more security, flexibility, publicity, and freedom to monetize. Unlike the conventional art market, NFT art market trades works of art in rapid succession and has no sluggish business cycle. Yet, the emergence of such a market has been considered to be packed with speculation and economic uncertainty,

[1]International Research Training Group 1792, School of Business and Economics, Humboldt-Universität zu Berlin. Unter den Linden 6, 10099 Berlin, Germany. Email: min-bin.lin@hu-berlin.de

[2]International Research Training Group 1792, School of Business and Economics, Humboldt-Universität zu Berlin. Unter den Linden 6, 10099 Berlin, Germany. Email: bingling.wang@hu-berlin.de

[3]Artnet, 233 Broadway, 26th Floor, New York, NY 10279-2600, USA. Email:fabian.bocart@gmail.com

[4]Louvain Institute of Data Analysis and Modelling in economics and statistics, ISBA, Université catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium. Email: christian.hafner@uclouvain.be

[5]Blockchain Research Center, Humboldt-Universität zu Berlin, Germany; Asia Competitiveness Institute, Lee Kuan Yew School of Public Policy, National University of Singapore, Singapore; Sim Kee Boon Institute for Financial Economics, Singapore Management University, Singapore; Faculty of Mathematics and Physics, Charles University, Czech Republic; National Yang Ming Chiao Tung University, Taiwan. Email: haerdle@hu-berlin.de

given the limited understanding toward this market. To precisely depict the NFT art market and offer a gauge for market volatility, we construct the price index – Digital Art Index (DAI) – using a hedonic regression on the top 10 liquid NFT art collections and provide an empirical overview on this emerging market. Due to the common impact of outlying phenomenon in the art markets, we propose two innovative alternative procedures – Huberization and DCS-t filtering to robustify the index. Moreover, by considering each artwork's time-variant and time-invariant characteristics, we are able to investigate price determinants and access the intrinsic value of the market.

All the supplementary materials and source codes can be found in the Q^2 ecosystem: Quantlet.com and Quantinar.com.

Keywords: Non-fungible token, NFT, blockchain technology, digital art, hedonic regression, index construction, robustness
JEL Classification: C14, C43, C51, G10, Z11

1 Introduction

Non-fungible token (NFT), a digital token as a proof of ownership and authenticity for assets, offers a futuristic possibility for art trading and thus impact on the art market. Its market capitalization has boomed up to US $ 41 billion in 2021, which is 30 % growth by last year; but the conventional art market meanwhile has shrunk by 22 % in sales from 2019 (Dailey, 2022; Chainalysis, 2022). A preliminary comparison on quarterly sales of four different art markets worldwide can be found in Figure 1.1, sourced from Artnet.com. We can observe a surge of NFT art since the COVID-19 pandemic in 2020; while it receives a decline later in the Q4 of 2021. Other art markets also receive a boost on sales since the pandemic.

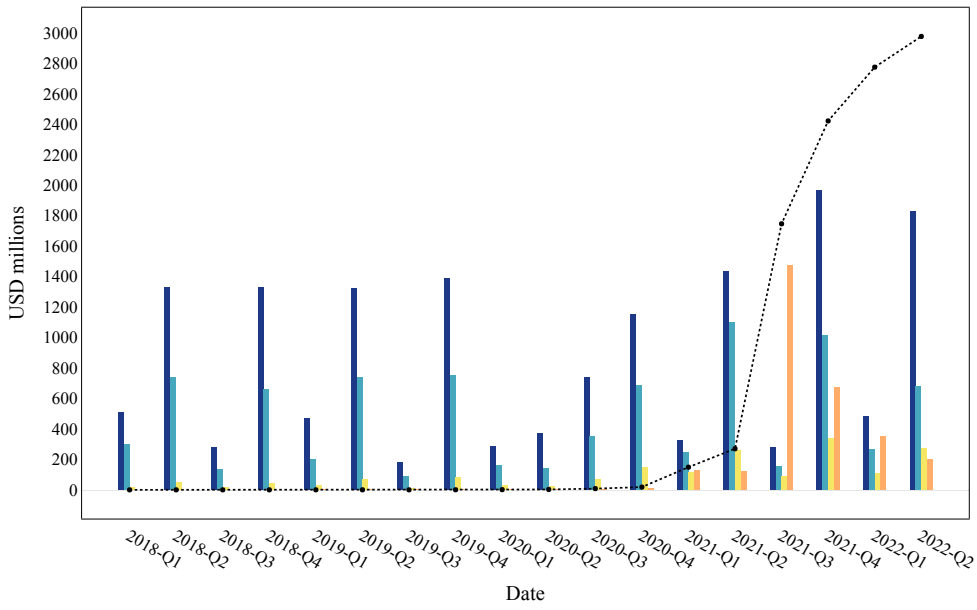


Figure 1.1: Quarterly sales in USD millions of **post war**, **contemporary**, **ultra contemporary**, **NFT** art markets, plotted in order. The dotted line is the cumulative sales on NFT.

Driven by blockchain (BC) technology and smart contracts (SC), NFTs

inherit their functional features – i.e. network decentralization, privacy and encryption, immutability of data, and programmability in contracts. NFT digital art market has been seen as a rising opportunity for the creators from professionals to amateurs to interact with under-served populations. It furnishes more equitable access to wealth and the inclusion potential to artworks. Yet, such a market is still under development and the public understanding towards it is not fully formed.

In this paper we investigate the top 10 NFT art collections ranked by transaction volume on the marketplace –Opensea, which is currently the most traded NFT marketplace. The top 10 collections have consisted of total 192,763 digital art pieces and represented around 84 % of total market trading volume. In order to outline the general trends of the NFT art market and develop a broad understanding of its risk and return potential, we construct a characteristic-free price index – Digital Art Index (DAI) and then conclude its price movement and determinants. We apply the hedonic modeling method – which is derived from data that control for differences in the characteristics of assets in various samples – and use the Kalman filtering to estimate the hidden index time-series process. Our results can be used for observing and evaluating this new market and for the comparison with asset classes, e.g. cryptocurrencies (see CRIX).

The construction of an art price index is different from building indices for stocks or similar assets, which are based on prices of ientical shares. For the art market, very heterogeneous objects are to be considered. McAndrew (2010) indicates that the essence of art indices are: representativeness, liquidity, and capacity, i.e. the potential and values of sales. Due to the nature of art market: high heterogeneity among artworks and infrequent trading behaviors (unbalanced panel data), the determination of changes in market value is difficult to ascertain (Ginsburgh et al., 2006; Beckert and Rössel, 2013). An additional feature one needs to cope with is the presence of of outliers in art pricing, i.e. Figure 1.2. An average (arithmetic mean or median) price index might thus not be sufficient. Most art indices (e.g. Artnet.com,

Sotheby's Mei Moses) are based on data reflecting the price of an artwork at a certain time, a function of the fixed characteristics of work – time-invariant characteristics – and elements that vary over time – time-variant characteristics. For the NFT digital art, we collect the metadata and historical prices of each work and feature it by its time-variant characteristic (i.e. number of sales, owner address) and time-invariant characteristics (i.e. collection name, NFT scheme).

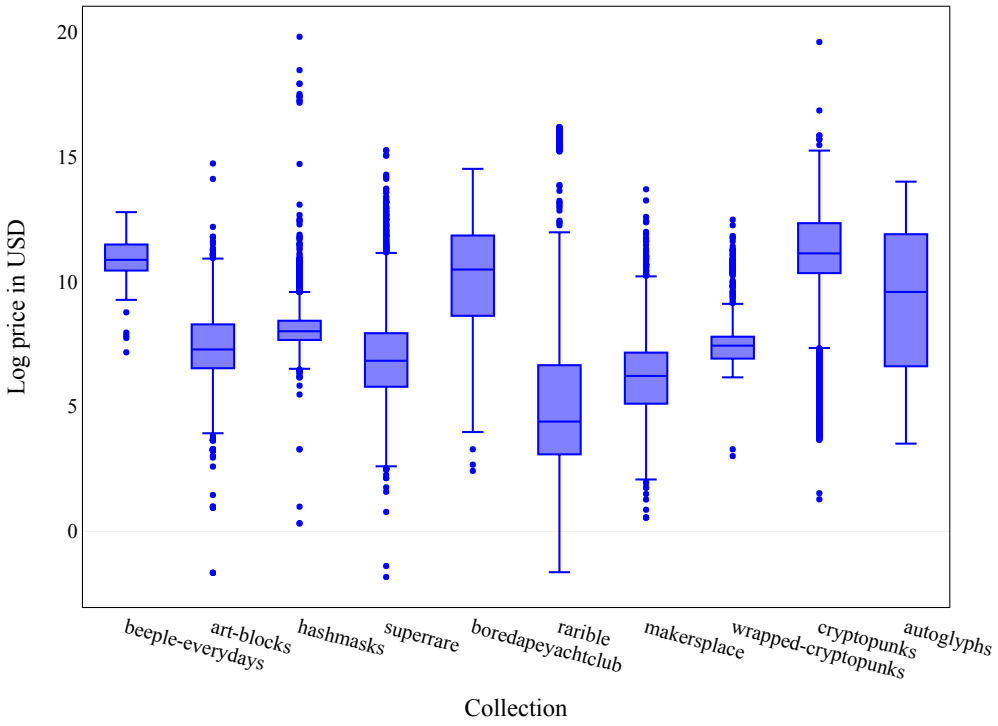


Figure 1.2: The existence of outliers in NFT digital art pricing.

Thus far, there is limited amount of literature having been developed to comprehend the economics of NFT art market and underpin its role to the conventional art market. Kräussl and Tugnetti (2022) overview NFT pricing and valuation methods, and identify NFTs having a similar investment profile – high yield and high risk – with classic collectibles. Goldberg et al. (2021);

Dowling (2022a) investigate virtual land pricing in Decentraland. Nadini et al. (2021) study the behaviors of NFT traders from a perspective of network theory and offer brief statistics for the market. Ante (2021); Dowling (2022b) look into cointegration relationship and time varying co-movements among the selected cryptos and NFTs. Borri et al. (2022) construct a market index based on repeat-sales regression (RSR) and compare it to the conventional asset and crypto markets. Due to the lack of a representative market index most of the studies are nevertheless restricted to the individual NFT assets. In order to mitigate such a research gap and facilitate the realization of market research, we develop a hedonic regression based price index. Based on a comprehensive dataset collected in this study we include the characteristics and historical prices of each art work. We apply two alternative methods of robustification of the price index, the first is a classical Huberization, the second an application of a dynamic conditional score (DCS) model to the filtering problem. By applying robustification to our index construction, we reduce the influence of the immanent outliers.

The paper proceeds as follows. Section 2 portrays the selling processes for NFT art and convention art markets and highlights their stylized facts. In addition, this section essentially compares it with the conventional art market. Section 3 exposes the common used estimation methods for constructing art indices. Section 4 elaborates the methodologies used. Section 5 presents the data attributes and data processing. We demonstrate the result in Section 6 and offer a discussion for interpretation in Section 7 Finally, Section 8 concludes the findings.

2 NFT art market

Since the emergence of blockchain technology, there have been many adoptions in different sectors across all applications – crypto utility tokens in particular. NFTs are different from utility tokens. They offer the possibility for digital assets to be owned and traded. We give a definition of NFT as – *a*

cryptographically secured and unique digital identifier minted through smart contracts on a blockchain that cannot be replicated, subdivided or substituted. It can be used to represent both tangible and intangible assets and to facilitate a programmable smart contract between seller and buyer (Zinovyeva et al., 2022). In such a case, an artwork can be traded directly among creators and buyers in an automated and flexible fashion and the transaction can be transparent and protected. An outlook for the NFT digital art market can be found in Allen et al. (2022); Pawelzik and Thies (2022).

On the other hand, in the conventional art market artists and creators are generally required to seek for a dealer or a gallery to access potential buyers, e.g. individual collectors, museums, or auction houses (Velthuis, 2003; Zorloni, 2005). Before a work enters a marketplace, it involves a series of valuation and curation processes. Due to the involvement of many third parties, it can therefore be costly and time consuming. A review literature of the contemporary art market and its economics is given by McEachern (2016). In the following, we briefly overview both the NFT and conventional art markets.

2.1 Selling process

In Figure 2.1 we illustrate the selling processes for the NFT and conventional art markets. The arrow lines flow with artwork transfers. The primary market is where each artwork issues when it is first made available for sale by a creator; and the secondary market is where the artwork get resales since the first sale from its creator. In the NFT art market, a creator can be an institute (e.g. Larva Labs, Dapper Labs) or an individual. Most of the artworks are traded in the secondary market via a marketplace (e.g. Opensea, Rarible) except the institutes that can support the architecture of a blockchain. Art creators generally preserve the copyright of artworks and are able to reproduce a series of similar works. Also, they can decide the listing price – the first sale price – and control the bidding process with buyers. Due to the

programmability in NFTs, a creator can later demand a royalty, a preset percentage of the sale price ranging from 5-10% typically. The crypto payment and the transfer of NFT are processed automatically and simultaneously in a single transaction between two mutually distrusting parties on a blockchain network. NFT consequently establishes diverse contented marketplaces and enables a vibrant and sustainable revenue stream for creators.

In a conventional art market art creators depend on a centralization dealer / gallery to proceed with the selling process. The listing price is decided via a valuation process and a negotiation between creators and dealers. After the first sale, the artwork can be traded in different channels (Zorloni, 2005; Towse and Hernández, 2020); however, each channel might be independent from each other. Moreover, there are a few causal linkages between different marketplaces across the world and no feedback relationships in between (Le Fur, 2020). The transaction history and pricing can be less transparent in comparison to the NFT art market, where all the transaction history are secured by blockchain technology. Resale royalty right, also called droit de suite in the conventional market might be difficult due to the lack of traceability in sales history.

