# 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 Karl Härdle

REPRINT | 2026 / 02

UCLouvain

LIDAM
ISBA | 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

# 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]

October 24, 2025

**Abstract**

The market of non-fungible tokens (NFTs), driven by blockchain and smart contracts, provides both artists and art collectors an unprecedented marketplace with more security, flexibility, publicity, and freedom to monetize. Yet, the emergence of such a market has been considered to be packed with speculation and economic uncertainty, given the limited understanding towards this market. To provide a precise depiction of the NFT art market and gauge market volatility, we construct the Digital Art Index (DAI), a novel price index using hedonic regression on the top 10 liquid NFT art collections (as of 2023). Addressing artwork price inequality, which often disrupts the price discovery process, this paper introduces two innovative alternative methods: Huberization and score-based filtering. These methods effectively mitigate the influence of outliers, particularly in an emerging market with limited accessible observations. In conclusion, the NFT art market presents significant opportunities for large gains, which are often favored by risk-takers, but also carries the potential for significant losses. Its pricing is necessarily determined by institutional creators and platforms, meaning that solo artists may not benefit significantly in the current market environment.

**Keywords:** NFT, digital art, Hedonic regression, outliers, index construction, robustnes, Huberization, DCS-t filtering
**JEL Classification:** C14, C43, C51, G10, Z11

[1]International Research Training Group 1792, School of Business and Economics, Humboldt-Universität zu Berlin. Unter den Linden 6, 10099 Berlin, Germany; IDA Institute Digital Assets, Bucharest University of Economic Studies, Romania. Email: linminbin0593@gmail.com

[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; National Yang Ming Chiao Tung University, Taiwan; IDA Institute Digital Assets, Bucharest University of Economic Studies, Romania. Email: haerdle@hu-berlin.de

# 1 Introduction

Non-fungible tokens (NFT) offer vast and innovative possibilities for art trading. An NFT is – *a cryptographically secured and unique digital identifier minted through smart contracts on a blockchain that cannot be replicated, subdivided or substituted.* In contrast to cryptocurrency or utility tokens, NFT is a distinct, non-interchangeable token that facilitates asset tokenization while guaranteeing verifiable ownership and authenticity. It can be used to represent both tangible and intangible assets and through a programmable smart contract on artwork is traded directly between creators and buyers. Kraizberg (2023) further concludes that NFTs could potentially substitute the reliance on legal protection provided by intellectual property rights (IPRs), but legal frameworks and public perception must be aligned. On the other hand, in the conventional art market artists and creators are generally required to seek for a third party, (i.e. a dealer or gallery) to access potential buyers (Zorloni, 2005).

The NFT market capitalization rose up to US$ 41 billion in 2021, a 30 % growth by the previous year; meanwhile, the conventional art market shrank by 22 % in sales from 2019 (Dailey, 2022; Chainalysis, 2022). A preliminary comparison on quarterly sales of four different art markets worldwide is illustrated in Figure 1.1, sourced from Artnet.com. At the time of writing this research, the NFT art market is in a developmental stage and exhibits strong bubble-like behaviors, with a lack of complete comprehension. After substantial growth in 2021, the market experienced a significant downturn in the summer of 2022.

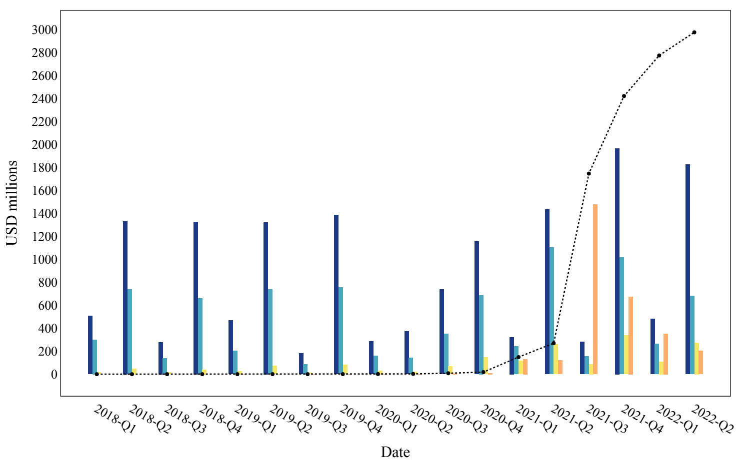



**Figure 1.1: Quarterly sales in USD millions of post war, contemporary, ultra contemporary, NFT art markets**. The dotted line is the cumulative quarterly sales on NFT.

This paper constructs a price index to depict the overall trend of the NFT art market and offer valuable insights into its risk and return potential, while also examining the determinants that influence the price movement of the artworks. The complexity of the art market, characterized by high heterogeneity among artworks and infrequent trading behaviors (i.e., unbalanced panel data), poses challenges in determining changes in market value when constructing an art price index, unlike stock or asset indices based on identical share prices (Ginsburgh et al., 2006; Beckert and Rössel, 2013). Chanel et al. (1996) further shows that relying on resales does not accurately reflect the market; instead, one should regress on the full set of sales. According to McAndrew (2010), the essence of art indices lies in their representativeness, liquidity, and capacity, which collectively determine the potential and values of art sales. Using an average (arithmetic mean or median) price index may be insufficient in capturing the intricacies of the data. Hedonic regression (HR) Rosen (1974), which considers the inherent attributes and characteristics of each artwork, is widely applied in this case. It helps identify the factors influencing pricing and provides a characteristic-adjusted price index.

However, price inequality, as evidenced by rather frequent outliers in NFT art pricing shown in Figure 1.2, frequently hinders the fair price discovery process in the market. This may be exemplified by CryptoPunk 9998's US$532M sale on October 28, 2021, which caused notable surges in the overall transaction volume (see Figure 2.1) that consequently may skew price indexes. To enhance the indices' resilience and mitigate the impact of outliers, we propose two innovative methods for index smoothing: Kalman filtering (KF) with robustification à la Huber (1964), and dynamic conditional score filtering using the student-$t$ distribution (DCS-$t$), see e.g. Harvey (2013). Through empirical analysis and robustness checks, both methods show their capacity to impart a certain level of robustness to indices.

The relevant literature includes Borri et al. (2022), who construct a market index using repeat-sales regression (RSR) and compares it to conventional assets and cryptocurrencies, and Horky et al. (2022), who employ hedonic regression for index construction and uses LASSO (least absolute shrinkage and selection operator) for identifying price determinants, similar to our study. However, neither of these papers considers how the heterogeneity of artwork influences pricing or accounts for the impact of price inequality. The characteristics considered in these studies are limited to metadata (e.g. cryptocurrencies used for trading, media formats) and do not extend to the artworks' stylistic properties (i.e. traits). As a result, their explanation of price determinants may lack resolution. In this paper, we incorporate NFT artwork metadata and historical prices to categorize each piece based on its time-variant characteristics (e.g., sales, owner address) and time-invariant characteristics (e.g., collection name, traits, NFT scheme). To address the robustness of the index in hedonic regression, we introduce the Digital Art

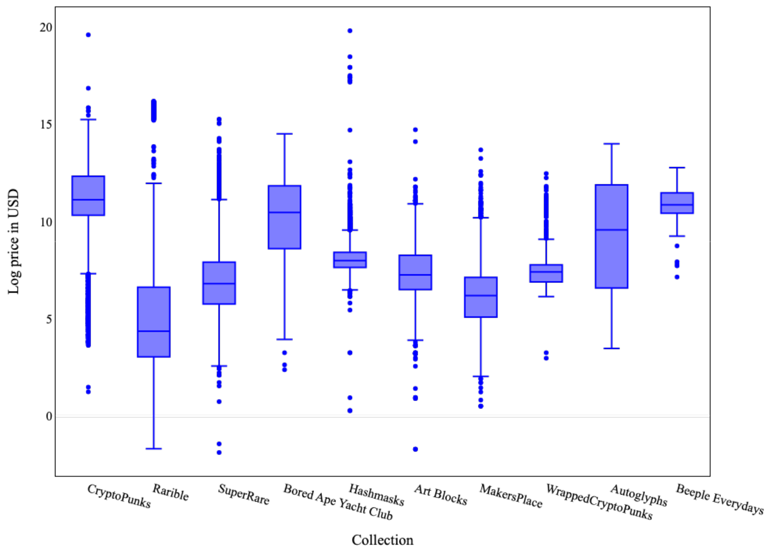



**Figure 1.2: Boxplots of log prices in USD on ten NFT digital art marketplaces. The data covers the period 4/1/2019 to 3/31/2022 and is described in Section 2.3.**

Index (DAI), which of course has to be distinguished from the stablecoin with the same acronym. Using LASSO, we identify the price determinants for sampled works, particularly focusing on the special traits highlighted by the creators.

The paper proceeds as follows. Section 2 portrays the selling processes for NFT and conventional art markets, highlights their stylized facts, and then describes the data attributes and data processing used in the paper. Section 3 elaborates the methodologies used in this study. We demonstrate the result in Section 4 and offer a discussion for interpretation in Section 5. Finally, Section 6 concludes our findings.

# 2 The art markets and the data

Art has been considered as one of the investment instruments that offers lower volatility and lower correlation with other assets, thus being used as an alternative asset for portfolio diversification, see Mei and Moses (2002a); Worthington and Higgs (2004); Campbell (2008). In the following we give an overview of conventional and NFT art markets, both from an economic and organisational point of view, and then describe the data we use in this paper.

## 2.1 The conventional art market

In a conventional art market, creators depend on a central dealer/gallery for the selling process, with the listing price decided through valuation and negotiation. Subsequently, the artwork is traded through various independent channels, typically lacking feedback relationships (Le Fur, 2020). The transaction history and pricing are often opaque, making resale royalty rights challenging.

For investments in conventional art markets, Bocart et al. (2020); Whitaker and Kräussl (2020) show that contemporary art and gallery portfolios 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) find that the art market is subject to frequent booms and busts, so that the behaviour of market agents can hardly be explained by rational consumer models. Goetzmann et al. (2011); Renneboog and Spaenjers (2013) direct attention to income inequality and art market sentiment and show their correlation with price trends. A wide body of literature analyzes 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). To summarize, the main stylized facts of the conventional art market include infrequent trading, price inequality, illiquidity, centralized patrons, limited transparency in price formulation, and the presence of multiple independent marketplaces.

