IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.
Digital Object Identifier 10.1109/ACCESS.2024.0429000

Draft version

NFT Market Regimes and Cross-Sectional Alpha Heterogeneity: Evidence from Structural Clusters

SATOSHI KONDO[1], RYOHEI BANNO[2], (Member, IEEE)
[1]Independent Researcher, Tokyo, Japan (e-mail: satoshi.kndo@gmail.com)
[2]Graduate School of Social Data Science, Hitotsubashi University, Kunitachi, Tokyo 186-8601, Japan (e-mail: banno@computer.org)

This work was partly supported by JSPS KAKENHI Grant No. 24H00691. The research reported in this paper is based on work conducted while Satoshi Kondo was enrolled in the Graduate School of Social Data Science, Hitotsubashi University.

ABSTRACT This paper studies regime-dependent cross-sectional alpha heterogeneity in NFT markets using weekly Ethereum secondary-market trades. We identify three broad market regimes (*pre-boom*, *boom*, and *post-boom*) using change-point detection and estimate collection-level alpha, defined as the intercept under a parsimonious baseline factor specification comprising an NFT market factor and an ETH/USD factor. We use this alpha as a descriptive benchmark and show that its cross-sectional dispersion is most pronounced in the *boom* regime, remains detectable but weaker in the *post-boom* regime, and is comparatively fragile in the *pre-boom* regime. To organize this heterogeneity, we construct structural clusters using collection-level features that summarize participation dispersion, trading persistence, concentration, and price-distribution properties. These clusters provide a useful descriptive organization of alpha differences, especially in the *boom* regime. Feature-level analyses further show that persistence and price-dispersion measures are, on average, positively associated with alpha, whereas concentration measures are negatively associated. Supplementary analyses indicate that observed collection categories contain some additional information, but their incremental contribution beyond structural clusters is generally modest in the central analyses. Overall, the paper provides a regime-aware descriptive framework for organizing cross-sectional heterogeneity in NFT markets, most clearly in the *boom* regime and more modestly in the *post-boom* regime. The findings should be interpreted as conditional associations under the baseline specification rather than as causal effects or evidence for a definitive asset-pricing model.

INDEX TERMS alpha, blockchain, change-point detection, clustering, factor model, market structure, NFTs, regime shift

I. INTRODUCTION

NON-FUNGIBLE TOKENS (NFTs) have emerged as a major class of blockchain-based digital assets spanning art, collectibles, gaming items, memberships, and utility-linked tokens. As NFT markets expanded through large secondary marketplaces such as OpenSea, they exhibited return patterns that differ markedly from those of traditional financial assets, including extreme volatility, strong concentration, rapid project turnover, and pronounced boom-bust dynamics [1]–[7]. These features suggest that NFT returns should not be understood solely through aggregate market behavior. Even within the same broad market environment, collections differ substantially in breadth of participation, trading persistence, concentration, and price dynamics.

A growing empirical literature has examined NFT pricing and return behavior from multiple perspectives, including market concentration, co-movement with crypto assets, sale mechanisms, item attributes, participant behavior, and attention dynamics [8]–[14]. However, less is known about how collection-level residual performance varies across market regimes and whether such heterogeneity can be organized systematically and interpreted clearly in markets shaped by symbolic value, community dynamics, and platform-specific trading structures.

The paper's contribution is a regime-aware structural organization framework for collection-level NFT heterogeneity rather than a new pricing, forecasting, or causal identification design. We use collection-level alpha, defined as the residual intercept under a parsimonious baseline factor specification, as a descriptive benchmark and ask whether its heterogeneity differs across market regimes and can be usefully organized by structural clusters and features. Observed collection categories are treated as a supplementary descriptive attribute rather than as the primary organizing principle.

The paper makes three contributions. First, it documents that collection-level alpha heterogeneity is regime-dependent: cross-sectional differences are most visible in the *boom* period, remain detectable but weaker in the *post-boom* period, and are comparatively fragile in the *pre-boom* period. Second, it shows that structural clusters constructed from collection-level market features provide a useful descriptive organization of this heterogeneity. Third, it provides feature-level evidence that participation dispersion, trading persistence, and concentration are systematically associated with cross-sectional alpha differences in a descriptive rather than causal sense.