2.2 Economics of art market

Art has been considered as one of the investment instruments that offers lower volatility and lower correlation with other assets, thus being used for portfolio diversification (Mei and Moses, 2002a; Worthington and Higgs, 2004; Campbell, 2008). In terms of the performance in returns, Bocart et al. (2020); Whitaker and Kräussl (2020) indicate that contemporary art and gallery portfolio can perform almost as well as the S&P 500, and the majority of art markets do not load significantly on momentum or liquidity factors. Penasse and Renneboog (2021) show that the art market is subject to frequent booms and busts so it is not applicable to rational consumer models and exhibits an opportunity for spectacular profits. Goetzmann et al. (2011); Renneboog and

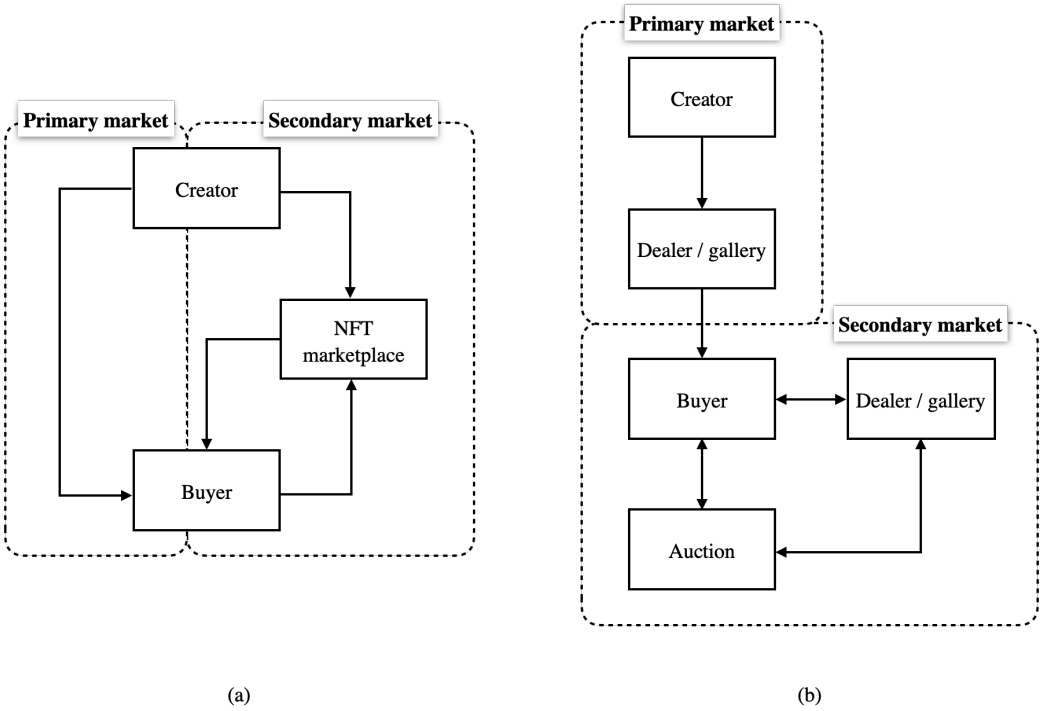


Figure 2.1: Selling processes for (a) NFT art market; and (b) conventional art market.

Spaenjers (2013) direct attention to income inequality and art market sentiment and show their correlation to art price trends. A wide body of literature analyzes art price determinants via – the work of art (size, material, genre), the artist (age, sex, place of residence), and the sales channel (location, affiliation) (Rengers and Velthuis, 2002; Li et al., 2021; Garay, 2021). This can be summarized on the main stylized facts of the conventional art market – i.e. *infrequent trading, price inequality, illiquidity, centralized patrons, less transparency in price formulation, multiple independent marketplaces.*

The exchanges and transfers of NFT art, compared with the conventional artworks, proceed safely and with (nearly) no intermediaries. Cryptos are essentially the means of payment in NFT markets – e.g. Ethereum (ETH), Solana (SOL), therefore, crypto pricing can affect NFT asset pricing and the markets (Ante, 2021). Dowling (2022b) remarks that NFT assets (i.e. Decentraland, Cryptopunks) have a limited volatility spillover from cryptos. Urom

et al. (2022) find that the connectedness of returns and volatility among NFT markets and other conventional markets (e.g. stock, commodity, crypto) are higher during extreme market conditions; and NFT markets usually is a net return shock receiver. In terms of liquidity, Wilkoff and Yildiz (2022) conclude that there is a positive relationship between ETH returns and NFT illiquidity. In other words, while ETH returns are high investors are unwilling to buy NFTs. In Figure 2.2, we can further observe that the downwards and upwards of total transaction volume of NFT art are often corresponding to dropping and rising of CRIX index proposed by Trimborn and Härdle (2018), an index for tracking the evolution of cryptocurrency market.

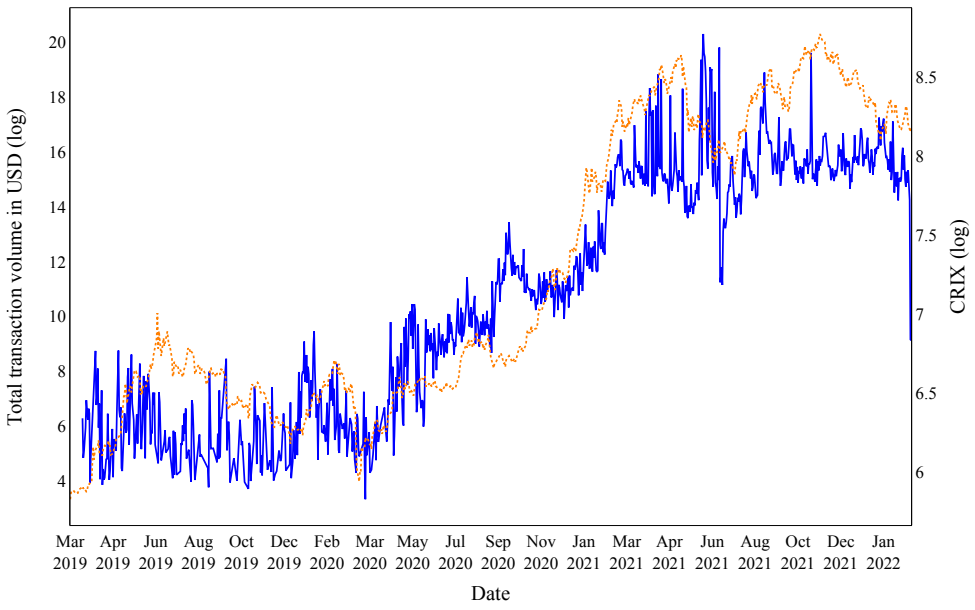


Figure 2.2: Time series of total transaction volume of NFT art in USD (solid) and CRIX (dotted).

Similar to the conventional art, NFT artworks are traded infrequently and there exists an imbalance in asset pricing. In the report of Vasan et al. (2022), they indicate that market first-mover creators, in general, own an advantage in pricing and sales; and a network effect (i.e. from twitter followers) can

be observed. Rarity and uniqueness of collectibles induce market demand and thus offer a higher price premium, known as rarity premium (Koford and Tschoegl, 1998). For instance, the on-chain art generative collection – Autoglyphs, due to its limited number of issues, has consistent pricing and resale rates in Figures 1.2 and 3.1. Mekacher et al. (2022) quantify rarity based on natural language descriptions (i.e. traits) in the metadata of NFT art and show that high rarity induces high returns. Moreover, in the real world art market curation and valuation play a critical role for accessing quality of artworks. Applying such a mechanism to the digital art market, ArtBlocks, for example, offers a curation board for evaluating generative artworks from their scripts.

We can observe there exist many adoptions from the conventional art market on the NFT art market. Transforming and evolving towards digitalization, the NFT art market has provided many distinctive technical properties – backed by blockchain technology and smart contracts – that attract different groups of art collectors (i.e. crypto investors) and consequently influence trading behaviors. In this study we dedicate to categorize the stylized facts for such a brand new market and empirically compare it to other existing markets.

3 Towards DAI, an Index for Digital Art

Two estimation methods – Repeat-sales regression (RSR) and hedonic regression (HR) are commonly applied to construct price indices for heterogeneous goods that comprise various characteristics. In absence of individual characteristics, RSR uses the prices traded within two distinct moments in time. Bailey et al. (1963) employs the RSR for real estate price index construction

and it is formulated as:

$$\underbrace{\log\left(\frac{p_{t+s}^i}{p_t^i}\right)}_{\text{Relative price}} = \delta_{t+s} d_{t+s}^i + u_{t+s}^i \tag{3.1}$$

where p_t^i is the observed price of object $i \in \{1, \ldots, N\}$ in period $t \in \{1, \ldots, T\}$ with $T \geq t + s > t$; and t and $t + s$ are the periods for the first sale and the second sale, respectively. δ_t is the index; $d_t^i \in \{0, 1\}$ is the time dummy; and ε_t^i is the error term and $u \overset{\text{iid}}{\sim} \mathcal{N}(0, {\sigma_\varepsilon}^2)$. One can see δ_t here as the average return at period t in the portfolio. RSR might suffer though from sample biases when the data is rather sparse. Moreover, it only considers relative prices (returns), so that the individual objects having more than two sales are not included.

Hedonic regression model regresses prices on various characteristics of assets transacted and is able to adapt the changing characteristics over time. Though the model has been proposed since the applications of Waugh (1928) and Court (1939), later Griliches (1961, 1971) and Rosen (1974) have justified it as a quality-adjusted price index and HR model has started widely being implemented in e.g. house pricing, art pricing. A general form for an HR model is written as:

$$\log\left(p_t^i\right) = \sum_{\tau=1}^{t} \beta_\tau d_\tau^i + \sum_{k=1}^{m} \gamma_k v_k^i + \sum_{\tau=1}^{t} \sum_{j=1}^{n} \theta_\tau^j w_\tau^{ij} + \varepsilon_t^i \tag{3.2}$$

where p_t^i is the observed price of object $i \in \{1, \ldots, N\}$ in period $t \in \{1, \ldots, T\}$; β_τ is the estimated coefficient of time dummy variable $d_\tau^i \in \{0, 1\}$ with $\tau \in \{1, \ldots, T\}$; and ε_t^i is the error term with $\varepsilon \overset{\text{iid}}{\sim} \mathcal{N}(0, {\sigma_\varepsilon}^2)$. p_t^i is a function of m time-invariant characteristics v_k^i with $k \in \{1, \ldots, m\}$ and of n time-variant characteristics w_τ^{ij} at period τ with $j \in \{1, \ldots, n\}$; and γ_k and θ_τ^j are the corresponding coefficients. Characteristics depend on the type of objects studied. We might encounter omitted bias and misspecification challenges (Wallace and Meese, 1997).

We can easily observe both the RSR and HR concepts coincide if there are no time-variant characteristics in the hedonic factors. Then, Equation 3.2 can be written as:

$$\log\left(p_t^i\right) = \sum_{\tau=1}^{t} \beta_\tau d_\tau^i + \sum_{k=1}^{m} \gamma_k v_k^i + \varepsilon_t^i \tag{3.3}$$

We can, thus, get the difference between the first sale log price p_t^i and the second sale p_{t+s}^i in an HR fashion as below. Note that d is being non zero at time t and $t+s$ at which the two observable p occur in succession.

$$\begin{aligned}\log\left(p_{t+s}^i\right) - \log\left(p_t^i\right) &= (\sum_{\tau=1}^{t+s} \beta_\tau d_\tau^i + \sum_{k=1}^{m} \gamma_k v_k^i + \varepsilon_{t+s}^i) - (\sum_{\tau=1}^{t} \beta_\tau d_\tau^i + \sum_{k=1}^{m} \gamma_k v_k^i + \varepsilon_t^i) \\ &= \beta_{t+s} d_{t+s}^i + (\varepsilon_{t+s}^i - \varepsilon_t^i)\end{aligned} \tag{3.4}$$

Similarly to other collectibles (e.g. equipment, traditional artworks, real estate), the NFT art market is rather non-liquid and often has low frequency in sales. From Figure 3.1 one observes that most of the collections listed have the medians of number of sales around 1. As discussed in Shiller (2008), if either the fraction of sales or the relative price of new assets vary a lot, the presence of new sales along with existing sales can be of essence. Hence, RSR concept does not fit the NFT artworks with only a single sale since it cannot represent a precise resolution of the market. One may argue that incorporating both new and existing asset prices can be an issue (Shiller, 2008), i.e. the supply of new NFTs varies over time as market conditions change. The NFT art market is emerging and uneven in assets' prices and sales, therefore, it is indeed less stable and alters with the market conditions.