## 2.2 The NFT art market

In the NFT art market, a creator can be either institutional (e.g. Larva Labs, Dapper Labs) or independent. The majority of artworks are traded in the secondary market through marketplaces (e.g., Opensea, Rarible), except for those from institutes supporting the architecture of a blockchain. Art creators usually retain copyright for their artworks, enabling them to produce a series of similar works. Additionally, they have the autonomy to set the listing price, which is the initial sale price, and maintain control over the bidding process with buyers. With NFTs' programmability, creators can enforce a royalty, typically 5-10% of the sale price, at a later stage. The crypto payment and NFT transfer occur automatically and simultaneously in a single transaction between two mutually distrusting parties on a blockchain network.

Transactions in NFT are usually settled in cryptocurrencies (CC) such as Ethereum (ETH) and Solana (SOL). Consequently, the evolution of CC influences NFT asset prices as well as their volatility, see Dowling (2022). Additionally, Urom et al. (2022) discovered that the connectedness between NFT markets and conventional markets such as stocks or commodities is higher during extreme market conditions. Furthermore, Figure 2.1 shows a clear positive correlation between the total NFT art transaction volume and the CRIX index proposed by Trimborn and Härdle (2018), an index for tracking the evolution of the CC market.

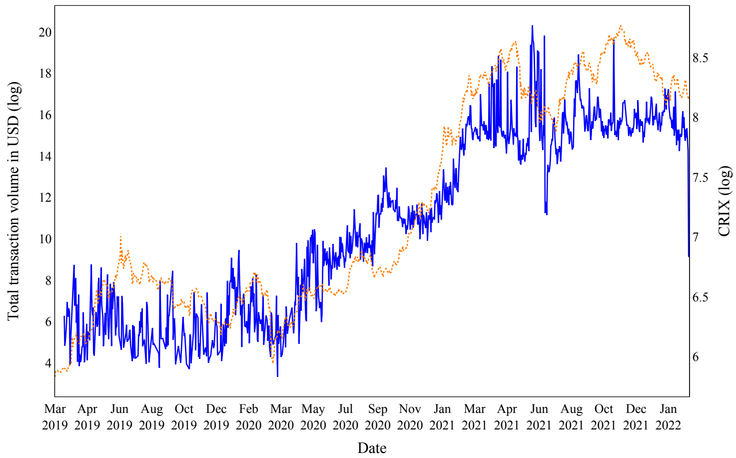



**Figure 2.1: Total transaction volume of NFT art in USD (solid) and CRIX (dotted).**

Similar to conventional art, NFT artworks are traded infrequently and there exists an imbalance in asset pricing. Vasan et al. (2022) demonstrate the existence of a price-advantage of first-mover creators and a network effect from social media followers. Rarity and uniqueness of collectibles induce market demand and thus offer a rarity premium (Koford and Tschoegl, 1998). 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.

Finally, we notice some adoptions from the conventional art market in the NFT art market. For example, in the conventional art market, curation processes play a critical role for the valuation of a work. Applying such a mechanism to the digital art market, *ArtBlocks* offers a curation board for evaluating generative artworks from their scripts.

## 2.3 Data description

We collect the data through the API provided by OpenSea, the first and largest NFT marketplace. To ensure representative constituents without an unwieldy sample size, we choose the (at that date) top 10 NFT art collections based on total transaction volumes, accounting for 83.71% of the NFT art market volume as of July 06, 2021. The data covers the period from April 1, 2019, to March 31, 2022 and comprises two data sets: artwork's transaction data and metadata, which are subsequently merged using the unique id of each NFT. Table 1 presents the descriptive statistics. The data comprises 91,841 works and 61,478 sales across the ten collections, with an average sale price of US$74,567.68 and a standard deviation of US$2,241,169.84. CryptoPunks stands out with the highest transaction volume and the highest mean sale price. Collections like Rarible, SuperRare, Art Blocks, and MakersPlace feature a large number of works, as they serve as digital art galleries and marketplaces where many independent creators submit their works. The remaining collections are owned by institutional creators.

Table A.1 presents the metadata along with their attributes. The metadata provides details about each work, including the collection slug, contract type, scheme name, creator, year of creation, and traits (characteristics of artworks). Collection-specific properties are integrated as interactive dummies, such as sub-collections in ArtBlocks and backgrounds in Bored Ape Yacht Club. To handle collections with extensive trait sets, such as CryptoPunks' accessories, and unstructured textual descriptions, like SuperRare, we employ natural language processing (NLP) techniques to mitigate the feature space expansion. Further details on this processing can be found in Appendix B. Lastly, the transaction-related information in the metadata for each work consists of both categorical and continuous variables. The categorical variables (e.g., auction type, payment token, seller, and winner addresses) are one-hot encoded. The continuous variables (e.g., frequency of the same creator, frequency of the same owner address) are divided into intervals according to their frequency distributions.

# 3 Towards DAI, an Index for Digital Art

In this section, we introduce the methodology for constructing a price index for NFT art, taking into account the pecularities of the NFT art market. Before presenting the hedonic regression framework, we have as a preliminary step the correction of potential sample selection biases.

## 3.1 Selection bias correction

Sample selection bias can be viewed as a particular form of omitted variable bias and is particularly relevant in hedonic regressions where the sample consists only of sold items, whereas unsold ones are unobserved, see e.g. Malpezzi et al. (2003); Collins et al. (2009). In our case, some NFT transactions cannot be included because of unobserved prices, for

**Table 1: Descriptive Statistics of NFT art collections** The statistics for the sale prices in USD include the mean, standard deviation, minimum, maximum, and median values.

| **Collection** | **CryptoPunks** | **Rarible** | **SuperRare** | **Bored Ape Yacht Club** | **Hashmasks** |
|---|---|---|---|---|---|
| **Number of assets** | 8,692 | 6,844 | 13,054 | 8,548 | 9,350 |
| **Number of sales** | 5,315 | 4,913 | 9,131 | 7,583 | 7,583 |
| **Transaction volume** | 466,690,540.69 | 209,234,594.07 | 89,732,333.24 | 86,227,401.78 | 72,123,445.50 |
| **Mean** | 212,097.24 | 344,586.17 | 7,792.95 | 87,112.27 | 120,073.14 |
| **Standard deviation** | 4,482,995.88 | 1,570,411.72 | 91,090.14 | 117,999.40 | 4,992,995.53 |
| **Minimum** | 3.52 | 0.19 | 0.16 | 10.99 | 1.33 |
| **Median** | 67,760.00 | 79.52 | 912.00 | 35,427.50 | 2,980.75 |
| **Maximum** | 325,702,901.08 | 10,692,800.00 | 4,265,693.70 | 2,012,465.31 | 402,498,000.00 |
| **Collection** | **Art Blocks** | **MakersPlace** | **Wrapped CryptoPunks** | **Autoglyphs** | **Beeple Everydays** |
| **Number of artworks** | 25,116 | 17,645 | 1,358 | 512 | 722 |
| **Number of sales** | 16,153 | 9,751 | 733 | 230 | 86 |
| **Transaction volume** | 47,802,950.66 | 32,166,275.87 | 20,798,316.41 | 20,344,441.74 | 12,038,702.98 |
| **Mean** | 3,807.87 | 1,761.09 | 7,565.45 | 135,730.28 | 73,655.38 |
| **Standard deviation** | 23,120.31 | 12,868.51 | 20,070.29 | 246,721.85 | 65,947.09 |
| **Minimum** | 0.19 | 1.69 | 19.90 | 32.80 | 1,279.13 |
| **Median** | 1,435.59 | 495.60 | 1,672.05 | 14,422.44 | 52,274.38 |
| **Maximum** | 2,496,000.00 | 886,557.00 | 261,437.30 | 1,203,815.40 | 354,222.80 |

example due to cancellations. To address this issue, we use Heckman's two-step estimator of the sample selection model, or Tobit-II model (Heckman, 1976, 1979).

Assume a latent outcome variable ${y^{\mathbf{o}}_{t,i}}^*$, for time periods $t \in \{1, \dots, T\}$ and sales $i \in \{1, \dots, N_t\}$, and another latent selectivity variable ${y^{\mathbf{s}}_{t,i}}^*$ that determines whether the outcome $y^o_{t,i}$ is observed (in which case $y^o_{t,i} = {y^{\mathbf{o}}_{t,i}}^*$) or not. Suppose that for each $t$ and $i$ we have a $(K \times 1)$ vector of covariates $x^o_{t,i}$ for the outcome variables, and another $(k \times 1)$ vector of covariates $x^s_{t,i}$ for the selectivity variables, assuming that this is a sub-vector of $x^o_{t,i}$. Consider the model

$$ {y^{\mathbf{o}}_{t,i}}^* = {x^{\mathbf{o}}_{t,i}}^\top \alpha^{\mathbf{o}}_t + \varepsilon^{\mathbf{o}}_{t,i} \tag{3.1} $$

$$ {y^{\mathbf{s}}_{t,i}}^* = {x^{\mathbf{s}}_{t,i}}^\top \alpha^{\mathbf{s}}_t + \varepsilon^{\mathbf{s}}_{t,i} \tag{3.2} $$

with error terms $\varepsilon^{\mathbf{o}}_{t,i}$, $\varepsilon^{\mathbf{s}}_{t,i}$ and coefficient vectors $\alpha^{\mathbf{o}}_t(K \times 1)$ and $\alpha^{\mathbf{s}}_t(k \times 1)$. The outcome $y^{\mathbf{o}}_{t,i}$ is only observed if the selection propensity variable is positive, i.e. ${y^{\mathbf{s}}_{t,i}}^* > 0$. Conditional on selection, the expected value of the outcome in (3.1) is given by:

$$ \mathsf{E}[y^{\mathbf{o}}_{t,i} | x^{\mathbf{s}}_{t,i}, \varepsilon^{\mathbf{s}}_{t,i}] = {x^{\mathbf{o}}_{t,i}}^\top \alpha^{\mathbf{o}}_t + \mathsf{E}[\varepsilon^{\mathbf{o}}_{t,i} | \varepsilon^{\mathbf{s}}_{t,i} \geq -{x^{\mathbf{s}}_{t,i}}^\top \alpha^{\mathbf{s}}_t] \tag{3.3} $$

Assuming a bivariate Gaussian distribution for the error terms, i.e. $\binom{\varepsilon^{\mathbf{o}}_{t,i}}{\varepsilon^{\mathbf{s}}_{t,i}} \sim \mathcal{N}\left(\binom{0}{0}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix}\right)$, the expectation on the right hand side of (3.3) can be calculated as $\mathsf{E}[\varepsilon^{\mathbf{o}}_{t,i} | \varepsilon^{\mathbf{s}}_{t,i} \geq -{x^{\mathbf{s}}_{t,i}}^\top \alpha^{\mathbf{s}}_t] = \rho\sigma\varphi({x^{\mathbf{s}}_t}^\top \alpha^{\mathbf{s}}_t)/\Phi({x^{\mathbf{s}}_t}^\top \alpha^{\mathbf{s}}_t)$, where $\varphi(\cdot)$ and $\Phi(\cdot)$ are the standard normal probability density function (pdf) and cumulative distribution function (cdf), respectively.

In a first step, $\alpha_t^{\mathbf{s}}$ is estimated via a probit model for the binary variable $y_{t,i}^{\mathbf{s}}$, which takes the value 1 if $y_{t,i}^{\mathbf{s}\,*} > 0$, and 0 otherwise. In a second step, the model (3.3) is estimated by running an OLS regression of $y_{t,i}^{\mathbf{o}}$ on $x_{t,i}^{\mathbf{o}}$ and $\varphi(x_t^{\mathbf{s}\top}\widehat{\alpha}_t^{\mathbf{s}})/\Phi(x_t^{\mathbf{s}\top}\widehat{\alpha}_t^{\mathbf{s}})$, where $\widehat{\alpha}_t^{\mathbf{s}}$ is the probit estimator of the first step. We use this bias correction in our subsequent models for constructing the price index.

## 3.2 Hedonic regression

An HR model regresses prices on various characteristics of assets transacted and can 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. The characteristics of NFT art pricing is fundamentally different from classical art markets. We therefore extend the HR model of Bocart and Hafner (2015) to account for the pecularities of NFT art markets. Our transaction data can be viewed as an unbalanced panel, since for every time point $t$, a potentially different number of transactions $N_t$ is recorded. Let us denote $y_t = (y_t^1, \ldots, y_t^{N_t})^\top$ where $y_t^i$ is the log price of the $i$-th sale at time $t \in \{1, \ldots, T\}$ with $i \in \{1, \ldots, N_t\}$; $\varepsilon_t = (\varepsilon_t^1, \ldots, \varepsilon_t^{N_t})^\top$ with $\varepsilon \overset{\text{iid}}{\sim} \mathcal{N}(0, {\sigma_\varepsilon}^2)$. Note that $i$ here refers to the sale rather than a particular object. The HR model can now be written as

$$y_t = d_t\beta_t + X_t^\top\alpha + \varepsilon_t \tag{3.4}$$

where $d_t = (1, \ldots, 1)^\top$ is a $(N_t \times 1)$ vector, $\beta_t = (\beta_t^1, \ldots, \beta_t^{N_t})^\top$ is the vector that constitutes the price index, $X_t$ is the $(N_t \times K)$ matrix of characteristics of NFT art, and $\alpha = (\alpha^1, \ldots, \alpha^K)^\top$ is the corresponding coefficient vector.

In our study, $X_t$ includes the factors listed in Table A.1. These factors comprise both one-hot encoded categorical variables and continuous variables, such as the number of sales and the CryptoPunks accessory's scarcity. Note that for a given artwork, variables such as traits and characteristics of a work are time-invariant, while others related to frequency and continuous attributes may change over time. The parameter vector $\alpha$ is interpreted as the implicit prices of the various characteristics. The fixed time effects $\beta_t$ and hedonic prices $\alpha$ can be estimated by OLS, and one obtains a price index $\hat{\beta}_t^{OLS}$ of a "neutral" or characteristic-free artwork.

There are, however, two issues with OLS estimation: First, the squared loss is not robust with respect to outliers, and second, it does not penalize for model complexity in high dimensions. The robustness problem will be treated in the next section, while for the second issue we propose to do a model regularization via LASSO (least absolute shrinkage and selection operator), where the estimator is given by

$$(\hat{\alpha}, \hat{\beta}) = \arg \min_{\beta \in \mathbb{R}^T, \alpha \in \mathbb{R}^K} \left\{ \sum_{t=1}^{T} \|(y_t - d_t \beta_t - X_t^\top \alpha)\|_2 + \lambda \|\alpha\|_1 \right\} \tag{3.5}$$

where $\lambda$ is the LASSO regularization parameter. LASSO regularization shrinks the least relevant coefficients to exactly zero, which enhances interpretability and reduces model complexity.

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

$$\text{Index}_t = 100 \frac{\exp\left(\hat{\beta}_t\right)}{\exp\left(\hat{\beta}_1\right)} \tag{3.6}$$

Another prevalent approach for constructing price indices for heterogeneous assets is repeat-sales regression, which focuses solely on assets with at least two sales in order to compute returns that are averaged for each time period. It has been used, for example, by Bailey et al. (1963) for real estate prices. While the advantage of RSR is that it does not depend on a choice of characteristics, its main drawback is that much of the data is discarded when most of the artworks are sold only once. In this paper, we refrain from using RSR, since the NFT art market is relatively illiquid with low sales frequencies, as shown in Figure 3.1 for various listed collections. Moreover, as highlighted by Shiller (2008), when there is substantial variation in either the fraction of sales or the relative price of new assets, it becomes essential to include new sales alongside existing sales. In this paper, we therefore concentrate on HR regression.

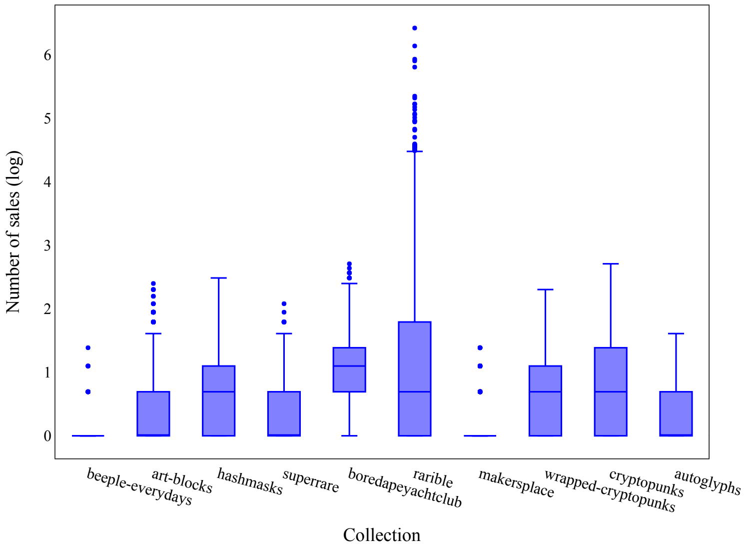



**Figure 3.1: Boxplots for the log number of sales in different NFT art collections**.

## 3.3 Huberization

Proceeding with the OLS regression, we observe that there exist many outliers in residuals $\hat{\varepsilon}_t$, see the QQ plot in Figure 3.2, and also Figure 3.3 where the OLS residuals are displayed via box plots over time. This is compared with the number of transactions per day in the lower panel of the same figure. While the mean of residuals stabilizes over time, there is clearly a positive association between number of transactions and outliers in residuals.

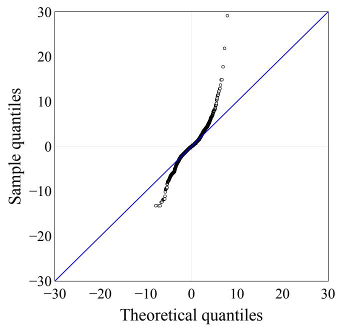



**Figure 3.2: QQ plot for OLS residuals against normal distribution**.

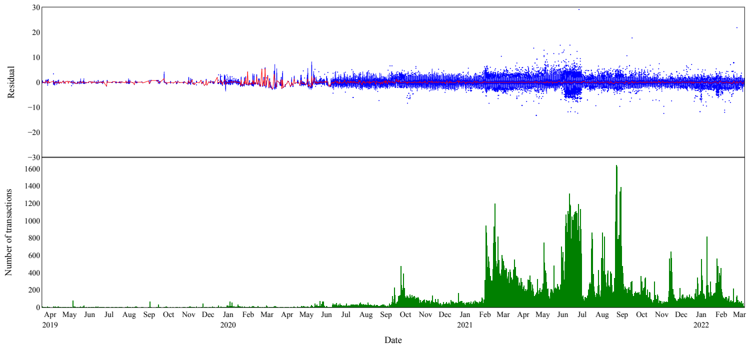



**Figure 3.3: OLS residual box plots over time with mean of residuals and number of transactions per day.**

In order to alleviate the impact of outliers on the estimates of $\beta$, we introduce a correction procedure. Härdle (1984) uses an influence curve, Huber's $\psi$-function proposed by Huber (1964), to bent down the influence of outlying observations. For implementation we construct observations $\tilde{y}_t = (\tilde{y}_t^1, \dots, \tilde{y}_t^{N_t})^\top$ based one a one-step correction of $y_t$:

$$\tilde{y}_t = d_t\hat{\beta}_t + X_t^\top \hat{\alpha} + \tilde{\varepsilon}_t \tag{3.7}$$

where $\hat{\beta}_t$ and $\hat{\alpha}$ are the estimates of the OLS or LASSO regressions, while $\tilde{\varepsilon}_t = (\tilde{\varepsilon}_t^1, \dots, \tilde{\varepsilon}_t^{N_t})^\top$ is a vector of bounded one-step residuals,

$$\tilde{\varepsilon}_t^i = \frac{\psi(\hat{\varepsilon}_t^i)}{\Delta} \tag{3.8}$$

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 residuals $\hat{\varepsilon}_t^i$. The $\psi$-function is the derivative of the $\rho$-function defined as

$$\rho_\tau(\varepsilon_t^i) = \begin{cases} \frac{1}{2}(\varepsilon_t^i)^2 & \forall |\varepsilon_t^i| \leq \tau \\ \tau(|\varepsilon_t^i| - \frac{1}{2}\tau) & \text{otherwise} \end{cases} \tag{3.9}$$

where $\tau \in \mathbb{R}^+$ is a hyperparameter that controls the effect of outliers. The huberized LASSO estimator is then obtained as

$$(\hat{\alpha}^*, \hat{\beta}^*) = \underset{\beta \in \mathbb{R}^T, \alpha \in \mathbb{R}^K}{\operatorname{argmin}} \left\{ \sum_{t=1}^{T} \|\tilde{y}_t - d_t\beta_t - X_t^\top \alpha\|_2^2 + \lambda \|\alpha\|_1 \right\}. \tag{3.10}$$

The estimator of $\beta$ in (3.10) can be used to construct a price index, taking into account sample selection bias, robustness, and regularization for model complexity. However, this estimator is based on assuming that $\beta$ is a fixed parameter, without dynamics driven by a stochastic process. In the following we relax this assumption by allowing for random effects.