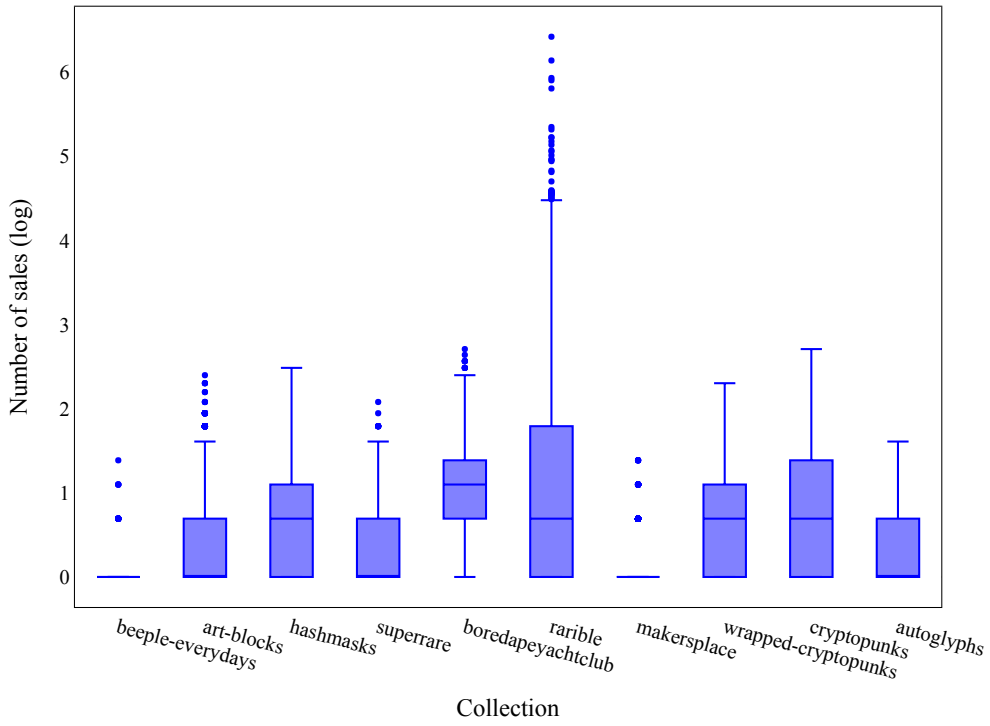


Figure 3.1: Low frequent sales in the NFT art market.

4 Methodology

After going through the preliminary for the methodology, in this section we explain in detail the regression model and the construction of art index. We explicitly consider the hedonic price index model and refine it by emphasizing outlying and influential data and issues on biased sampling.

4.1 Hedonic regression

To consider the characteristics of NFT art and understand its pricing fundamentally, we apply a hedonic regression model and extend it from the work of Bocart and Hafner (2015). Due to the infrequent trading of NFT

art, we actually have to deal with an unbalanced panel data structure, i.e. panel data where the individual time series have unequal length of observations. Writing $y_t = (y_t^1, \ldots, y_t^{N_t})^\top$ where y_t^i is the log price of the i-th sale at time $t \in \{1, \ldots, T\}$ with $i \in \{1, \ldots, N_t\}$; $\beta_t = (\beta_t^1, \ldots, \beta_t^{N_t})^\top$; and $\varepsilon_t = (\varepsilon_t^1, \ldots, \varepsilon_t^{N_t})^\top$ with $\varepsilon \overset{\text{iid}}{\sim} \mathcal{N}(0, {\sigma_\varepsilon}^2)$. Note that i here refers to the sales instead of an object. Equation 3.2 can be equivalently expressed as:

$$y_t = d_t \beta_t + X_t^\top \alpha + \varepsilon_t \tag{4.1}$$

where $d_t = (1, \ldots, 1)^\top$ is a $(N_t \times 1)$ vector. $X_t = (v_k, w_t^j)$ is of dimension $(N_t \times K)$ and contains K explanatory variables including time-variant and time-invariant characteristics at time t; and $\alpha = (\alpha^1, \ldots, \alpha^K)^\top$ is the corresponding vector parameter. Note that N_t changes with time t.

Once we obtain the parameter α, interpreted as implicit prices of the various characteristics, β_t can be estimated:

$$\hat{\beta_t}^{OLS} = \underset{\beta_t \in \mathbb{R}}{\operatorname{argmin}} \left\{ \|y_t - d_t \beta_t - X_t^\top \alpha\|_2^2 \right\} \tag{4.2}$$

Since the model marginalizes over X_t, one creates the characteristic-free price index $\hat{\beta_t}^{OLS}$.

Incorporating with lasso (least absolute shrinkage and selection operator) for model regularization, we subject Equation 4.1 to: $\|\alpha\|_1 < b$ and the lasso determined coefficient $\hat{\beta_t}^{lasso}$ in the equivalent Lagrangian form:

$$\hat{\beta_t}^{lasso} = \arg \min_{\beta_t \in \mathbb{R}, \alpha \in \mathbb{R}^K} \left\{ \|(y_t - d_t \beta_t - X_t^\top \alpha)\|_\rho + \lambda \|\alpha\|_1 \right\} \tag{4.3}$$

where b is a pre-defined parameter that determines the degree of regularization; ρ is a certain loss function, i.e. Huber loss, and λ is regularization parameters.

Given the estimate $\hat{\beta}_t$, the price index with base 100 in $t = 1$ is defined

as:

$$\text{Index}_t = 100\frac{\exp\left(\hat{\beta}_t\right)}{\exp\left(\hat{\beta}_1\right)} \tag{4.4}$$

4.2 Selection bias correction

We apply Heckman's sample selection model, known as Tobit-II model (see Heckman (1976, 1979)) to address the potential problem of sample selection bias – especially omitted variable bias (Malpezzi et al., 2003; Collins et al., 2009). In our case, we only observe a sub sample of the NFT transactions that have been successfully transacted but other transaction event types (i.e. bid entered, cancelled) cannot be included in the models of Section 4.1 due to unobserved dependent variables (i.e. price). The Heckman model utilizes the regression covariates to approximate selection, transforms the prediction, and then includes that transformation in the model.

Assume two latent variables ${y^{\mathbf{o}}_{t,i}}^*$ and ${y^{\mathbf{s}}_{t,i}}^*$ where the time $t \in \{1, \ldots, T\}$ and the sale $i \in \{1, \ldots, N_t\}$. $y^{\mathbf{o}*}_t = ({y^{\mathbf{o}}_{t,1}}^*, \ldots, {y^{\mathbf{o}}_{t,N_t}}^*)^\top$ links the covariates of interest to the outcomes and $y^{\mathbf{s}*}_t = ({y^{\mathbf{s}}_{t,1}}^*, \ldots, {y^{\mathbf{s}}_{t,N_t}}^*)^\top$ is a vector of the selection propensity variables, as seen below:

$$y^{\mathbf{o}*}_t = {x^{\mathbf{o}}_t}^\top \beta^{\mathbf{o}}_t + \varepsilon^{\mathbf{o}}_t \tag{4.5}$$

$$y^{\mathbf{s}*}_t = {x^{\mathbf{s}}_t}^\top \beta^{\mathbf{s}}_t + \varepsilon^{\mathbf{s}}_t \tag{4.6}$$

where $x^{\mathbf{o}}_t$ and $x^{\mathbf{s}}_t$ are matrices of explanatory variables. Here, assume $x^{\mathbf{s}} \subseteq x^{\mathbf{o}}$, i.e. predictor variables using for the main outcomes of interest $y^{\mathbf{o}*}$ can be used to predict the selection $y^{\mathbf{s}*}$. $\varepsilon^{\mathbf{o}}_t$ and $\varepsilon^{\mathbf{s}}_t$ are error terms. $\beta^{\mathbf{o}}_t$ and $\beta^{\mathbf{s}}_t$ are coefficient parameters. $y^{\mathbf{o}}_{t,i}$ is only observed if the selection propensity variable is positive such as:

$$y^{\mathbf{s}}_{t,i} = \begin{cases} 1 & \text{if } {y^{\mathbf{s}}_{t,i}}^* > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.7}$$

Given selection to the main sample, the expected value of the outcome in Equation 4.5 is given by:

$$\mathsf{E}[y_t^{\mathbf{o}}|x_t^{\mathbf{s}}, \varepsilon_t^{\mathbf{s}}] = x_t^{\mathbf{o}\top}\beta_t^{\mathbf{o}} + \mathsf{E}[\varepsilon_t^{\mathbf{o}}|\varepsilon_t^{\mathbf{s}} \geq -x_t^{\mathbf{s}\top}\beta_t^{\mathbf{s}}] \tag{4.8}$$

Estimating the model above by OLS gives generally biased results, hence assume $\varepsilon_t^{\mathbf{o}}$ and $\varepsilon_t^{\mathbf{s}}$ are mean independent and follow a bivariate normal distribution, i.e. $\binom{\varepsilon^{\mathbf{o}}}{\varepsilon^{\mathbf{s}}} \sim \mathcal{N}\Big(\binom{0}{0}, \begin{pmatrix} 1 & \rho \\ \rho & \sigma^2 \end{pmatrix}\Big)$. Accordingly, we can write:

$$\begin{aligned} y_t^{\mathbf{o}} &= \beta_t^{\mathbf{o}} x_t^{\mathbf{o}\top} + \mathsf{E}[\varepsilon_t^{\mathbf{o}}|\varepsilon_t^{\mathbf{s}} \geq -x_t^{\mathbf{s}\top}\beta_t^{\mathbf{s}}] + u_t \\ &\overset{def}{=} \beta_t^{\mathbf{o}} x_t^{\mathbf{o}\top} + \rho\sigma\lambda(\beta_t^{\mathbf{s}} x_t^{\mathbf{s}\top}) + u_t \end{aligned} \tag{4.9}$$

where $\lambda(\cdot) = \varphi(\cdot)/\Phi(\cdot)$ is the inverse Mills ratio (IMR), being $\varphi(\cdot)$ the standard normal probability density function (pdf) and $\Phi(\cdot)$ the complementary cumulative distribution function (cdf). u_t is an error term which is independent from both $x_t^{\mathbf{s}}$ and $x_t^{\mathbf{o}}$. $\beta_t^{\mathbf{s}}$ is estimated via a probit model, assuming $\mathsf{P}(y_t^{\mathbf{s}} = 1|x_t^{\mathbf{s}}) = \Phi(x_t^{\mathbf{s}\top}\beta_t^{\mathbf{s}})$ and performing MLE, and thereafter, we access the estimates of $\lambda(\beta_t^{\mathbf{s}} x_t^{\mathbf{s}\top})$. Since $\rho \neq 0$ the unknown multiplicator $\rho\sigma$ can be estimated by OLS regressing $y_t^{\mathbf{o}}$ on $x_t^{\mathbf{o}}$ given the selection sample and $\lambda(\hat{\beta}_t^{\mathbf{s}} x_t^{\mathbf{s}\top})$. Further details on the consistency of such estimators are discussed in Miranda and Rabe-Hesketh (2006).

4.3 Huberization

Proceeding with the OLS regression, we observe that there exist many outliers in residuals $\hat{\varepsilon}_t$, see Figure 4.1. This is also visible in Figure 4.2 where the OLS residuals are displayed via box plots over time in comparison with number of transaction per day. The number of transactions of NFT artworks can induce the outlying observations in residuals; and the mean of residuals gets stablized by time.

In order to alleviate such an impact of outliers and offer a robust estimates of β, we introduce a correction procedure – Huberization to reflect the ac-

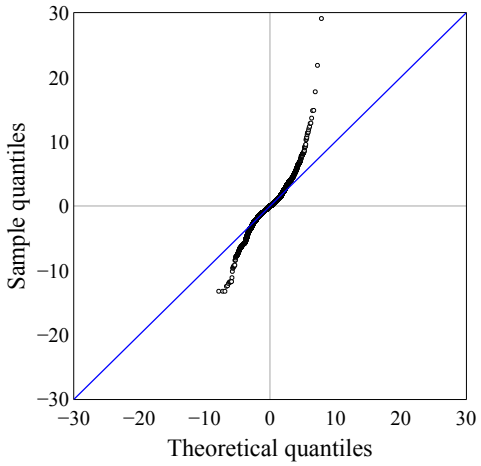


Figure 4.1: QQ plot for OLS standard residuals against standard normal distribution.

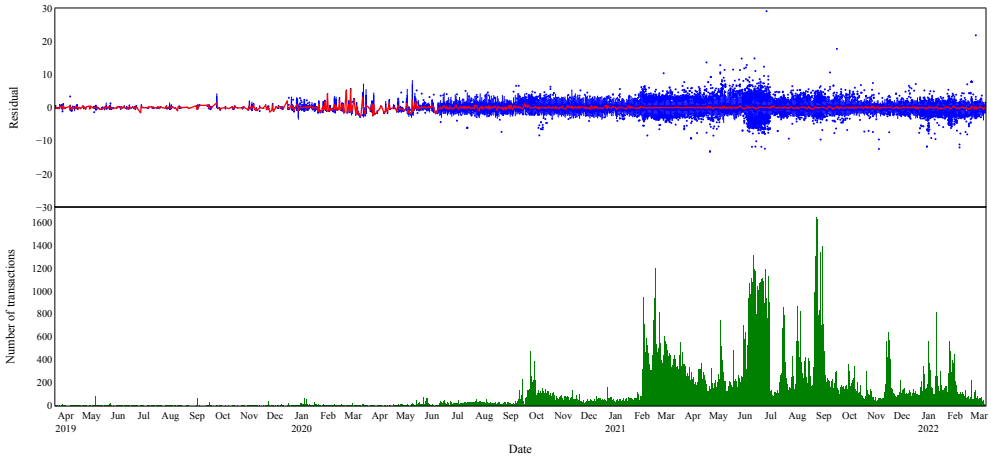


Figure 4.2: OLS residual box plots over time with mean of residuals and number of transactions per day.

tual outlyingness. Härdle (1984) uses an influence curve, Huber's ψ-function from Huber (1964), to bent down the influence of outlying observations. We construct pseudo observations $\tilde{y}_t = (\tilde{y}_t^1, \ldots, \tilde{y}_t^{N_t})^\top$ as a one-step correction of y_t:

$$\begin{aligned} \tilde{y}_t &= y_t + \tilde{\varepsilon}_t \\ &= d_t\beta_t + X_t^\top\alpha + \tilde{\varepsilon}_t \end{aligned} \tag{4.10}$$

where y_t can be estimated using $\hat{y}_t = d_t\hat{\beta}_t + X_t^\top\alpha$ from Equations 4.2 or 4.3. $\tilde{\varepsilon}_t = (\tilde{\varepsilon}_t^1, \ldots, \tilde{\varepsilon}_t^{N_t})^\top$ is a vector of the bounded one-step residuals:

$$\tilde{\varepsilon}_t^i = \frac{\psi(\varepsilon_t^i)}{\Delta} \tag{4.11}$$

where the scaling parameter $\Delta = \mathsf{E}[\psi'(\varepsilon_t^i)]$ is estimated via $\hat{\Delta} = T^{-1}\sum_{t=1}^{T}[n_t]^{-1}\sum_{i=1}^{n_t}[\psi'_\tau(\hat{\varepsilon}_t^i)]$ from OLS or lasso. The ψ-function is simply the derivative of the ρ-function.

$$\rho_\tau(\varepsilon_t^i) = \begin{cases} \frac{1}{2}(\varepsilon_t^i)^2 & \forall|\varepsilon_t^i| \leq \tau \\ \tau(|\varepsilon_t^i| - \frac{1}{2}\tau) & \text{otherwise} \end{cases} \tag{4.12}$$

where $\tau \in \mathbb{R}^+$ is a hyperparameter that controls the effect of outliers. The huberized estimation of β is obtained: $\hat{\beta}_t^* = \text{argmin}_{\beta_t\in\mathbb{R}}\left\{\|\tilde{y}_t - d_t\beta_t - X_t^\top\alpha\|_2^2\right\}$.