## 3.4 Random effects models

Rather than considering the time effects $\beta_t$ as fixed parameters, one can alternatively specify random effects as in Bocart and Hafner (2015). This is particularly attractive as it allows to introduce temporal dependencies in the price index series, for example via autoregressive or random walk type processes. Because the underlying process is unobserved, some filtering procedure is necessary, the most popular of which is the Kalman filter. An alternative filtering method is based on the recently developed dynamic conditional score (DCS) models. In the following we discuss these two approaches.

### 3.4.1 Kalman filtering

We treat the parameter $\beta_t$ of (3.4) as random effects in a state space form and apply Kalman filtering. Extending the classic random-effects models and imposing exogeneity of the regressors regarding the time component such that $\beta_t \sim \mathcal{N}(0, \sigma_\beta^2)$ with $\mathsf{E}[\beta_t|X] = 0$, we rewrite (3.4) as $y_t = X_t^\top \alpha + \eta_t$ with a composite error term $\eta_t = (\eta_t^1, \dots, \eta_t^{N_t})^\top$ and

$$\eta_t = d_t \beta_t + \varepsilon_t \tag{3.11}$$

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

If Huberization and the LASSO are used, then in a first step an estimator of $\alpha$ is obtained as

$$\hat{\alpha} = \underset{\alpha \in \mathbb{R}^K}{\operatorname{argmin}} \left\{ \sum_{t=1}^{T} \|\tilde{y}_t - X_t^\top \alpha\|_2^2 + \lambda \|\alpha\|_1 \right\}. \tag{3.12}$$

and then a model is fitted to the obtained residuals $\hat{\eta}_t = \tilde{y}_t - X_t^\top \hat{\alpha}$ in a second step.

Allowing for time series dynamics of the index $\beta_t$, we fit an autoregressive process of order one, AR(1), including a random walk as special case:

$$\beta_t = \phi \beta_{t-1} + \xi_t \tag{3.13}$$

with $|\phi| \leq 1$ and $\xi \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\xi{}^2)$. As $\phi = 1$, $\xi_t$ is the excess over the average price – interpreted as returns of a market portfolio – and $\sigma_\xi$ is hence seen as market volatility. Here, we explicitly focus on the random walk process in (3.13) with $\phi = 1$.

Given the linear Gaussian state space representation in (3.11) and (3.13), we use the Kalman filter to estimate $\beta_t$ for given model parameters, while $\sigma_\varepsilon$ and $\sigma_\xi$ are estimated by maximum likelihood. Denote the unknown parameter vector by $\theta = \left(\sigma_\varepsilon, \sigma_\xi\right)$ within the parameter space $\Theta = \{\theta : \sigma_\varepsilon^2 > 0, \sigma_\xi^2 > 0\}$.

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 $\theta$ as:

$$\mathcal{L}(\theta) = \frac{1}{2} \sum_{t=1}^{T} \Big\{ \log|\Sigma_t(\theta)| + e_t(\theta)^\top \Sigma_t(\theta)^{-1} e_t(\theta) \Big\} \tag{3.14}$$

and the maximum likelihood estimator is:

$$\hat{\theta} = \underset{\theta \in \Theta = \mathbb{R}_+^2}{\operatorname{argmax}} \, \mathcal{L}(\theta) \tag{3.15}$$

Using the estimator $\hat{\theta}$, we obtain predictions $\beta_{t|t-1}$ and $\eta_{t|t-1}$, updates $\beta_{t|t}$ and $\eta_{t|t}$ and smoothed estimates $\beta_{t|T}$ and $\eta_{t|T}$ via the Kalman filter. Given the full sample information $(t = 1, \ldots, T)$, we therefore infer the underlying $\beta_t$ using the smoothed etimates $\beta_{t|T}$.

### 3.4.2 DCS-t filtering

The Kalman filter is optimal in a linear Gaussian state space model, see e.g. Durbin and Koopman (2012). In the presence of outliers, however, it may overweight the impact of these outliers on the changes in the index $\beta_t$. 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 $\nu$ degrees of freedom. Assuming a random walk for $\beta_t$ , the model becomes:

$$\eta_t = d_t\beta_t + v_t, \quad v_t \sim t_\nu(0, \sigma_v^2 I_{n_t}) \tag{3.16}$$

$$\beta_t = \beta_{t-1} + \kappa u_{t-1} \tag{3.17}$$

$$u_t = \frac{1}{w_t}\frac{\nu + n_t}{\nu}\frac{1}{\sigma_v^2} d_t^\top e_t \tag{3.18}$$

$$w_t = 1 + \frac{1}{\nu\sigma_v^2} e_t^\top e_t \tag{3.19}$$

Here, $u_t$ is the score w.r.t. $\beta_t$, i.e.

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

where $f_t$ is the conditional density of $\eta_t$,

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

Figure 3.4 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 Kalman filtering, and in which case there is no downweighting of outliers.

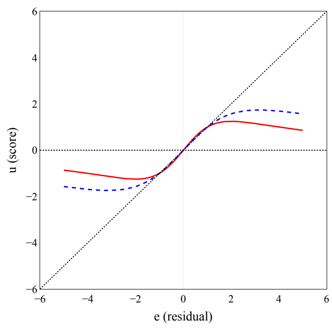



**Figure 3.4: Score $u_t$ with respect to $\beta_t$.** It is a function of the one-step prediction error for a multivariate t distribution with $\nu = 4$ (solid) and $\nu = 10$ (dashed).

Estimation of the model 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 $\theta = (\sigma_v^2, \kappa, \nu)^\top$.

# 4 Results

In this section we present the results of our empirical analysis and study the performance of Kalman and DCS-t filters when applied to our NFT transaction data. We perform a residual analysis to demonstrate how huberizing before applying KF reduces the influence of outliers, and compare it with the alternative DCS-t filter that dynamically reduces this impact via the score function.

## 4.1 KF and DCS-t filterings

The score-driven filter (DCS-t) shows adaptability in estimating time-varying parameters based on past observations within a diverse range of nonlinear models, as highlighted in a significant body of literature (Harvey, 2022; Artemova et al., 2022). It is capable of handling complex time series dynamics, such as outliers or time-varying volatilities (Harvey, 2013). In our empirical study of the NFT art market (Figure 4.1), the score update increases in magnitude after the midpoint of the observation period and shows significant autocorrelations up to a week. In contrast, the classic Kalman filtering may fail to respond adequately to a sudden change due to its reliance on an accurate state space model and prior distribution of noise. The empirical results shown in Figure 4.2 illustrate that the filtered index $\hat{\beta}_{t|t}^{\mathrm{KF}}$ is volatile during the first half of the observation period, gradually smoothing thereafter, whereas $\hat{\beta}_{t|t-1}^{\mathrm{DCS\text{-}t}}$ shows the opposite pattern. Moreover, we present a simulation study in Appendix C to illustrate the estimation performances of both filters under various scenarios. The results demonstrate that the DCS-t filter has superior adaptability to volatile time series.

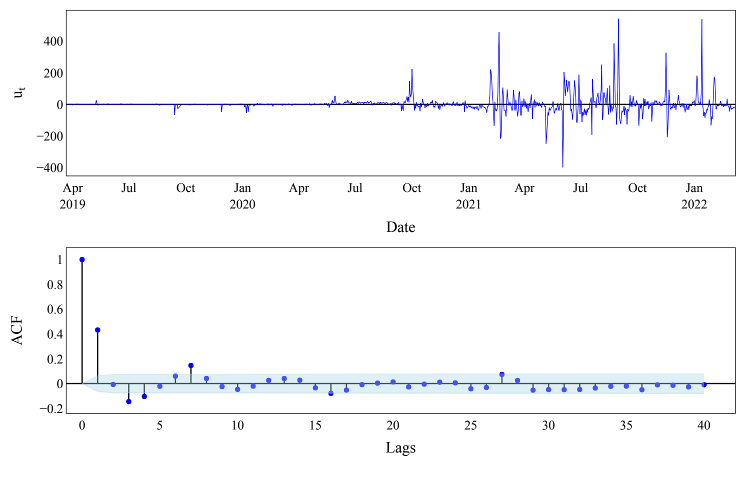



**Figure 4.1: The update and ACF of score $u$.**

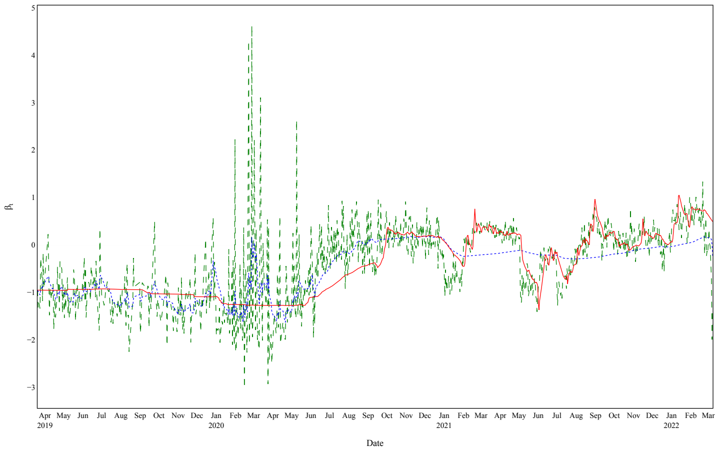


**Figure 4.2: KF predicted $\hat{\beta}^{\mathbf{KF}}_{t|t-1}$ (dashed), KF corrected $\hat{\beta}^{\mathbf{KF}}_{t|t}$ (dotted), and DCS-t predicted $\hat{\beta}^{\mathbf{DCS\text{-}t}}_{t|t-1}$ (solid).**

## 4.2 Residual analysis

The residuals of OLS estimation of the fixed effects model after the Heckman correction exhibit heavy tails as illustrated in Figure 3.2. Figure 4.3 shows how Huberization for a given $\tau = q_{0.01}(\hat{\varepsilon})$ effectively truncates residuals over time and substantially reduces the influence of outliers.

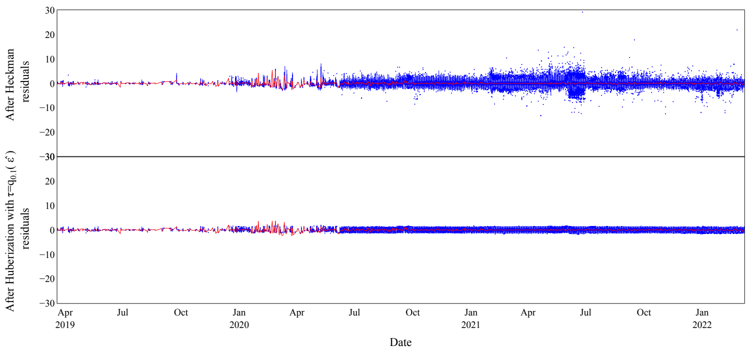
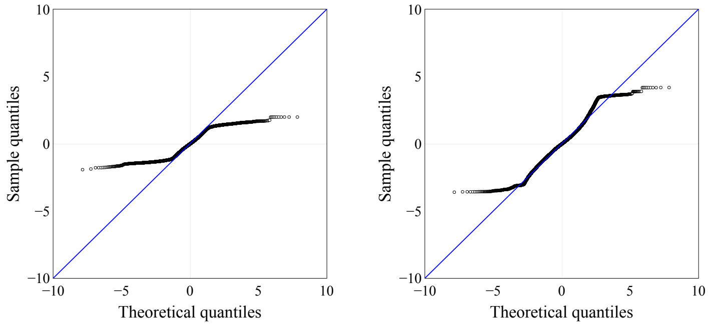



**Figure 4.3: Residual box plots after Heckman correction (upper panel) and after Huberization (lower panel) over time with mean of residuals.**

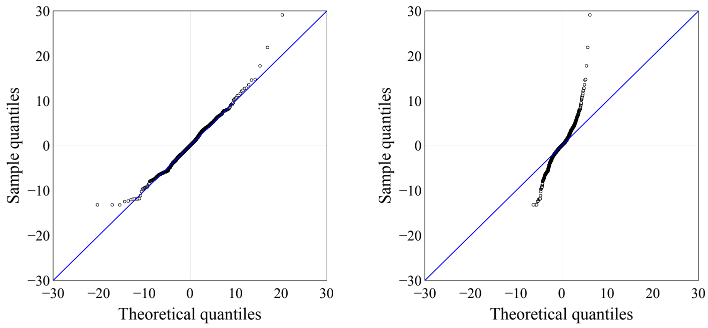
In Figure 4.4 we compare two different quantile levels $\tau$ of residuals after Heckman correction – $q_{0.1}(\hat{\varepsilon})$ and $q_{0.01}(\hat{\varepsilon})$. For $\tau = q_{0.1}(\hat{\varepsilon})$, residuals are capped severely to about $[-2, 2]$, while $\tau = q_{0.01}(\hat{\varepsilon})$ gives a more gentle Huberization with residuals kept roughly in $[-4, 4]$. Overall, $\tau = q_{0.01}(\hat{\varepsilon})$ delivers residuals that align well with the normal distribution for a wide range of central quantiles, while having downweighted extremes.




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

**Figure 4.4: QQ plots for OLS residuals against standard normal distribution after Huberization with different quantile levels**.

As an alternative to using Huberization and Kalman filtering, DCS-t filtering uses the scores of the conditional distribution to update the filter. Given the existence of outliers and heavy-tailed distributions, a multivariate $t$-distribution is employed, as in (3.16). In Figure 4.5, we present the QQ plot for the OLS residuals post Heckman correction, compared against the standard $t$-distribution with two different degrees of freedom $\nu$. A heavier tail in the $t$-distribution (i.e. $\nu = 4$) provides greater robustness compared to the normal distribution, and a good fit to the data. On the other hand, choosing $\nu = 10$ underestimates the presence of extremes and does not yield a good fit. We therefore proceed by choosing $\nu = 4$ for the DCS-t filter.




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

**Figure 4.5: QQ plot for OLS residuals against student $t$-distribution with two different degrees of freedom $\nu$**.

## 4.3 Comparison of KF and DCS-t

The two proposed innovative procedures enhance robustness by downweighting outlier influence, crucial for constructing indices for highly heterogeneous goods. We now discuss the construction of price indices using these procedures under various parameter settings.

Figure 4.6 visually demonstrates the index variants employing Huberization with various settings on the hyperparameter $\tau$. The Heckman corrected index (without any outlier handling) is significantly affected by outliers, particularly during the time of March to April 2020, where it shows a relatively sharp spike. For Huberization, all three considered levels of $\tau$ lead to substantial reduction of this spike, and are quite similar in other periods. As sales have been increasing since late 2020, see Figure 1.1, the Huberized indices show a smoother trend, which may be a concern as the indices do not adapt quickly to a highly active market.

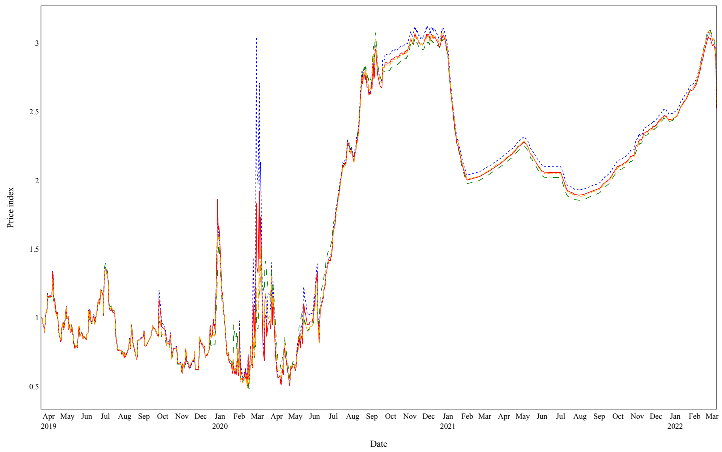



**Figure 4.6: DAI – after Heckman correction (dotted), after Huberization with $\tau = q_{0.1}(\hat{\varepsilon})$ (dashed), after Huberization with $\tau = q_{0.05}(\hat{\varepsilon})$ (dash-dotted), and after Huberization with $\tau = q_{0.01}(\hat{\varepsilon})$ (solid).**

Presenting an alternative to Huberization and Kalman filtering, Figure 4.7 compares the alternatives using DCS-t filtering with optimized parameters (via MLE) and four degrees of freedom to Huberizing with $q_{0.01}(\hat{\varepsilon})$. Following the fourth quarter of 2020 with an increasing number of transactions (see Figure 3.3), the indices employing DCS-t filtering demonstrate higher reactivity comparison to Huberized indices, while being more stable and less affected by a small number of outliers when there are fewer observations. This feature is especially desirable for an emerging market, as it facilitates an index that retains robustness to outliers while considering their influence, especially in cases where they occur frequently.

As demonstrated earlier in Section 4.2, both procedures offer robustness. Huberization in a one-step robust regression, crops off the influence from the outliers at once, followed by the Kalman filtering to estimate the unobserved components. On the other hand, DCS-t filtering fits the data using a heavy-tailed distribution and gradually adapts through

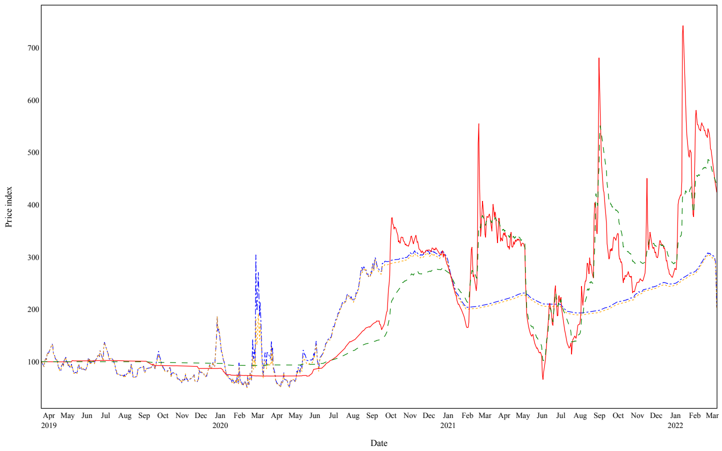



**Figure 4.7: DAI variants – after Heckman correction (dash-dotted), after Huberization with $\tau = q_{0.01}(\hat{\varepsilon})$ (dotted) DCS-t filtering (solid), and DCS-t filtering with $\nu \sim \infty$ (dashed).**

one-step updates via the scores of the conditional distribution. Given the rapid evolution of the NFT art market, incorporating the dynamics of the conditional distribution offers an adaptive and straightforward approach. We will therefore concentrate on the index constructed using DCS-t filtering in the following section.

# 5 The DAI

In this section, we begin by exploring the factors influencing the prices of the collected artworks. As NFTs are predominantly traded using cryptocurrencies, particularly ETH, their return dynamics can significantly impact the overall performance of the NFT art market (Dowling, 2022). We, thus, examine the causal relationships between DAI and other relevant crypto variables, such as CRIX, ETH, and trading volume (ETH vol.). Finally, we present a summary of key stylized facts related to the NFT art markets.

## 5.1 Price determinant

We perform a hedonic LASSO analysis (3.5) on the prices of the collected NFT artwork. The hedonic factors listed in Table A.1 are considered as independent variables, and the analysis is conducted with a penalty parameter of $\lambda = 0.001$. A $\lambda$ value of 0.001 was selected based on its superior performance metrics, including an adjusted R-squared of 0.87, an Akaike Information Criterion (AIC) of 7.22, and a Bayesian Information Criterion (BIC) of 7.05. In comparison, a $\lambda$ value of 0.01 yielded an adjusted R-squared of 0.78, an AIC of 13.27, and a BIC of 13.65.

Tables C.1 and C.2 display the active variables and their corresponding coefficients. Active variables are factors with non-zero regression coefficients in the final model after applying LASSO regularization. In summary, we have 129 active variables out of the 391 factors considered. To clarify the price determination among these artworks, we categorize the active variables into two groups: 37 external factors and 92 internal factors.