4.4 Kalman filtering

To efficiently obtain estimates of the unobserved components and to refine prior estimates, we treat the parameter β_t of Equation 4.1 in a state space form and apply Kalman filtering. Extending the classic random-effects models and imposing exogeneity of the regressors with respect to the time component such that $\beta_t \sim \mathcal{N}(0, \sigma_\beta^2)$ with $\mathsf{E}[\beta_t] = 0$, we can rewrite (4.1) as $y_t = X_t^\top\alpha + \eta_t$ with a composite error term $\eta_t = (\eta_t^1, \ldots, \eta_t^{N_t})^\top$ and

$$\eta_t = d_t\beta_t + \varepsilon_t \tag{4.13}$$

where $\varepsilon \overset{\text{iid}}{\sim} \mathcal{N}(0, {\sigma_\varepsilon}^2)$. σ_ε can be interpreted as idiosyncratic volatility such that it reflects the variation around the predicted price using the market index β_t and characteristics X_t.

Given the recursive nature and normality assumption of error terms, we fit an autoregressive order one process, AR(1), of β_t with a random walk:

$$\beta_t = \phi\beta_{t-1} + \xi_t \tag{4.14}$$

with $|\phi| \leq 1$ and $\xi \overset{\text{iid}}{\sim} \mathcal{N}(0, {\sigma_\xi}^2)$. As $\phi = 1$, ξ_t is the excess over the average price – interpreted as returns of a market portfolio – and σ_ξ can be hence seen as market volatility.

Given the linear Gaussian state space representation in 4.13 and 4.14, we now use Kalman filtering to access the underlying β_t via the parameter estimation of σ_ε and σ_ξ. We apply maximum likelihood estimation (MLE) to calibrate these parameters. Denote the unkown parameter vector by $\theta = \left(\sigma_\varepsilon, \sigma_\xi\right)$ within the parameter space $\Theta = \{\theta : \sigma_\xi^2 > 0, \sigma_u^2 > 0\}$. ϕ can be estimated jointly with σ_ε and σ_ξ while the AR(1) process of Equation 4.14 is stationary, i.e. $|\phi| < 1$. Here, we explicitly feature the unit root process in Equation 4.14 as $\phi = 1$.

Denote conditional mean and variance by $\eta_{t|t-1}$ and $\Sigma_\eta(t|t-1)$, respectively. Let $e_t(\theta) = \eta_t - \eta_{t|t-1}$ and $\Sigma_t(\theta) = \Sigma_\eta(t|t-1)$. We therefore define the log-likelihood function of θ as:

$$\mathcal{L}(\theta) = \frac{1}{2}\sum_{t=1}^{T}\Big\{\log|\Sigma_t(\theta)| + e_t(\theta)^\top\Sigma_t(\theta)^{-1}e_t(\theta)\Big\} \tag{4.15}$$

and the maximum likelihood estimator is:

$$\hat{\theta} = \underset{\theta\in\Theta=\mathbb{R}_+^2}{\operatorname{argmax}}\,\mathcal{L}(\theta) \tag{4.16}$$

Obtained $\hat{\theta}$ through MLE we can predict $\beta_{t|t-1}$ and $\eta_{t|t-1}$ and correct the prediction with $\sigma_\beta(t|t-1)$ and $\Sigma_\eta(t|t-1)$. Given the full sample information

$(t = 1, \ldots, T)$, we can therefore infer the underlying β_t.

4.5 DCS-t filtering

The Kalman filtering is optimal in a linear Gaussian state space model. In the presence of outliers, however, it may overweight the impact of these outliers on the changes in the index β_t. In consequence, we proposed to huberize residuals. Instead of Huberization, an alternative is based on dynamic conditional score (DCS) models introduced by Creal et al. (2011, 2013) and Harvey (2013). The idea is to replace the one step ahead prediction error in the updating equation of the Kalman filtering by a likelihood score with respect to the index. For heavy-tailed distributions this simply downweighs the impact of extremes. For example, consider a multivariate t-distribution with ν degrees of freedom. Assuming a random walk for β_t , the model becomes:

$$\begin{aligned} \eta_t &= d_t\beta_t + v_t, \quad v_t \sim t_\nu(0, \sigma_v^2 I_{n_t}) && (4.17)\\ \beta_t &= \beta_{t-1} + \kappa u_{t-1} && (4.18)\\ u_t &= \frac{1}{w_t}\frac{\nu + n_t}{\nu}\frac{1}{\sigma_v^2} d_t^\top e_t && (4.19)\\ w_t &= 1 + \frac{1}{\nu\sigma_v^2} e_t^\top e_t && (4.20) \end{aligned}$$

Here, u_t is the score w.r.t. β_t, i.e.

$$u_t = \frac{\partial \log f_t}{\partial \beta_t}$$

where f_t is the conditional density of η_t,

$$\log f_t(\eta_t; \theta) = \log\Gamma(\frac{\nu + n_t}{2}) - \log\Gamma(\frac{\nu}{2}) - \frac{n_t}{2}\log\Big(\pi\nu\sigma_v^2\Big) - \frac{\nu + n_t}{2}\log w_t.$$

Figure 4.3 depicts the score u_t as a function of e_t for the case $n_t = \sigma_v = 1$, $\nu = 4$ and $\nu = 10$. We see that the downweighting is stronger the heavier the

tails of the distribution. Note that in the Gaussian case ($\nu \to \infty$), the score reduces to $u_t = \frac{1}{\sigma_v^2} d_t^\top e_t$, which is analogous to the Kalman filtering, and in which case there is no downweighting of outliers. Estimation of the model

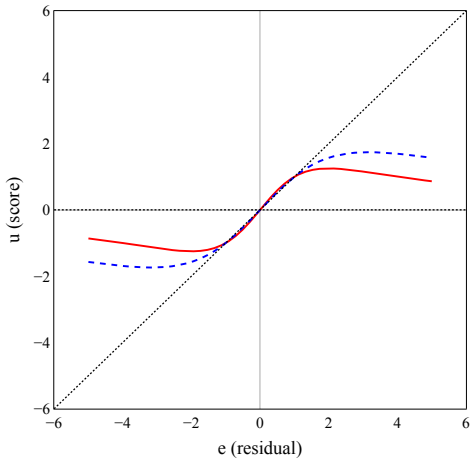


Figure 4.3: Score u_t with respect to β_t as a function of the one-step prediction error for a multivariate t distribution with $\nu = 4$ (solid) and $\nu = 10$ (dashed) degrees of freedom.

parameters is again performed by maximizing numerically the log likelihood function

$$\mathcal{L}(\theta) = \sum_{t=1}^{T} \log f_t(\eta_t; \theta)$$

with respect to $\theta = (\sigma_v^2, \kappa, \nu)^\top$.

5 Data description

Our data are collected using the API from OpenSea, which is the first and largest marketplace of NFTs. The data are also available at the Blockchain Research Center (BRC). To consider the representative constituents yet not explode in sample size, we select the top 10 NFT art collections ranked by the total transaction volume, accounting for 83.71% of the NFT art market

volume as of July 2021. The data comprise two data frames – NFT artwork metadata and transaction histories, which are then joined for the purpose of Hedonic regression, by the unique id of each work. Despite the continuous variables, i.e. number of sales and punk accessory, all the other variables are one-hot encoded. Table A.1 shows the attributes of each variable.

Artwork metadata describe the properties of each work, inclusive of the collection slug, contract type, schema name, creator, created year and traits (characteristics of artworks). Among which, creator ids and addresses are only available in collection MakersPlace and SuperRare because the rest are algorithm generated works. We summarize the creators into the number of their works have been listed, i.e. the only 1 piece, within 2 and 100 pieces, within 100 and 1,000 pieces, and more than 1000 pieces.

Collection specific properties are included as interactive dummies, for example, 'art_blocks_subcollection' in Art Blocks, 'bdap_Background' in Bored Ape Yacht Club. For the collections with too large set of words included in traits – e.g. accessories for CryptoPunks – and with unstructured textual descriptions – e.g. SuperRare, we adopt the natural language processing (NLP) techniques to avoid the explosion in feature space. CryptoPunks are in 5 types and each work has a set of accessories – such as headgears, glasses, and earrings – as of the total 92 categories. The traits within these categories are converted into the scarcity index using the concept of the TF-IDF (term frequency–inverse document frequency) in Jones (1972). The traits of the p-th punk's sale are denoted by T_p where $p \in (1, 2, \ldots, P)$, i.e. $P = 5466$. The TF of a word w in trait T_p is calculated as:

$$\mathrm{TF}(w, T_p) = \frac{f_{w,T_p}}{\sum_{w' \in T_i} f_{w',T_p}}$$

where f_{w,T_p} is the raw count of a word in a given trait.

The IDF of the word is:

$$\text{IDF}(w, T_p) = \log \frac{P}{1 + \sum_{p=1}^{P} \mathbf{I}(w, T_p)}.$$

The scarcity index of T_p is defined as the sum of the product of TF and IDF of all the words within a trait as:

$$\text{SI}_{T_p} = \sum_{w \in T_p} \text{TF}(w, T_p) * \text{IDF}(w, T_p).$$

Due to the consistent structure of CryptoPunks' traits, the scarcity of each punk can be measured by adding the scarcity of each accessory it equips. Here, SI_{T_p} of each accessory is the importance of an accessory for a given punk adjusted by the rarity of this accessory owned by all traded punks.

Regarding SuperRare, each artwork has a trait, or a set of heterogeneous tags; and the number of words within its trait varies from each other. Limited number of these words recurs but most of them only appear once. We, thus, incorporate the pre-trained word embedding model – Global vectors for word representation (GloVe) in Pennington et al. (2014) to handle such the inconsistency for SuperRare. Based on global word-word co-occurrence statistics from a corpus, the training result showcases linear substructures in word vector space. It implies that the similarity between words can be measured by the linear distance of two resulting word vectors, and those vectors are additive. Hence, we create GloVe index for each SuperRare trait T_s, $p \in (1, 2, \ldots, S)$,

$$\text{GI}_{T_s} = \frac{1}{f_{w,T_s}} \sum_{w \in T_s} \text{GloVe}_{w,T_s},$$

where f_{w,T_s} denotes the raw count of a word within a given trait; and GloVe_{w,T_s} denotes the output of the pre-trained GloVe model with the input of word w, trained on Wikipedia corpuses in a 50 dimensional vector. Taking the weighted sum of the GloVe vectors for each word, we obtain the GloVe index for each artwork. We can use this index to consider the simi-

larity among works. In order to further reduce the dimensionality, we apply k-means clustering on the GloVe index of each work and yield in dummies for 30 clusters.

Last, transaction histories of artworks among all the collections include the categorical variables (e.g. auction type, payment token, seller and winner addresses) and the continuous variables (e.g. number of historical sales). The categorical variables are one hot encoded and the continuous variables are aggregated by its frequency into different intervals.

6 Result

As pointed out previously, we observe the frequent presence of outliers and it significantly influences the stability and robustness of indices. In this paper we have proposed two alternative procedures – to resolve the overweighted impact from these outliers, as seen Figure 6.1. We first conduct the residual analysis to showcase how Huberization can assist to diminish such an impact. Diametrically, considering the evolution of conditional distribution of an observation via a score function, DCS-t filtering offers outlier insensitivity in a dynamic fashion.