### 5.1.1 External factors

External factors encompass transactional and contextual information that describes trading details, such as transaction time, inflow and outflow addresses, and the number of sales that have occurred.

First, we see that the auction type[1]: English auction – has a positive effect on NFT art price. The English auction is an open-outcry ascending dynamic bids method, which is more commonly used and straightforward for sellers and buyers. It is the most commonly used auction type on Opensea. We observe that trading with commonly used cryptos – Ethereum (ETH), Wrapped Ether (WETH), Dai result in higher prices for NFTs. Additionally, the widely adopted NFT standard, ERC721, has a positive influence, while the ERC-1155, which enables trading multiple tokens using a single contract, has not yet become mainstream.

Regarding seller experience, the most substantial influence on price is observed from sellers with 2 to 100 transaction records, while even those with only one transaction record show a positive influence. That is, an inexperienced seller may still have an opportunity to achieve a high sale price. The market remains welcoming to new entrants. Buyers with experience between 100 and 500 sales exhibit a significant negative impact on price, while less experienced buyers contribute to higher deal prices. Amid the NFT art market's emerging nature, lack of clarity in price formulation, and presence of asymmetric information, inexperienced buyers find it challenging to comprehend market dynamics, resulting in higher bids on assets when making their initial purchases. This can also be attributed to the variable frequency of the same owner address; assets owned by individuals with only one asset are more likely to be sold at lower prices. In contrast, individuals owning more than 100 assets exert a positive influence on prices. Overall, experienced sellers and buyers tend to gain advantages in the NFT art market, likely due to their proficient bidding skills and keen market sensibility.

[1]English auction: Each asset is advertised for sale, and the highest bidder wins the item; Dutch auction: Each asset is bid upon during the auction period, but the price decreases over time.

In terms of creators, the frequencies of same creator address and ID have a negative impact. Listing multiple works does not lead to a higher price for creators. Moreover, the number of sales for a work exhibits a positive influence, indicating that frequently sold or popular NFT artworks are reflected in their price.

The variable contract address is given by the launching platforms, such as the example where 0x8c9f364bf7a56ed058fc63ef81c6cf09c833e656 corresponds to the auction house, SuperRare. As such, its influence is pertinent to the collection slugs, which we will further address when discussing internal factors.

### 5.1.2 Internal factors

These factors, often, correspond to the intrinsic values of artworks, encompassing details such as the collection or auction house of origin, the creation date, and the visual properties of a work.

Two types of art collections can be distinguished: ones like Rarible and Makersplace, which mainly feature works from independent creators, and others like CryptoPunks and Autoglyphs, which shows artworks from institutional creators. In general, artworks from independent creators tend to be sold at lower prices compared to those from institutional creators, as evidenced by the negative coefficients of Rarible and Makersplace. This result is also consistent with Figure 1.2, where the medians of log prices for Rarible and Makersplace are lower than the others. Moreover, these institutionally initiated works within the same collection exhibit a tendency to share similar visual compositions, as they are generated by an algorithm. In contrast to the conventional art market, where the reputation of an artist or authenticity evidence usually drives price increases (Renneboog and Spaenjers, 2013), the price of an NFT artwork is highly influenced by its publisher and the collection to which it belongs, rather than being primarily determined by the artists themselves.

The year of creation exhibits a negative influence, with earlier works, particularly those from 2019, having a more pronounced impact on prices. This effect could be attributed to the oversupply and uneven distribution of artworks among collections, similar to the concept of the tragedy of the commons discussed in Hardin (1998). After identifying the interactive variables that describe the properties of each work in each collection, we draw the following conclusions:

**Art Blocks.** This collection primarily showcases demand-oriented generative works. Buyers choose their desired style and subsequently receive a corresponding randomly generated artwork produced by an algorithm. The artwork can take the form of a static image, a 3D model, or an interactive animation. Within this collection, the selection of style holds significant importance (see Table C.2), i.e. Archetypes by Kjetil Golid, Ringers by Dmitri Cherniak. The Art Blocks section Playground, which consists of curated artists' experimental projects, has a positive impact, while works from the section Factory, lacking curation and valuation processes, exert a negative influence on prices. So, a well-implemented curation process and quality control of artworks can positively impact their pricing.

**Bored Ape Yacht Club (BAYC).** It comprises 10,000 limited unique Bored Ape NFTs and places significant emphasis on forming an online community membership that grants access to exclusive members-only benefits. BAYC's features, such as the background, earring, fur, hat, eyes, clothes, and mouth of an ape, all play a crucial role in influencing the pricing of a work. The higher the scarcity and bolder the features (e.g. trippy fur, king's crown, beam or laser eyes, black suit) of an ape, the higher its price tends to be. This finding may be attributed to the buyers' preference for unique features that can function as status symbols within the online community.

**Hashmasks.** The collection includes 16,384 digital portraits created by over 70 artists worldwide. Similar to BAYC, rare features such as item–Shadow Monkey, masks–Abstract, Animal, Steampunk, and characters–Golden Robot, Mystical, and Puppet, exhibit a positive influence on prices. Works having no distinguishable appearance or features (e.g. no items, male and female characters) have a lower price.

**CryptoPunks.** The collection is one of the earliest adopters of an NFT with the ERC-721 standard and consists of limited $10,000$ unique collectible characters. The scarcity of attributes, also, plays a significant role in determining the pricing of artworks within the collection. The variable CryptoPunk accessory is derived from the scarcity index for the accessories of a punk (see Appendix B), and it has a notably positive impact on pricing. The most common punk type–Male influences negatively.

**SuperRare.** This collection's features (i.e. tags) are clustered into groups based on the similarity (see Appendix B). In contrast to CryptoPunks, SuperRare serves as a marketplace for independent creators, and its artworks' tags are more sparse and inconsistent. While the tag group to which an artwork belongs does have some influence on the price, this influence is relatively weak.

**Autoglyphs.** In total, it consists of 512 algorithmically generated works. The collection slug demonstrates a notably positive influence, whereas none of the patterns show significance in their pricing. Due to the limited number of works in Autoglyphs, the collection slug has a significant influence over the artworks' style.

In conclusion, we observe that the scarcity of NFT artworks' internal factors plays a role in determining their prices, resembling the dynamics seen in the conventional art market. The collection to which a work belongs can often hold more influence over its price than the individual creators and artworks themselves.

## 5.2 Granger causality analysis

In order to understand the relationship between NFT and CC markets, we now perform a Granger causality analysis. In particular, we look at daily returns based on the DAI and CRIX indices, ETH and the transaction volume of ETH. For each pair of variables, we specify an autoregressive model of order $\tau$ for the first variable, including $\tau$ lags of the second variable, and then test for joint significance of the coefficients of the second variable using an F-test. We do this for different number of lags, $\tau$. Based on the results presented in Table 2, there is no significant evidence to suggest that CRIX, ETH, and ETH vol. have a causal influence on DAI, nor vice versa; thus, the NFT art market might serve as a potential financial instrument for risk diversification concerning cryptocurrencies, given the limited causal impact on DAI.

**Table 2: Pairwise Granger causality test results**.

| **Number of lags $\tau$** | **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 $\nRightarrow$ DAI | 0.0019 | 0.9649 | 0.2195 | 0.8030 | 0.1465 | 0.9319 | 0.5682 | 0.6858 |
| DAI $\nRightarrow$ CRIX | 0.0299 | 0.8628 | 0.1481 | 0.8624 | 0.2746 | 0.8437 | 0.2032 | 0.9366 |
| ETH $\nRightarrow$ DAI | 0.0690 | 0.7928 | 0.0795 | 0.9236 | 0.4170 | 0.7409 | 0.5071 | 0.7305 |
| DAI $\nRightarrow$ ETH | 0.0369 | 0.8477 | 0.0359 | 0.9647 | 0.2721 | 0.8456 | 0.3202 | 0.8645 |
| ETH vol. $\nRightarrow$ DAI | 1.0194 | 0.3131 | 0.4592 | 0.6320 | 0.4014 | 0.7521 | 0.5235 | 0.7185 |
| DAI $\nRightarrow$ ETH vol. | 0.3906 | 0.5323 | 0.7010 | 0.4965 | 0.7884 | 0.5007 | 0.6632 | 0.6178 |

## 5.3 DAI and the NFT art market

This section aims to interpret DAI and provide a narrative that helps understand its evolution in the context of the emerging NFT art market. By employing hedonic regression, we discount the implicit price derived from the characteristics of an artwork, allowing

DAI to be perceived as a premium or characteristic-free price of NFT artworks, thereby providing a more accurate representation of market trends.

Figure 5.1 presents the evolution of DAI and its return, ranging from April 2019 to March 2022. The index remains relatively stable until March 2020, after which it begins a notable upward trend with several peaks, particularly from early 2021 onwards. The volatility in DAI returns increased alongside the growing popularity of NFTs. In 2020, the COVID-19 pandemic led to an economic recession. As observed by Mei and Moses (2002b), economic recessions and declines in art prices can create investment opportunities, as artworks have the potential to outperform traditional assets such as equities and bonds. This trend is reflected in the growth observed since the end of 2020.

In terms of potential risks and losses, the returns fell below the 90%, 95%, and 99% Value-at-Risk (VaR) thresholds in May and June 2021; that is, the actual market conditions were more volatile and negative than predicted. The VaR values are estimated using a 30-day rolling window of historical DAI returns, assuming a student's t-distribution with 4 degrees of freedom. In Table 3, the price index demonstrates a broad range of values with a large standard deviation. The returns exhibit a high kurtosis, indicating heavy tails, which implies a greater likelihood of extreme values and an increased risk for the NFT art investment. Both the index and returns have medians smaller than their mean values, indicating right-skewness. In particular, the returns exhibit a skewness of 4.630, showing a distribution with a long right tail, which means that there are more frequent large positive returns than negative ones. Thus, the NFT art market attracts investors with a high risk tolerance.

Moreover, we find a moderate autocorrelation in market returns, see Figure 5.2. According to Goetzmann (1995), a persistent trend in the conventional art market, specifically serial dependence in returns, is linked to market inefficiency. Despite advancements, the NFT art market remains relatively inefficient and immature, leading to considerable uncertainty regarding resale values. This prevailing tendency poses a significant price risk for investors. Overall, while the NFT art market presents opportunities for substantial returns, it also entails significant risks and is characterized by frequent boom and bust cycles. Therefore, investors should consider their risk appetite and investment strategy before investing in the NFT market.