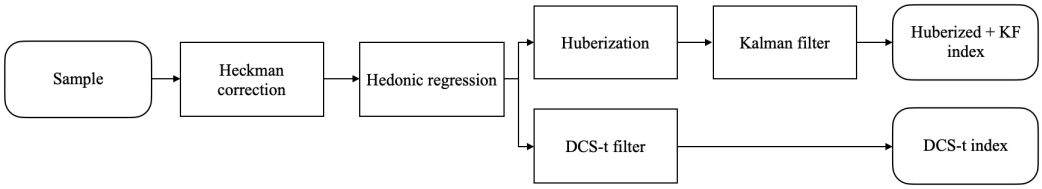
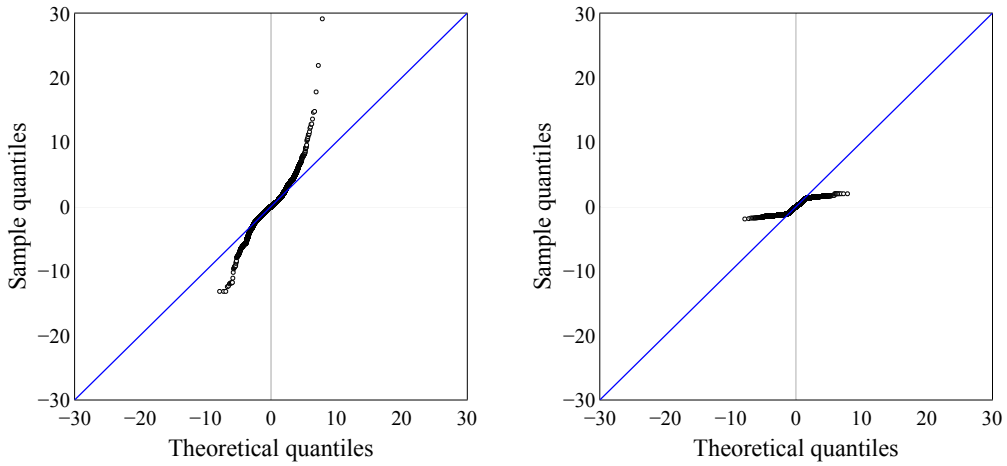


Figure 6.1: Two proposed procedures to handle the impact of outliers

6.1 Residual analysis

The OLS estimation of the DAI exhibits a fat-and heavy-tail behavior on residuals, see Figure 4.1. In order to achieve robustness and account for

the influence of outliers, Huberization is proposed in this paper as an one-step robust regression. Given $\tau = q_{0.1}(\hat{\varepsilon})$, after Huberization the scale of residuals has shrunk significantly in Figure 6.2b. We compare the evolution of residuals after Heckman correction and after Huberization with $\tau = q_{0.1}(\hat{\varepsilon})$ over time. Figure 6.3 illustrates that Huberization truncates the residuals and the influence of outlying observations is significantly downweighted.



(a) After Heckman correction **(b)** After Huberization with $\tau = q_{0.1}(\hat{\varepsilon})$

Figure 6.2: QQ plots for OLS standard residuals against standard normal distribution after Heckman correction and after Huberization.

Since τ here is a hyperparameter in Equation 4.12, we select three different quantile levels of residuals after Heckman correction – $q_{0.1}(\hat{\varepsilon})$, $q_{0.05}(\hat{\varepsilon})$, and $q_{0.01}(\hat{\varepsilon})$ – to examine their sensitiveness to outliers, being against standard normal distribution. In Figure 6.4, Huberizing with the three quantile levels, they all curve off in the extremities which indicates fat tails are still observed. Given $\tau = q_{0.01}(\hat{\varepsilon})$ within theoretical quantiles around $[-3, 3]$ $\tau = q_{0.01}(\hat{\varepsilon})$ it demonstrates a wider scope of fitting, however, in the right extremities it has a a spike where residuals aligned nearly horizontally. For $\tau = q_{0.1}(\hat{\varepsilon})$ and $\tau = q_{0.05}(\hat{\varepsilon})$, both have a slight jump on the right tail. All in all, $\tau = q_{0.01}(\hat{\varepsilon})$ resembles better the normal distribution compared with the others.

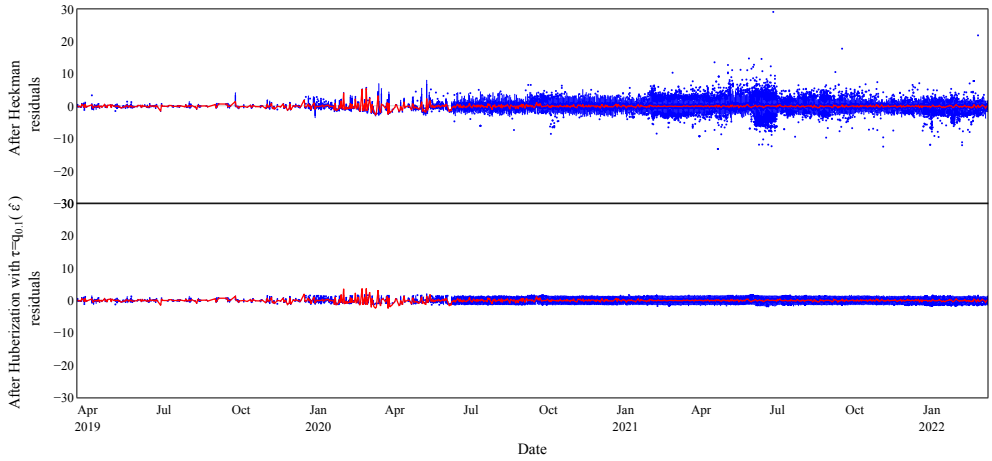


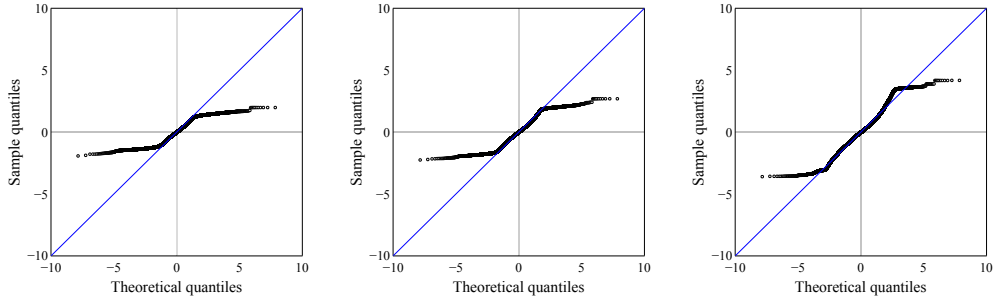
Figure 6.3: Residual box plots for after Heckman correction (upper panel) and after Huberization (lower panel) over time with mean of residuals.

As an alternative to using Huberization and Kalman filtering, we propose DCS-t filtering to adaptively accommodate the updates of conditional distribution. Considering the presence of outliers or heavy tailed distributions, one takes a multivariate t-distribution, as in Equation 4.17. Here, we plot the QQ plot for the OLS residuals after Heckman correction against the standard t-distribution with two settings of degrees of freedom ν in Figure 6.5. As discussed through Figure 4.3 previously, having heavier tail in the t-distribution (i.e. $nu = 4$) can be more robust than the normal distribution. However, we observe some gaps in the density in the extremities. The larger is the degree of freedom in Figure 6.5b, the more is it approaches to the case of normal distribution as in Figure 6.2b.

6.2 Digital art index

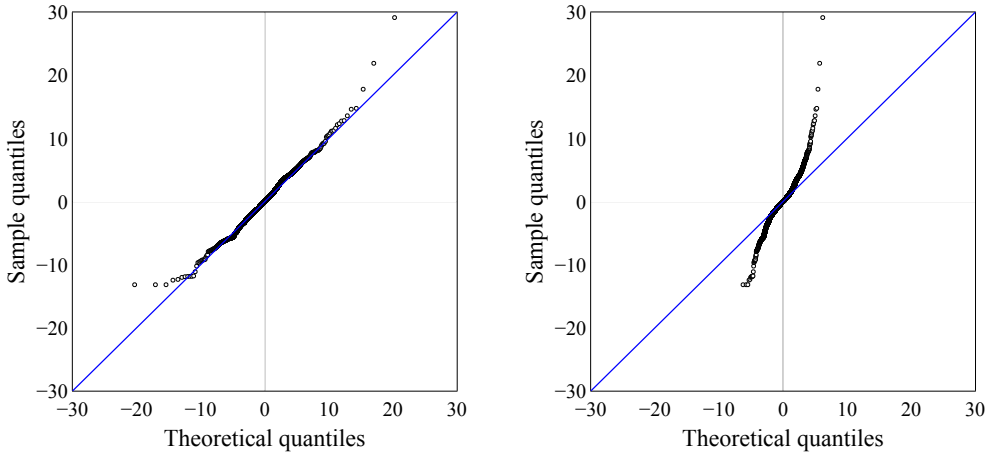
In this section we showcase the price indices computed by both proposed procedures in different parameter settings.

Figure 6.6 illustrates the indices using Huberization with different settings



(a) $\tau = q_{0.1}(\hat{\varepsilon})$ **(b)** $\tau = q_{0.05}(\hat{\varepsilon})$ **(c)** $\tau = q_{0.01}(\hat{\varepsilon})$

Figure 6.4: QQ plots for OLS residuals against standard normal distribution after Huberization with different quantile levels.



(a) $\nu = 4$ **(b)** $\nu = 10$

Figure 6.5: QQ plot for after Heckman against standard t-distribution with two different ν.

on τ and filtering. The Heckman corrected index is highly influenced by extremes, i.e. the time 2020 March to April exhibiting a relatively sharp spike. Huberization with $\tau = q_{0.1}(\hat{\varepsilon})$ is still slightly sensitive compared with the DAI lower on $\tau = q_{0.05}(\hat{\varepsilon})$, $q_{0.01}(\hat{\varepsilon})$. It is natural that with the rise of the sales in the NFT art market, see Figures 1.1 and 4.2, the DAI develops smoothly in its four variants.

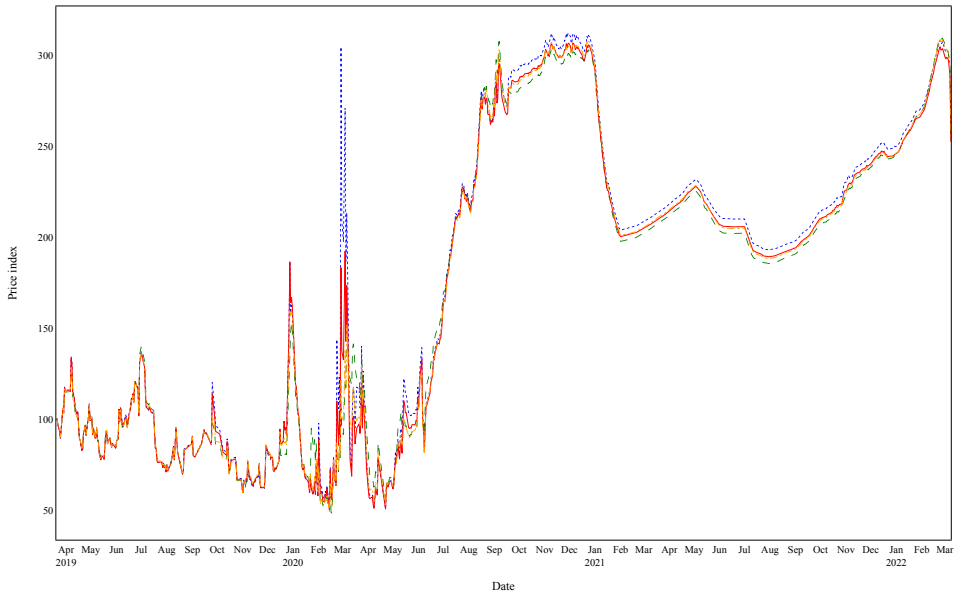


Figure 6.6: DAI – after Heckman correction (dotted), after Huberization with $\tau = q_{0.1}(\hat{\varepsilon})$ (dashed), after Huberization with $\tau = q_{0.05}(\hat{\varepsilon})$ (dash-dotted), and after Huberization with $\tau = q_{0.01}(\hat{\varepsilon})$ (solid).

Offering an alternative to Huberization and filtering, in Figure 6.7 we compare the DAI variants using DCS-t filtering given optimized parameters and infinite degree of freedom to huberizing with $\tau = q_{0.05}(\hat{\varepsilon}), q_{0.01}(\hat{\varepsilon})$. After the forth quarter of 2020 the indices applying DCS-t filtering are more reactive. Setting $\nu \sim \infty$, the t-distribution in Equation 4.17 approaches the normal distribution and it smooths the index.

As has been shown previously in Section 6.1, both procedures provide

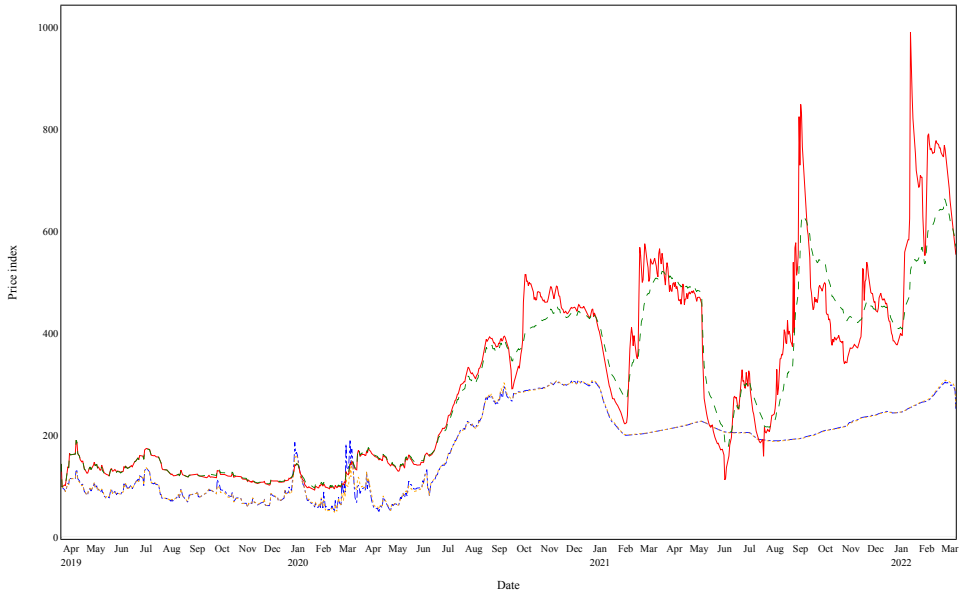


Figure 6.7: DAI variants – DCS-t filtering (solid), DCS-t filtering with $\nu \sim \infty$ (dashed), after Huberization with $\tau = q_{0.05}(\hat{\varepsilon})$ (dotted), and after Huberization with $\tau = q_{0.01}(\hat{\varepsilon})$ (dash-dotted).

robustness. Huberization in a one-step robust regression, crops off the influence from the outliers at once. After that, we use Kalman filtering to approach the unobserved components. Instead, DCS-t filtering fits the data with a heavy-tailed distribution and gradually adjusts by one step updates of conditional distributions. In this case the alternative procedures are not that alternative but rather complimentary. However, as the NFT art market is rapidly changing, including the dynamics of conditional distribution may be more adaptive and straightforward. In the following, we therefore only consider the DAI variant with DCS-t filtering.

7 The DAI

The works of NFT digital art are transacted by the means of cryptocurrencies, in particular ETH. In order to examine the dynamics between the DAI and other relevant crypto variables like the CRIX, ETH or trading volume (ETH vol.), we explore their causal relationships. We showcase the price determinants for collected NFT artworks, concluding the result from Lasso as a variant. Focusing on art markets, we discuss the DAI with the quarterly sales of – Post war, contemporary and ultra contemporary markets.

7.1 Causal inference

Evaluate as the returns, $(p_t^i - p_{t-1}^i)/p_{t-1}^i$ where $i \in \{\text{DAI, CRIX, ETH, ETH vol.}\}$. Given two time series $y = \{y_t\}_{t=1}^T$ and $x = \{x_t\}_{t=1}^T$, we test whether including one can help to predict the other in (any) unidirectional (i.e. $x \Rightarrow y$, $y \Rightarrow x$) or bidirectional (i.e.$x \Leftrightarrow y$) fashion. That is, in 7.1 and 7.2 $\eta_t > \varepsilon_t \succ x \Rightarrow y$. This implies $\sigma^2(y|\mathcal{U}') < \sigma^2(y|\mathcal{U}' - x')$ where $\mathcal{U}' \subset \mathcal{U}$ is a set of prior values of the all causative variables; and $x' \subset x$ is the prior values of time series x. We examine each pair with different number of time lags, τ. There is no evidence that CRIX, ETH, and ETH vol. significantly cause DAI, as well as vice versa. Consequently, NFT art market can possibly act as a financial instrument for risk diversification toward cryptos as their causal influence on DAI is restricted.

$$y_t = \sum_{j=1}^{\tau} a_j y_{t-j} + \eta_t \tag{7.1}$$

$$y_t = \sum_{j=1}^{\tau} a_j y_{t-j} + \sum_{j=1}^{\tau} b_j x_{t-j} + \varepsilon_t \tag{7.2}$$

Table 1: Pairwise Granger causality test results.

Number of lags τ	1		2		3		4	
Null hypothesis H_0	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value
CRIX $\not\Rightarrow$ DAI	1.7602	0.1851	2.1543	0.1169	1.6460	0.1777	1.2583	0.2853
DAI $\not\Rightarrow$CRIX	0.0320	0.8581	0.0386	0.9622	0.3657	0.7778	0.3520	0.8427
ETH $\not\Rightarrow$ DAI	1.0671	0.3020	1.0194	0.3615	1.1518	0.3276	0.9851	0.4151
DAI $\not\Rightarrow$ ETH	0.4160	0.5192	0.1414	0.8681	0.6584	0.5779	0.7233	0.5762
ETH vol. $\not\Rightarrow$DAI	0.0831	0.7733	0.1205	0.8865	0.5234	0.6663	0.4400	0.7797
DAI $\not\Rightarrow$ ETH vol.	0.2585	0.6113	1.1683	0.3117	0.9408	0.4206	0.7729	0.5431

7.2 Price determinant

We examine the regression results and investigate price determinants of NFT artworks by applying Hedonic Lasso regression to the pseudo observations $\tilde{y}_t$ from Huberization and the hedonic factors in Table A.1. Having the penalty parameter $\lambda = 0.001$ in Lasso, we show the active variables and their coefficients in Tables C.1 and C.2. As a variable receives a positive coefficient, it is marked in orange; otherwise, it is in blue. In order to better explain price determination among these works, we categorize the active variables into two parts – external factors and internal factors.

7.2.1 External factors

External factors are relevant to transaction information, at which it describes the trading details such as transaction time, inflow and outflow addresses, and number of sales occurred.

First, we see that the auction type as English auction has a positive effect on price. As English auction is via open-outcry and is ascending bids, it is more common and straightforward for sellers and buyers. In terms of seller experience, a seller by its address with less than 100 transaction records has the upmost influence on price. An inexperienced seller may still have an opportunity to have a high sale price. That is, the market is still

open to the new entrants. Conversely, buyers who have experienced between 100 and 500 sales have a significant negative impact on price; whereas less experienced ones may contribute to a higher deal price. Those artworks owned by experienced owners or professional collectors who have between 100 and 1000 transaction records are likely to be sold at higher prices; while the opposite is also valid. Consequently, this may be owing to the bidding skill and market sensibility of experienced buyers.

Moreover, we observe that trading with commonly used cryptos (i.e. ETH, WETH) and stable coins (i.e. DAI) may induce higher price. As NFT scheme ERC721 is the most common standard used in the market, it presents a positive impact here. The contract addresses can actually be referred to the launching platforms – i.e. 0x8c9f364bf7a56ed058fc63ef81c6cf09c833e656 links to the auction house, SuperRare. The influence, therefore, are relevant to the collection slugs which we discuss later in the internal factors. Last, looking at number of sales of works, we find it has positive influence but it is limited.

7.2.2 Internal factors

These factors are corresponding to the intrinsic values of artworks. They present i.e. at which collection or auction house a work has been created or published, by whom and when it is created, what visual properties it has.

One of our major findings from the collection slugs is that artworks (e.g. in Rarible, Marketplace) created by independent artists are, in general, priced lower than institutional initiated works (e.g. in Autoglyphs by Lava Lab). This result can correspond to Figure 1.2 at which the log price medians for Rarible and Marketplace are lower than the others. This may be explain why in the external factors of Figure C.1 frequency of creator addresses (i.e. number of works created) and creator id can have a negative impact on price as the institutional initiated works are more in favor. Interestingly, these institutional initiated works within a same collection tend to own a similar

visual composition; and they are commonly algorithm generated. Also, the finding might be against the pricing mechanism (Renneboog and Spaenjers, 2013) in the conventional art market where the reputation of an artist or the authenticity evidence such as the signature of the artist usually significantly raise the price. Instead, the price of an artwork is highly influenced by its publisher and collection. Also, we find that the works created during the last 3 years induce a lower price. This might be due to the oversupply and unfair distribution of artworks among collections, similarly to the tragedy of the commons in Hardin (1998).

Spotting the interactive variables that describe the properties of each work, we conclude them by collections as follows:

Art Blocks is a collection of demand-oriented generative works. Buyers can select their desired style and later receive a corresponding randomly generated work from the algorithm. The artwork can be a static image, 3D model, or an interactive anime. Consequently, the selection of style is crucial within this collection, see in art_blocks_subcollection of Table C.2. For instance, the styles in Archetypes and Ringers give a positive influence on price. In addition, the works from Art Blocks section – Playground where include curated artists' experimental projects impact positively In contrast, the works within Factory where has no curation and valuation process have a negative influence.

Bored Ape Yacht Club (BAYC) is a collection of 10,000 unique Bored Ape NFTs, as well as a membership of an online community which grants access to members-only benefits. For this collection, properties such as background, earring, fur, hat, eyes, clothes and mouth of an ape all matters on pricing of a work. Apes with bolder and more scarce features – i.e. trippy fur, king's crown, beam or laser eyes, black suit, grin multi-colored mouth – can raise the price. This result might relate to the buyers' preference for unique feature in order to boast within the online community.

Hashmasks is created by over 70 artists worldwide. The collection in-

cludes 16,384 digital portraits. Here, item of 'Shadow Monkey'; masks of 'Abstract', 'Animal', and 'Steampunk'; characters of 'Golden Robot', 'Mystical', 'Puppet'; skin color of 'Freak'; eye color of 'Heterochromatic'; backgrounds of 'Doodle', and 'Pixel' offer a positive impact on price. However, Hashmasks having an indistinguishable appearance or features (i.e. without any item, as male and female characters) lead to a lower price.

CryptoPunks is a collection with limited 10,000 unique collectible characters. It is one of the earliest adopter of an NFT with the ERC-721 standard. For this collection, the scarcity of attributes play a significant role in pricing. The variable 'punk_accessory' is based on the scarcity index for the accessories of punks (see Section 5) and has significantly positive impact on price. On the other hand, the most common punk type – 'Male' has a negative impact.

For **SuperRare**, we cluster these tags into groups based on their similarity (see Section 5) and we find that these tags have a significant influence on price. Last, none of the patterns used in **Autoglyphs** are significant in pricing within this collection. On a final note, we see that the scarcity of NFT artworks' internal factors is of importance on their price determination.

7.3 DAI and other art markets

In this section we intent to interpret the DAI and provide a narrative to understand its evolution through the emerging of the NFT art market. Applying Hedonic regression, we discount the implicit price given by the characteristics of an artwork; therefore, DAI, seen as a premium (characteristic-free price) of NFT artworks, can better represent the trend of the market. First, we examine the return of DAI and find that it becomes more volatile since 2021 as the growing in popularity of NFTs. In Figure 7.1 it presents a limited autocorrelation. However, as suggested by Goetzmann (1995) for the conventional art market, there should exist the possibility of persistent trends – which can be considered as a measure of market efficiency captured by

serial dependency in returns. That is, the NFT art market might be rather inefficient and immature and consequently it suffers from instantaneous uncertainty about the immediate resale value of a work of art. Such a tendency of the market produces a price risk for investors.

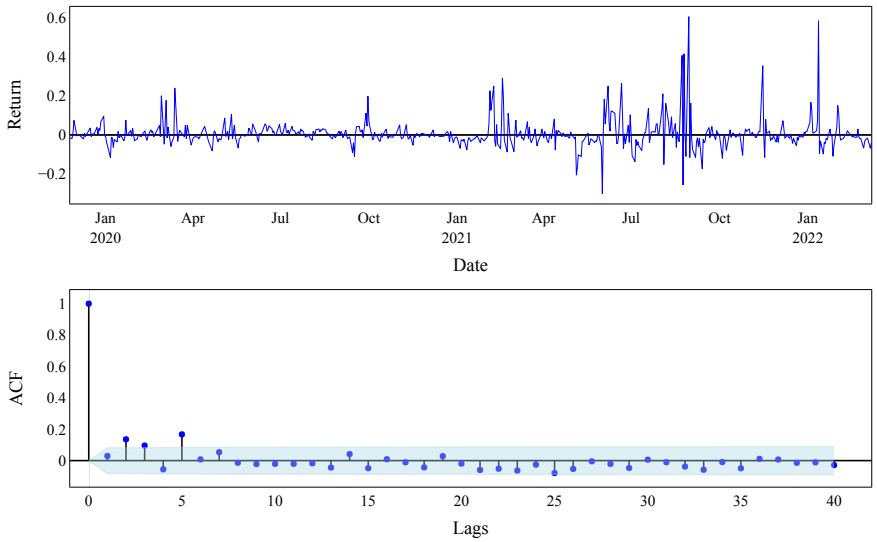


Figure 7.1: Returns and autocorrelation function for DAI with 95% CIs.

Due to the limitation of data availability for the other markets, we are not able to access their corresponding indices directly. Here, we illustrate the DAI with the sales shares of different art markets – post war, contemporary, ultra contemporary, and NFT art markets – in the background of DAI in Figure 7.2. In the beginning of the COVID-19 pandemic outbreak in 2020, all the art markets receive a recession; and these markets recover anon. As discussed by Mei and Moses (2002b), during an economic recession the declines on art prices offer an opportunity to investors and later these artworks might outperform the classic investment assets, i.e. stocks, equities, bonds.

The growing of NFT art sales brings up the DAI since the end of 2020. In the mid of 2021 it starts suppressing the other markets but surpassingly the DAI does not reflect this trend. The sales of NFT art market does not

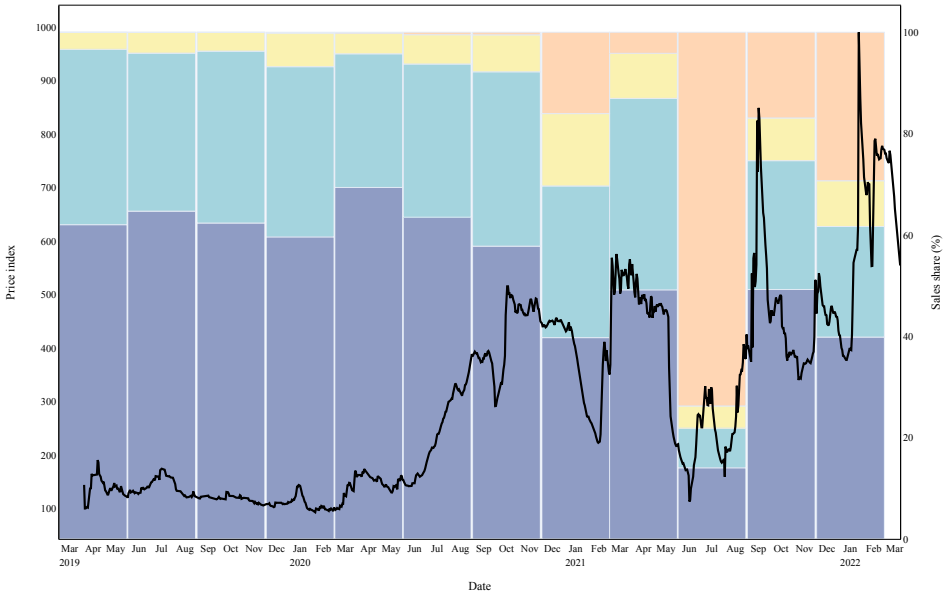


Figure 7.2: DAI (price index) and the sales shares of – **post war**, **contemporary**, **ultra contemporary**, **NFT** art markets (background).

fully equivalent to the performance of the market on price or return. That is, it is of essence to have the DAI that can reflect this market. After the mid of 2021, we can see several spikes on DAI and the sales of these four markets remain relatively steady. It is hard to distinguish how the NFT art would affect the other markets but yet we do not see it as substitute or complementary goods to the conventional art.