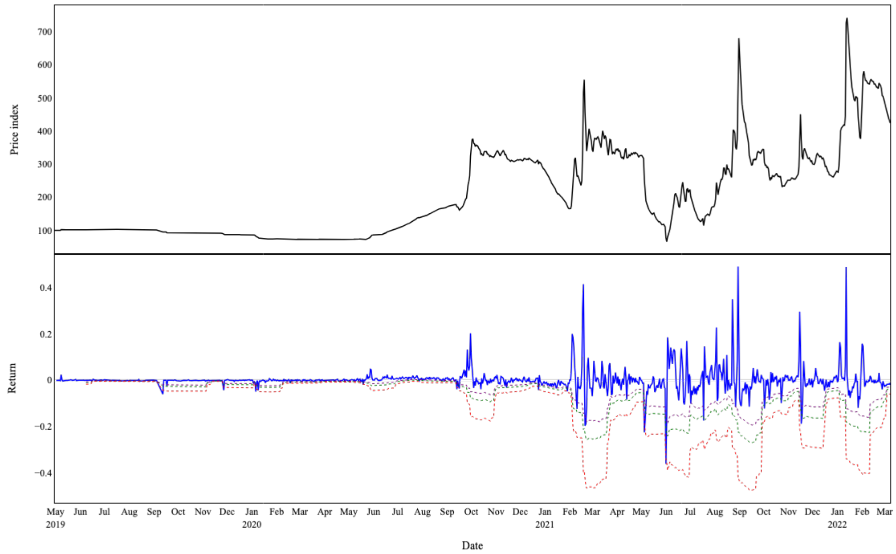



**Figure 5.1: DAI price index and its returns**. The dotted lines represent: 90% VaR, 95% VaR, and 99% VaR.

**Table 3: Descriptive Statistics for DAI and its returns**.

| **Statistic** | **Index** | **Return (%)** |
|---|---|---|
| Mean | 215.602 | 0.293 |
| Standard Deviation | 134.953 | 5.693 |
| Minimum | 65.847 | -30.275 |
| Median | 171.781 | -0.033 |
| Maximum | 742.231 | 63.311 |
| Kurtosis | 0.340 | 45.153 |
| Skewness | 0.904 | 4.630 |

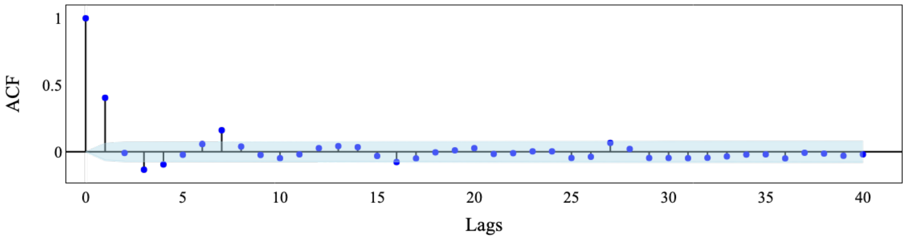



**Figure 5.2: Autocorrelation function for DAI returns with** 95% **CIs**.

# 6 Conclusion

Experiencing the emergence of photography, Benjamin (1968) describes the unique existence of artworks within a limited time period at a specific place, which becomes an untenable category in the era of art reproduction. With the emergence of NFT art and the ease of perfectly reproducing the aesthetics of digital artworks through copying and pasting, it is essential to adopt a different approach when dealing with NFT art.

We propose a Digital Art Index (DAI) using hedonic regression, highlighting the heterogeneity of artworks and the challenges that price inequality in the NFT art market poses to the price discovery process, particularly the significant influence of outlier prices on index construction. To assess the robustness of the hedonic price index, we introduce two innovative alternative methods: Huberization with Kalman filter and DCS-t filtering. We demonstrate that Huberization serves as a one-step robust regression method that can be promptly and efficiently applied. Meanwhile, DCS-t filtering is a stepwise method that incorporates changes in the conditional distribution of an observation in the Kalman filter's state space formulation. Both procedures offer a degree of robustness to the indices; however, in a fast-changing environment with infrequent and volatile observations, DCS-t filtering proves to be more adaptive and straightforward.

By conducting pairwise causality tests between DAI and other market returns, we discover that NFT art can serve as an option for risk diversification alongside cryptocurrencies. However, DAI returns exhibit limited autocorrelation and show high skewness and heavy tails, indicating that the market remains inefficient and immature. The NFT art market is particularly attractive to risk-takers due to its potential for significant gains and substantial losses. Upon examining price determinants observed from hedonic LASSO, we draw conclusions regarding the possible price formulation involving external and internal factors. Notably, external factors such as payment tokens and the frequency of seller addresses have a significant influence on price determination; that is, platforms for selling NFT art play a crucial role. Regarding internal factors, the scarcity of artwork's traits has a limited impact, whereas collection slugs hold relatively higher influence on pricing. Overall, institutionally initiated collections, such as CryptoPunks by Lava Lab, tend to achieve better pricing compared to works from independent creators. Consequently, it remains challenging for solo artists to profit in this market.

# References


Artemova, M., Blasques, F., van Brummelen, J. and Koopman, S. J. (2022). *Score-Driven Models: Methodology and Theory*, Oxford University Press.

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.

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

Chanel, O., Gérard-Varet, L.-A. and Ginsburgh, V. (1996). The relevance of hedonic price indices: The case of paintings, *Journal of Cultural Economics* **20**: 1–24.

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, *The Dynamics of Automobile Demand*, General Motors Corporation, pp. 99–117.

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

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*, Oxford University Press.

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* **21**(6): 25–34.

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

Griliches, Z. (1961). Hedonic price indexes for automobiles: An econometric of quality change, *The Price Statistics of the Federal Goverment*, National Bureau of Economic Research, Inc, pp. 173 – 196.

Griliches, Z. (1971). Introduction: Hedonic price indexes revisited, *Price Indexes and Quality Change*, Harvard University Press, pp. 3–15.

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*, Cambridge University Press.

Harvey, A. C. (2022). Score-driven time series models, *Annual Review of Statistics and Its Application* **9**: 321–342.

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*, Vol. 5, National Bureau of Economic Research, Inc, pp. 475–492.

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

Horky, F., Rachel, C. and Fidrmuc, J. (2022). Price determinants of non-fungible tokens in the digital art market, *Finance Research Letters* **48**: 103007.

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* **28**(1): 11–21.

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

Kraizberg, E. (2023). Non-fungible tokens: A bubble or the end of an era of intellectual property rights, *Financial Innovation* **9**(1): 32.

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* **28**(5): 1139–1198.

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*, John Wiley & Sons.

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, *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). Heterogeneous rarity patterns drive price dynamics in nft collections, *Scientific reports* **12**(1): 13890.

Penasse, J. and Renneboog, L. (2021). Speculative trading and bubbles: Evidence from the art market, *Management Science* **69**(7): 4939–4963.

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, *National Bureau of Economic Research Inc* .

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, *United Nations University - Maastricht Economic and Social Research Institute on Innovation and Technology (MERIT)* .

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

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.

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.

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

In this section, we list all the 391 variables we considered in the hedonic regression. Each variable is derived from NFT artwork transaction data and metadata. These variables are categorized by their type or divided into different intervals for frequency-based data. The number of sales and the CryptoPunks accessory's scarcity are continuous variables, while the remaining variables are one-hot encoded. For example, the Autoglyphs collection includes 6 traits (symbols) that form a work. Each artwork is characterized by a specific symbol scheme, such as "+-|", which is assigned a value of 1, while the unused symbol schemes are assigned a value of 0.

**Table A.1: Data attributes.**

| Type of variables | Auction type | Collection slug | Event type | Payment token | Frequency of same recipient account | Frequency of same seller address | Frequency of same winner address | Contract address |
|---|---|---|---|---|---|---|---|---|
| **Variable** | dutch | Art Blocks | cancelled | ABST | $(1000, \infty]$ | $(1000, \infty]$ | $(500, \infty]$ | 0x0ba51d9c015a7544e3560081ceb16ffe222dd64f |
| | english | Autoglyphs | successful | ASH | $(100, 1000]$ | $(500, 1000]$ | $(100, 500]$ | 0x131aebbfe55bca0c9eaad4ea24d386c5c082dd58 |
| | min_price | Beeple Everydays | | ATRI | $[2, 100]$ | $(100,500]$ | $[2, 100]$ | 0x1f0678d1238921dc99309360668b16a17098aa2a |
| | | Bored Ape Yachtclub | | DAI (crypto) | $= 1$ | $[2,100]$ | $= 1$ | 0x2947f98c42597966a0ec25e92843c09ac17fbaa7 |
| | | CryptoPunks | | ETH | | $= 1$ | | 0x41a322b28d0ff354040e2cbc676f0320d8c8850d |
| | | Hashmasks | | KLTR | | | | 0x7a6425c9b3f5521bfa5d71df710a2fb80508319b |
| | | MakersPlace | | LUX | | | | 0x7be8076f4ea4a4ad08075c2508e481d6c946d12b |
| | | Rarible | | NCT | | | | 0x7e3abde9d9e80fa2d1a02c89e0eae91b233cde35 |
| | | SuperRare | | RARI | | | | 0x7f268357a8c2552623316e2562d90e642bb538e5 |
| | | Wrapped CryptoPunks | | TRSH | | | | 0x8c9f364bf7a56ed058fc63ef81c6cf09c833e656 |
| | | | | USDC | | | | 0x93f2a75d771628856f37f256da95e99ea28aafbe |
| | | | | WETH | | | | 0xa5af48b105ddf2fa73cbaac61d420ea31b3c2a07 |
| | | | | WHALE | | | | 0xb47e3cd837ddf8e4c57f05d70ab865de6e193bbb |
| | | | | nan | | | | 0xcd4ec7b66fbc029c116ba9ffb3e59351c20b5b06 |
| | | | | | | | | 0xf2ee97405593bc7b6275682b0331169a48fedec7 |
| | | | | | | | | nan |

| Type of variables | Asset contract type | Schema name | Frequency of same owner address | Frequency of same creator address | Number of sales | CryptoPunks accessory's scarcity | ArtBlocks section |
|---|---|---|---|---|---|---|---|
| **Variable** | non-fungible | CryptoPunks | $(1000, \infty]$ | $(2000, \infty]$ | Mean | Mean | curated |
| | semi-fungible | ERC1155 | $(500,1000]$ | $(1000,2000]$ | 2.24 | 51.75 | factory |
| | | ERC721 | $(100,500]$ | $(100,1000]$ | Std. | Std. | nan |
| | | | $[2,100]$ | $[2,100]$ | 5.69 | 13.82 | playground |
| | | | $= 1$ | $= 1$ | | | |

Continued on next page

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

| Type of variables | ArtBlocks sub-collection | Autoglyphs trait | Bored Ape background | Bored Ape earring | Bored Ape fur | Bored Ape 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** | Bored Ape eyes | Bored Ape clothes | Bored Ape mouth | Hashmasks item | Hashmasks mask | Hashmasks 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** | Hashmasks skin color | Hashmasks eye color | Hashmasks background | Hashmasks glyph | Frequency of same creator ID | Superrare tag | CryptoPunks type | Created year |
|---|---|---|---|---|---|---|---|---|
| **Variable** | 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 | | |
| | | | | | | 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 Data processing

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 TF-IDF (term frequency–inverse document frequency) in Jones (1972). The traits of the $p$-th punk's sales 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 computed by:

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

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

The IDF of the word is:

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

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:

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

Having a consistent structure of CryptoPunks' traits, the scarcity of each punk is measured by summing the scarcity index of each accessory equipped. $\mathrm{SI}_{T_p}$ of each accessory is the importance of an accessory for a given punk, which is adjusted by the rarity of this accessory owned by the existing punks.

In SuperRare each artwork has a single tag (i.e. a word) or a set of heterogeneous tags in its trait. The number of tags within these traits varies from the others. Most of these tags only appear once. We, thus, apply the pre-trained word embedding model – Global vectors for word representation (GloVe) from Pennington et al. (2014) to handle these inconsistent traits. Based on global word-word co-occurrence statistics from the corpus, the training result shows a linear substructure in word vector space. It implies that the similarity between words can be measured by the linear distance of two resulting word vectors, and these vectors are additive. We create GloVe index for each SuperRare trait $T_s$, $p \in (1, 2, \ldots, S)$,

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

where $f_{w,T_s}$ denotes the raw count of a tag within a given trait; and $\mathrm{GloVe}_{w,T_s}$ is 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 use this index to consider the similarity among works in SuperRare. To further reduce the dimensionality, we apply $k$-means clustering on the GloVe indices and yield the dummy variables to represent the resulting 30 clusters.

## C Simulation for KF and DCS-t filterings

Offering a better resolution to compare them, we conduct simulation with a univariate state space model wherein the state variable follows an AR(1) process, see (3.11) and (3.13). At each time $t$ there exist $N_t$ observations and $t \in \{1, \ldots, 500\}$. Parameters $\varepsilon \overset{\text{iid}}{\sim} \mathcal{N}(0, {\sigma_\varepsilon}^2)$ and $\xi \overset{\text{iid}}{\sim} \mathcal{N}(0, {\sigma_\xi}^2)$ are the error terms for the observable and state variables, respectively. An example with $N_t = [5, 20]$, $\sigma_\varepsilon = 0.1$ and $\sigma_\xi = 0.2$ is shown in Figure C.1.

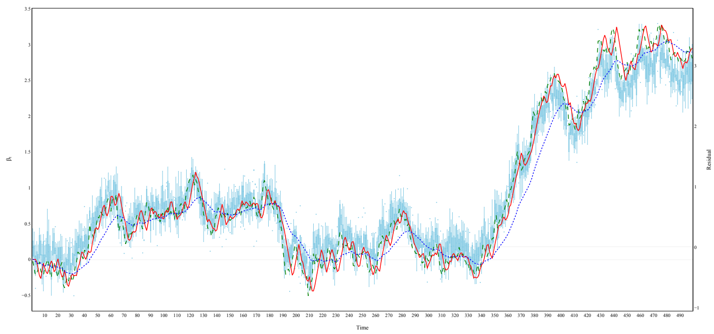



**Figure C.1:** $\beta_t$ **(dashed),** $\hat{\beta}^{\text{KF}}_{t|t}$ **(dotted),** $\hat{\beta}^{\text{DCS-t}}_{t|t-1}$ **(solid) and residual box plot over time (background) with** $N_t = [5, 20]$**,** $\sigma_\varepsilon = 0.1$ **and** $\sigma_\xi = 0.2$.

In Table B.1, the KF achieves a better outcome than the DCS-t in terms of mean squared errors (MSE) if the number of observations are fixed at each time $t$. Under random observations, the DCS-t surpasses the KF while having higher standard deviations in the model. According to the simulation result, we find that the DCS-t adapts better a volatile time series. The implementation of Huberization before applying Kalman filtering avoids deterioration caused by outliers. Huber (1964) shows that minimizing the Huber loss can be interpreted as MLE, i.e. its asymptotic variance

$$\mathsf{E}_F \psi^2(\varepsilon)/[\mathsf{E}_F \psi'(\varepsilon)]^2 \geq 1/\mathsf{E}_F[(f'(\varepsilon)/f(\varepsilon))^2]$$

has the equality when $\psi \propto -f'/f$ where $\varepsilon \sim F$ with density $f$. That is, once we obtain the optimal threshold $\tau$ over time it offers robust estimates of $\beta_t$. However, an optimizing update like this can be computationally costly.

**Table B.1: DF and DCS-t simulation results.**

| $N_t$ | $\sigma_\varepsilon$ | $\sigma_\xi$ | MSE | |
|---|---|---|---|---|
| | | | **KF** | **DCS-t** |
| 5 | 0.01 | 0.02 | 2.00e-04 | 2.64e-04 |
| 5 | 0.1 | 0.2 | 1.03e-02 | 2.56e-02 |
| 50 | 0.01 | 0.02 | 2.58e-03 | 4.17e-03 |
| 50 | 0.1 | 0.2 | 1.95e-01 | 2.19e-02 |
| [5, 20] | 0.01 | 0.02 | 7.66e-04 | 2.33e-04 |
| [5, 20] | 0.1 | 0.2 | 7.44e-02 | 2.75e-02 |
| [20, 100] | 0.01 | 0.02 | 2.72e-03 | 2.34e-04 |
| [20, 100] | 0.1 | 0.2 | 2.16e-01 | 2.45e-02 |

# D Price determinants

Here, we present the variables that retain non-zero coefficients after LASSO regularization, indicating their significant contributions to the model. These key predictors highlight their relevance in the underlying data patterns and their impact on NFT art pricing. The red cells represent variables with a positive influence, while the blue cells indicate variables with a negative impact. The external factors (Figure C.1) refer to variables related to transactional and contextual aspects, while the internal factors (Figure C.2) pertain to the artwork's collection and intrinsic characteristics.

**Table C.1: External factors.**

| Type of variables | Variable | Coeff. |
|---|---|---|
| Auction type | English | 0.5364 |
| Event type | Successful | -1.2119 |
| Payment token | DAI (crypto) | 4.5703 |
| | ETH | 1.6079 |
| | WETH | 1.3497 |
| Frequency of same seller address | (500, 1000] | 0.3941 |
| | (100, 500] | 0.8425 |
| | [2, 100] | 1.0586 |
| | = 1 | 0.8854 |
| Frequency of same 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 |

Continued on next page

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

| Type of variables | Variable | Coeff. |
|---|---|---|
| Frequency of same owner address | (500, 1000] | 0.5347 |
| | (100, 500] | 0.2085 |
| | =1 | -0.3033 |
| Frequency of same creator address | (1000, 2000] | -0.7539 |
| | (100, 1000] | -1.1146 |
| | [2, 100] | -1.1746 |
| | =1 | -0.6828 |
| Frequency of same creator ID | (100, 1000] | -0.1855 |
| Number of sales | | 0.0023 |

**Table C.2: Internal factors.**

| Type of variables | Variable | Coeff. |
|---|---|---|
| Collection slug | Autoglyphs | 2.9271 |
| | Beeple-everydays | 2.4190 |
| | Bored Ape Yacht Club | 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 sub-collection | 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 |
| Bored Ape background | Aquamarine | 0.0337 |
| | Gray | 0.0022 |
| Bored Ape earring | Cross | 0.0136 |
| | Gold Stud | 0.0142 |
| | Silver Stud | 0.0305 |

Continued on next page

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

| Type of variables | Variable | Coeff. |
| --- | --- | --- |
| Bored Ape 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 |
| Bored Ape 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 |
| Bored Ape 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 |

Continued on next page

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

| Type of variables | Variable | Coeff. |
|---|---|---|
| Bored Ape clothes | Black Suit | 0.2845 |
| | Black T | -0.0118 |
| | Guayabera | -0.0083 |
| | Hip Hop | 0.0042 |
| | Pimp Coat | 0.0093 |
| | Tie Dye | 0.0450 |
| Bored Ape 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 |
| Hashmasks item | Golden Toilet Paper | 0.2682 |
| | No Item | -0.6320 |
| | Shadow Monkey | 0.1354 |
| Hashmasks mask | Abstract | 0.0270 |
| | Animal | 0.0388 |
| | Basic | 0.0156 |
| | Doodle | -0.0598 |
| | Indian | -0.0169 |
| | Steampunk | 0.0755 |

Continued on next page

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

| Type of variables | Variable | Coeff. |
|---|---|---|
| Hashmasks character | Female | -0.0339 |
| | Golden Robot | 0.9744 |
| | Male | -0.0508 |
| | Mystical | 1.0136 |
| | Puppet | 0.1841 |
| Hashmasks skin color | Dark | -0.0212 |
| | Freak | 0.1429 |
| Hashmasks eye color | Blue | -0.3567 |
| | Dark | -0.4657 |
| | Green | -0.3512 |
| | Heterochromatic | 0.0739 |
| Hashmasks background | Doodle | -0.0065 |
| | Pixel | 0.0492 |
| Superrare tag | cluster_10 | 0.0369 |
| | cluster_18 | -0.3642 |
| | cluster_2 | -0.0648 |
| | cluster_22 | 0.0207 |
| | cluster_23 | 0.0113 |
| CryptoPunks type | Male | -0.1262 |
| Created year | 2019 | -2.8126 |
| | 2020 | -1.5785 |
| | 2021 | -1.0743 |
| CryptoPunks' accessory scarcity | | 0.0011 |