8 Conclusion

The discussion regarding NFT art often comes to the point – *What is art? and what is the definition for aesthetics here?* Experiencing the emerging of photography, Benjamin (1968) describes the unique existence of artworks within a limited time period at a specific place becomes an untenable category in the era of art reproduction. NFT artworks in this study are digitalized

are highly accessible through internet and replicable despite the ownership of tokens. While the aesthetics of digital artworks can be perfectly reproduced by copying and pasting, we might need to approach NFT art in a different manner compared to the tradition art. In particular, the NFT art market has really distinguishable consumers and market behaviors. This points out the essence of our study that we not only offer an index – DAI for such a unique market but also seek to highlight the transformation of the overall art industry.

In this study in order to construct a robust price index for heterogeneous goods, we include the time-variant and time-invariant characteristics of each work of art's through the Hedonic regression model. Avoiding the common issue for such a model: Selection bias, we apply Heckman correction to refine our samples. As the existence of outlying observations can significantly impact on the index, two alternative procedures – Huberization with Kaman filter and DCS-t filtering are proposed to robustify the results of Hedonic model. Consequently, we can see Huberization as an one-step robust regression that can be applied promptly and efficiently. Meanwhile, DCS-t filtering is a stepwise method which includes the changes of conditional distribution of observations in state space formulation of Kalman filtering. Both procedures can offer a degree of robustness to the indices but under the fast changing environment of NFT art market we consider the index from the DCS-t filtering more adaptive and straightforward to the case.

Through the causality tests we find that NFT art can be an option for risk diversification to cryptos. Also, the return of DAI presents limited autocorrelation which is different from the conventional art market. Looking into price determinants observed from Lasso, we conclude the possible price formulation with the external and internal factors. The external factors (i.e. payment tokens, frequency of seller addresses) influence much on price determination. For the internal factors the traits of NFT artworks have limited impact and collection slugs are relatively influential. In addition, institutional initiated collections (i.e. CryptoPunks by Lava Lab) tend to have

better pricing than the independent creators's works.

Regarding the role of NFT art as to whether it acts as a positive or negative stimulus to the other art markets, it is still undefinable due to the inefficiency and immaturity in the market. It requires a further follow-up study on its later development and application. As one of the first indices of such an emerging market, we do not aim to define its role within the art or finance contexts. Instead, DAI is an index helping us to foster understanding and witness the market evolution. For future studies, it can be meaningful to look at the NFT art from its visual perspective – e.g. visual autocorrelation (Özdilek, 2013) and visual link & knowledge discovery (Castellano et al., 2021) – and understand the value of art images towards pricing.

Acknowledgments

We thank the cloud computing infrastructure from Artnet.com.

In addition, this research is supported by the Deutsche Forschungsgemeinschaft via IRTG 1792 “High Dimensional Nonstationary Time Series”, Humboldt-Universität zu Berlin, the European Union’s Horizon 2020 research and innovation program “FIN-TECH: A Financial supervision and Technology compliance training programme”under the grant agreement no. 825215 (Topic: ICT-35-2018, Type of action: CSA), the European Cooperation in Science & Technology COST Action grant CA19130 - Fintech and Artificial Intelligence in Finance - Towards a transparent financial industry, the Yushan Scholar Program of Taiwan; and the Czech Science Foundation’s grant no. 19-28231X / CAS: XDA 23020303.

References

Allen, S., Juels, A., Khaire, M., Kell, T. and Shrivastava, S. (2022). NFTs for art and collectables: Primer and outlook.

Ante, L. (2021). The non-fungible token (NFT) market and its relationship with bitcoin and ethereum, *Available at SSRN 3861106* .

Bailey, M. J., Muth, R. F. and Nourse, H. O. (1963). A regression method for real estate price index construction, *Journal of the American Statistical Association* **58**(304): 933–942.

Beckert, J. and Rössel, J. (2013). The price of art: Uncertainty and reputation in the art field, *European societies* **15**(2): 178–195.

Benjamin, W. (1968). The work of art in the age of mechanical reproduction, 1936, *Illuminations: Essays and Reflections*, Houghton Mifflin Harcourt, pp. 217–252.

Bocart, F. Y., Ghysels, E. and Hafner, C. M. (2020). Monthly art market returns, *Journal of Risk and Financial Management* **13**(5): 100.

Bocart, F. Y. and Hafner, C. M. (2015). Volatility of price indices for heterogeneous goods with applications to the fine art market, *Journal of Applied Econometrics* **30**(2): 291–312.

Borri, N., Liu, Y. and Tsyvinski, A. (2022). The economics of non-fungible tokens, *Available at SSRN* .

Campbell, R. (2008). Art as a financial investment, *The Journal of alternative investments* **10**(4): 64–81.

Castellano, G., Lella, E. and Vessio, G. (2021). Visual link retrieval and knowledge discovery in painting datasets, *Multimedia Tools and Applications* **80**(5): 6599–6616.

Chainalysis (2022). The chainalysis 2021 NFT market report: Everything you need to know about the NFT market and its most successful collectors, *Report*.
URL: *https://go.chainalysis.com/nft-market-report.html*

Collins, A., Scorcu, A. and Zanola, R. (2009). Reconsidering hedonic art price indexes, *Economics Letters* **104**(2): 57–60.

Court, A. (1939). *Hedonic price indexes with automotive examples*.

Creal, D., Koopman, S. J. and Lucas, A. (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations, *Journal of Business & Economic Statistics* **29**(4): 552–563.

Creal, D., Koopman, S. J. and Lucas, A. (2013). Generalized autoregressive score models with applications, *Journal of Applied Econometrics* **28**(5): 777–795.

Dailey, N. (2022). NFTs ballooned to a $41 billion market in 2021 and are catching up to the total size of the global fine art market.

URL: *https://markets.businessinsider.com/news/currencies/nft-market-41-billion-nearing-fine-art-market-size-2022-1*

Dowling, M. (2022a). Fertile land: Pricing non-fungible tokens, *Finance Research Letters* **44**: 102096.

Dowling, M. (2022b). Is non-fungible token pricing driven by cryptocurrencies?, *Finance Research Letters* **44**: 102097.

Garay, U. (2021). Determinants of art prices and performance by movements: Long-run evidence from an emerging market, *Journal of Business Research* **127**: 413–426.

Ginsburgh, V., Mei, J. and Moses, M. (2006). The computation of prices indices, *Handbook of the Economics of Art and Culture* **1**: 947–979.

Goetzmann, W. N. (1995). The informational efficiency of the art market, *Managerial Finance* .

Goetzmann, W. N., Renneboog, L. and Spaenjers, C. (2011). Art and money, *American Economic Review* **101**(3): 222–26.

Goldberg, M., Kugler, P. and Schär, F. (2021). The economics of blockchain-based virtual worlds: A hedonic regression model for virtual land, *Available at SSRN 3932189* .

Griliches, Z. (1961). Hedonic price indexes for automobiles: An econometric of quality change.

Griliches, Z. (1971). Introduction: Hedonic price indexes revisited.

Hardin, G. (1998). Extensions of" the tragedy of the commons", *Science* **280**(5364): 682–683.

Härdle, W. (1984). Robust regression function estimation, *Journal of Multivariate Analysis* **14**(2): 169–180.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*, Vol. 52, Cambridge University Press.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models, *Annals of economic and social measurement, volume 5, number 4*, NBER, pp. 475–492.

Heckman, J. J. (1979). Sample selection bias as a specification error, *Econometrica: Journal of the econometric society* pp. 153–161.

Huber, P. J. (1964). Robust estimation of a location parameter, *The Annals of Mathematical Statistics* **35**(1): 73–101, 29.

Jones, K. S. (1972). A statistical interpretation of term specificity and its application in retrieval, *Journal of documentation* .

Koford, K. and Tschoegl, A. E. (1998). The market value of rarity, *Journal of Economic Behavior & Organization* **34**(3): 445–457.

Kräussl, R. and Tugnetti, A. (2022). Non-fungible tokens (NFTs): A review of pricing determinants, applications and opportunities, *Applications and Opportunities (May 17, 2022)* .

Le Fur, E. (2020). Dynamics of the global fine art market prices, *The Quarterly Review of Economics and Finance* **76**: 167–180.

Li, Y., Ma, M. X. and Renneboog, L. (2021). Pricing art and the art of pricing: On returns and risk in art auction markets, *European Financial Management* .

Malpezzi, S. et al. (2003). Hedonic pricing models: a selective and applied review, *Housing economics and public policy* **1**: 67–89.

McAndrew, C. (2010). *Fine art and high finance: expert advice on the economics of ownership*, Vol. 36, John Wiley & Sons.

McEachern, W. A. (2016). *Economics: A contemporary introduction*, Cengage Learning.

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

Mei, J. and Moses, M. (2002b). Economic activity and painting performance, *Technical report*, Working paper, Stern School of Business, New York University.

Mekacher, A., Bracci, A., Nadini, M., Martino, M., Alessandretti, L., Aiello, L. M. and Baronchelli, A. (2022). How rarity shapes the NFT market, *arXiv preprint arXiv:2204.10243* .

Miranda, A. and Rabe-Hesketh, S. (2006). Maximum likelihood estimation of endogenous switching and sample selection models for binary, ordinal, and count variables, *The stata journal* **6**(3): 285–308.

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

Özdilek, Ü. (2013). Visual autocorrelation of prices, *Journal of Cultural Economics* **37**(2): 203–223.

Pawelzik, L. and Thies, F. (2022). Selling digital art for millions-a qualitative analysis of NFT art marketplaces.

Penasse, J. and Renneboog, L. (2021). Speculative trading and bubbles: Evidence from the art market, *Management Science* .

Pennington, J., Socher, R. and Manning, C. D. (2014). Glove: Global vectors for word representation, *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543.

Rengers, M. and Velthuis, O. (2002). Determinants of prices for contemporary art in dutch galleries, 1992–1998, *Journal of cultural economics* **26**(1): 1–28.

Renneboog, L. and Spaenjers, C. (2013). Buying beauty: On prices and returns in the art market, *Management Science* **59**(1): 36–53.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition, *Journal of political economy* **82**(1): 34–55.

Shiller, R. J. (2008). Derivatives markets for home prices, *Technical report*, National Bureau of Economic Research.

Towse, R. and Hernández, T. N. (2020). *Handbook of cultural economics*, Edward Elgar Publishing.

Trimborn, S. and Härdle, W. K. (2018). CRIX an index for cryptocurrencies, *Journal of Empirical Finance* **49**: 107–122.

Urom, C., Ndubuisi, G. and Guesmi, K. (2022). Quantile return and volatility connectedness among non-fungible tokens (nfts) and (un)conventional assets, *MERIT Working Papers 2022-017*, United Nations University - Maastricht Economic and Social Research Institute on Innovation and Technology (MERIT).
URL: *https://EconPapers.repec.org/RePEc:unm:unumer:2022017*

Vasan, K., Janosov, M. and Barabási, A.-L. (2022). Quantifying NFT-driven networks in crypto art, *Scientific reports* **12**(1): 1–11.

Velthuis, O. (2003). Symbolic meanings of prices: Constructing the value of contemporary art in amsterdam and new york galleries, *Theory and society* **32**(2): 181–215.

Wallace, N. E. and Meese, R. A. (1997). The construction of residential housing price indices: a comparison of repeat-sales, hedonic-regression, and hybrid approaches, *The Journal of Real Estate Finance and Economics* **14**(1): 51–73.

Waugh, F. V. (1928). Quality factors influencing vegetable prices, *Journal of farm economics* **10**(2): 185–196.

Whitaker, A. and Kräussl, R. (2020). Fractional equity, blockchain, and the future of creative work, *Management Science* **66**(10): 4594–4611.

Wilkoff, S. and Yildiz, S. (2022). The behavior and determinants of illiquidity in the non-fungible tokens (NFTs) market, *Available at SSRN 4097200* .

Worthington, A. C. and Higgs, H. (2004). Art as an investment: Risk, return and portfolio diversification in major painting markets, *Accounting & Finance* **44**(2): 257–271.

Zinovyeva, E., Reule, R. C. and Härdle, W. K. (2022). Understanding smart contracts: Hype or hope?, *FinTech Research and Applications: Challenges and Opportunities*, World Scientific (Europe).

Zorloni, A. (2005). Structure of the contemporary art market and the profile of italian artists, *International Journal of Arts Management* pp. 61–71.

Appendices

A Hedonic factors

Table A.1: Data attributes.

Type of variables	auction_type	collection_slug	event_type	payment_tocken	to_account	seller_address	winner_address	contract_address
Variable	dutch english min_price	art-blocks autoglyphs beeple-everydays boredapeyachtclub cryptopunks hashmasks makersplace rarible superrare wrapped-cryptopunks	cancelled successful	ABST ASH ATRI DAI (crypto) ETH KLTR LUX NCT RARI TRSH USDC WETH WHALE nan	$(1000, \infty]$ $(100, 1000]$ $[2, 100]$ $= 1$	$(1000, \infty]$ $(500, 1000]$ (100,500] [2,100] $= 1$	$(500, \infty]$ $(100, 500]$ $[2, 100]$ $= 1$	0x0ba51d9c015a7544e3560081ceb16ffe222dd64f 0x131aebbfe55bca0c9eaad4ea24d386c5c082dd58 0x1f0678d1238921dc99309360668b16a17098aa2a 0x2947f98c42597966a0ec25e92843c09ac17fbaa7 0x41a322b28d0ff354040e2cbc676f0320d8c8850d 0x7a6425c9b3f5521bfa5d71df710a2fb80508319b 0x7be8076f4ea4a4ad08075c2508e481d6c946d12b 0x7e3abde9d9e80fa2d1a02c89e0eae91b233cde35 0x7f268357a8c2552623316e2562d90e642bb538e5 0x8c9f364bf7a56ed058fc63ef81c6cf09c833e656 0x93f2a75d771628856f37f256da95e99ea28aafbe 0xa5af48b105ddf2fa73cbaac61d420ea31b3c2a07 0xb47e3cd837ddf8e4c57f05d70ab865de6e193bbb 0xcd4ec7b66fbc029c116ba9ffb3e59351c20b5b06 0xf2ee97405593bc7b6275682b0331169a48fedec7 nan

Type of variables	asset_contract_type	schema_name	owner_address	creator_adress	art_blocks_section
Variable	non-fungible semi-fungible	CRYPTOPUNKS ERC1155 ERC721	$(1000, \infty]$ (500,1000] (100,500] [2,100] $= 1$	$(2000, \infty]$ (1000,2000] (100,1000] [2,100] $= 1$	curated factory nan playground

Continued on next page

Table A.1: Data attributes (cont'd).

Type of variables	art_blocks_subcollection	autoglyphs_trait	bdap_Background	bdap_Earring	bdap_Fur	bdap_Hat
Variable	All 27-Bit Digitals	Symbol Scheme: +-|	Aquamarine	Cross	Black	Army Hat
	All 720 Minutes	Symbol Scheme: /\	Army Green	Diamond Stud	Blue	Baby's Bonnet
	All Aerial Views	Symbol Scheme: O	Blue	Gold Hoop	Brown	Bandana Blue
	All AlgoRhythms	Symbol Scheme: O|-	Gray	Gold Stud	Cheetah	Bayc Flipped Brim
	All Algobots	Symbol Scheme: X/\	New Punk Blue	Silver Hoop	Cream	Bayc Hat Black
	All Apparitions	Symbol Scheme: [ ]	Orange	Silver Stud	Dark Brown	Bayc Hat Red
	All Archetypes	Symbol Scheme: [ ]O	Purple		Death Bot	Beanie
	All Bubble Blobbys	Symbol Scheme: [ ]|-+	Yellow		Dmt	Bowler
	All CENTURYs	Symbol Scheme: \			Golden Brown	Bunny Ears
	All Chromie Squiggles	Symbol Scheme: -/			Gray	Commie Hat
	All Construction Tokens				Noise	Cowboy Hat
	All Cryptoblots				Pink	Faux Hawk
	All Dreams				Red	Fez
	All Dynamic Slices				Robot	Fisherman's Hat
	All Elementals				Solid Gold	Girl's Hair Pink
	All Elevated Deconstructions				Tan	Girl's Hair Short
	All Fidenzas				Trippy	Halo
	All Frammentis				White	Horns
	All Genesis				Zombie	Irish Boho
	All HyperHashs					King's Crown
	All Ignitions					Laurel Wreath
	All Inspirals					Party Hat 1
	All NimBuds					Party Hat 2
	All Ringers					Police Motorcycle Helmet
	All Singularitys					Prussian Helmet
	All Spectrons					Sm Hat
	All Subscapes					Safari
	All Synapses					Sea Captain's Hat
	All The Blocks of Arts					Seaman's Hat
	All Unigrids					Short Mohawk
	All Watercolor Dreams					Spinner Hat
	nan					Stuntman Helmet
						Sushi Chef Headband
						Trippy Captain's Hat
						Vietnam Era Helmet
						Ww2 Pilot Helm

Continued on next page

Table A.1: Data attributes (cont'd).

Types of variables	bdap_Eyes	bdap_Clothes	bdap_Mouth	hash_Item	hash_Mask	hash_Character
	3d	Admirals Coat	Bored	Book	Abstract	Female
	Angry	Bandolier	Bored Bubblegum	Bottle	African	Golden Robot
	Blindfold	Bayc T Black	Bored Cigar	Golden Toilet Paper	Animal	Male
	Bloodshot	Bayc T Red	Bored Cigarette	Mirror	Aztec	Mystical
	Blue Beams	Biker Vest	Bored Dagger	No Item	Basic	Puppet
	Bored	Black Holes T	Bored Kazoo	Shadow Monkey	Chinese	Robot
	Closed	Black Suit	Bored Party Horn	Toilet Paper	Crayon	
	Coins	Black T	Bored Pipe		Doodle	
	Crazy	Blue Dress	Bored Pizza		Hawaiian	
	Cyborg	Bone Necklace	Bored Unshaven		Indian	
	Eyepatch	Bone Tee	Bored Unshaven Bubblegum		Mexican	
	Heart	Caveman Pelt	Bored Unshaven Cigar		Pixel	
	Holographic	Cowboy Shirt	Bored Unshaven Cigarette		Steampunk	
	Hypnotized	Guayabera	Bored Unshaven Dagger		Street	
	Laser Eyes	Hawaiian	Bored Unshaven Kazoo		Unique	
	Robot	Hip Hop	Bored Unshaven Party horn		Unmasked	
	Sad	Kings Robe	Bored Unshaven Pipe			
	Scumbag	Lab Coat	Bored Unshaven Pizza			
	Sleepy	Leather Jacket	Discomfort			
	Sunglasses	Leather Punk Jacket	Dumbfounded			
	Wide Eyed	Lumberjack Shirt	Grin			
Variable	X Eyes	Navy Striped Tee	Grin Diamond Grill			
	Zombie	Pimp Coat	Grin Gold Grill			
		Prison Jumpsuit	Grin Multicolored			
		Prom Dress	Jovial			
		Puffy Vest	Phoneme ooo			
		Rainbow Suspenders	Phoneme L			
		Sailor Shirt	Phoneme Oh			
		Service	Phoneme Vuh			
		Sleeveless Logo T	Phoneme Wah			
		Sleeveless T	Rage			
		Smoking Jacket	Small Grin			
		Space Suit	Tongue Out			
		Striped Tee				
		Stunt Jacket				
		Tanktop				
		Tie Dye				
		Toga				
		Tuxedo Tee				
		Tweed Suit				
		Vietnam Jacket				
		Wool Turtleneck				
		Work Vest				

Continued on next page

Table A.1: Data attributes (cont'd).

Types of variables	hash_Skin Color	hash_Eye Color	hash_Background	hash_Glyph	creators	super_tag	punk_type	created_year
	Blue	Blue	Book	Egyptian Hieroglyph	$(1000, \infty]$	cluster_0	Alien	2018
	Dark	Dark	Davinci	Greek Symbol	$(100, 1000]$	cluster_1	Ape	2019
	Freak	Freak	Doodle	Mannequin	$[2, 100]$	cluster_10	Female	2020
	Gold	Glass	Expressionist	Planetary	$= 1$	cluster_11	Male	2021
	Gray	Green	Mystery Night	Zodiac Sign		cluster_12	Zombie	
	Light	Heterochromatic	Pixel			cluster_13		
	Mystical	Mystical	Street Art			cluster_14		
	Steel	Painted	Waves			cluster_15		
	Transparent					cluster_16		
	Wood					cluster_17		
						cluster_18		
						cluster_19		
						cluster_2		
						cluster_20		
Variable						cluster_21		
						cluster_22		
						cluster_23		
						cluster_24		
						cluster_25		
						cluster_26		
						cluster_27		
						cluster_28		
						cluster_29		
						cluster_3		
						cluster_4		
						cluster_5		
						cluster_6		
						cluster_7		
						cluster_8		
						cluster_9		

B Price determinants

Table C.1: External factors.

Type of variables	Variable	Coeff.
auction_type	english	0.5364
event_type	successful	-1.2119
payment_tocken	DAI (crypto)	4.5703
	ETH	1.6079
	WETH	1.3497
frequency of 'seller_address'	(500, 1000]	0.3941
	(100, 500]	0.8425
	[2, 100]	1.0586
	= 1	0.8854
frequency of 'winner_address'	(100, 500]	-0.1205
	= 1	0.2120
contract_address	0x2947f98c42597966a0ec25e92843c09ac17fbaa7	0.3141
	0x7be8076f4ea4a4ad08075c2508e481d6c946d12b	-0.2917
	0x7e3abde9d9e80fa2d1a02c89e0eae91b233cde35	0.3086
	0x8c9f364bf7a56ed058fc63ef81c6cf09c833e656	0.5100
	0xcd4ec7b66fbc029c116ba9ffb3e59351c20b5b06	0.0714
schema_name	ERC721	0.6775
frequency of 'owner_address'	(500, 1000]	0.5347
	(100, 500]	0.2085
	=1	-0.3033

Continued on next page

Table C.1: External factors (cont'd)

Type of variables	Variable	Coeff.
	(1000, 2000]	-0.7539
frequency of	(100, 1000]	-1.1146
'creator_address'	[2, 100]	-1.1746
	=1	-0.6828
frequency of 'creator_id'	(100, 1000]	-0.1855
num_sales		0.0023

Table C.2: Internal factors.

Type of variables	Variable	Coeff.
collection_slug	autoglyphs	2.9271
	beeple-everydays	2.4190
	boredapeyachtclub	1.2886
	cryptopunks	1.4028
	hashmasks	0.4400
	makersplace	-0.7214
	rarible	-2.8638
	superrare	0.1091
	wrapped-cryptopunks	1.6581
art_blocks_section	factory	-1.3547
	playground	0.1033

Continued on next page

Table C.2: Internal factors (cont'd).

Type of variables	Variable	Coeff.
art_blocks_subcollection	All 27-Bit Digitals	-1.8213
	All 720 Minutes	-0.6192
	All Aerial Views	-3.0553
	All AlgoRhythms	-1.2175
	All Algobots	-0.1761
	All Apparitions	-2.2533
	All Archetypes	0.3477
	All Bubble Blobbys	-1.2125
	All CENTURYs	-1.2927
	All Chromie Squiggles	-1.6814
	All Cryptoblots	-2.3973
	All Dreams	-1.1674
	All Elementals	-1.1412
	All Fidenzas	-0.3117
	All Frammentis	-1.0715
	All Inspirals	-1.5496
	All Ringers	1.0678
	All Subscapes	-0.4786
	All Synapses	-1.4979
	All The Blocks of Arts	-1.6330
	All Watercolor Dreams	-2.1511

Continued on next page

Table C.2: Internal factors (cont'd).

Type of variables	Variable	Coeff.
bdap_Background	Aquamarine	0.0337
	Gray	0.0022
bdap_Earring	Cross	0.0136
	Gold Stud	0.0142
	Silver Stud	0.0305
bdap_Fur	Black	-0.0437
	Brown	-0.0451
	Cheetah	0.0293
	Cream	-0.0210
	Dark Brown	-0.0066
	Death Bot	0.1723
	Solid Gold	0.2578
	Trippy	0.8243
bdap_Hat	Bayc Hat Black	0.0133
	Bayc Hat Red	0.0009
	Fisherman's Hat	-0.0586
	King's Crown	0.4068
	Seaman's Hat	-0.0071
	Trippy Captain's Hat	0.0239

Continued on next page

Table C.2: Internal factors (cont'd).

Type of variables	Variable	Coeff.
bdap_Eyes	3d	0.0133
	Blue Beams	0.3318
	Closed	-0.1473
	Crazy	-0.0127
	Laser Eyes	0.2793
	Sleepy	-0.0187
	Wide Eyed	-0.0647
bdap_Clothes	Black Suit	0.2845
	Black T	-0.0118
	Guayabera	-0.0083
	Hip Hop	0.0042
	Pimp Coat	0.0093
	Tie Dye	0.0450
bdap_Mouth	Bored	-0.0795
	Bored Cigarette	-0.0063
	Bored Unshaven	-0.1011
	Bored Unshaven Cigarette	-0.0240
	Dumbfounded	-0.0428
	Grin	-0.0147
	Grin Multicolored	0.0836

Continued on next page

Table C.2: Internal factors (cont'd).

Type of variables	Variable	Coeff.
hash_Item	Golden Toilet Paper	0.2682
	No Item	-0.6320
	Shadow Monkey	0.1354
hash_Mask	Abstract	0.0270
	Animal	0.0388
	Basic	0.0156
	Doodle	-0.0598
	Indian	-0.0169
	Steampunk	0.0755
hash_Character	Female	-0.0339
	Golden Robot	0.9744
	Male	-0.0508
	Mystical	1.0136
	Puppet	0.1841
hash_Skin Color	Dark	-0.0212
	Freak	0.1429
hash_Eye Color	Blue	-0.3567
	Dark	-0.4657
	Green	-0.3512
	Heterochromatic	0.0739

Continued on next page

Table C.2: Internal factors (cont'd).

Type of variables	Variable	Coeff.
hash_Background	Doodle	-0.0065
	Pixel	0.0492
super_tag	cluster_10	0.0369
	cluster_18	-0.3642
	cluster_2	-0.0648
	cluster_22	0.0207
	cluster_23	0.0113
punk_type	Male	-0.1262
created_year	2019	-2.8126
	2020	-1.5785
	2021	-1.0743
punk_accessory		0.0011