This study is descriptive in scope. The reported relationships should be interpreted as conditional associations rather than causal effects, and the baseline factor model should be understood as a parsimonious benchmarking device rather than a definitive asset-pricing specification. In addition, the structural clusters should be interpreted as an ex post taxonomy that organizes observed residual performance heterogeneity, not as real-time predictors, trading signals, causal mechanisms, or unique latent classes of NFT collections. The analysis, therefore, aims to summarize and interpret cross-sectional variation across regimes, rather than to provide a predictive classification system or a fully specified structural model of NFT returns.

The remainder of the paper is organized as follows. Section II reviews related literature. Section III describes the data and preprocessing. Section IV presents the empirical design. Section V reports the main results. Section VI discusses interpretation and limitations. Section VII concludes.

II. RELATED WORK

Prior research has shown that NFT markets are highly uneven, with a strong concentration of activity and substantial heterogeneity across collections and participants. Using large-scale transaction data, Nadini et al. (2021) document rapid market expansion together with extreme concentration in trading volume and market outcomes [8]. Related work also finds that NFT markets exhibit partial co-movement with broader crypto-asset markets while retaining market-specific dynamics [9]. Taken together, these studies suggest that NFT markets are not well characterized by a single representative price process and instead motivate a collection-level perspective.

A second line of work examines NFT pricing and return dynamics. Prior studies report that NFT outcomes are associated with market conditions, project-level characteristics, liquidity, trading activity, speculative participation, and attention dynamics [10], [12], [15]–[19]. For some NFT segments, aesthetic and item-specific characteristics also appear to matter [20]. More broadly, this literature suggests that NFT returns reflect a mixture of market-wide and collection-specific dimensions rather than broad crypto-market exposure alone. More generally, the asset-pricing literature has long emphasized the trade-off between parsimony and completeness in factor-model design [21]. In this paper, we therefore treat the baseline factor specification as a benchmarking device rather than as a definitive model of NFT expected returns.

A third relevant strand concerns time variation in market conditions. NFT markets display pronounced boom-bust behavior, and the broader econometric literature has long emphasized that financial series may transition across distinct regimes over time [22]. Recent review-based work also highlights the breadth of empirical findings on investor behavior and market dynamics in NFT markets, which further underscores the need for more focused organizing frameworks for collection-level heterogeneity [23]. Yet, while prior NFT studies document concentration, pricing frictions, speculation, and attention effects, less is known about how collection-level residual performance varies across major market regimes and whether this variation can be organized using structural characteristics observed at the collection level.

This paper contributes to that gap. Rather than offering a broad review of NFT investor behavior or proposing a new asset-pricing theory, we provide a regime-aware descriptive framework for studying cross-sectional alpha heterogeneity across NFT collections. Our focus is on whether structural clusters and features—such as participation dispersion, trading persistence, and concentration—provide a useful way to organize this heterogeneity. Observed categories are therefore treated as a supplementary attribute rather than as the central object of comparison.

III. DATA AND PREPROCESSING

We study secondary-market NFT trades on the Ethereum blockchain. Each trade record contains at least a timestamp, the transaction price in ETH, the buyer and seller wallet addresses, the token identifier, and the collection identifier. To account for ETH/USD conversion effects in return construction, we merge daily ETH/USD reference prices from Etherscan [24]. All variables are aggregated at a weekly frequency, which reduces microstructure noise in trade-level observations and provides a common time unit for regime identification and factor model estimation.

The primary unit of analysis is the collection. For each collection, we aggregate trade-level observations into weekly outcomes such as median price, trading value, and return series, and we construct collection-level measures of participation breadth, concentration, and trading persistence. The empirical analysis is based on a versioned processed dataset constructed from the underlying blockchain records and the associated preprocessing pipeline. Additional implementation details, including identifier mapping, protocol coverage, time alignment, and handling of missing fields, are reported in Appendix A.

Because NFT markets are highly concentrated and heavy-tailed [8], [9], preprocessing is designed to be minimal and deterministic. We remove clearly invalid records, apply basic outlier filtering to trade prices, and then construct weekly collection-level aggregates. This procedure is intended as a basic data-cleaning step rather than as a comprehensive suspicious-trading detection system. Exact filtering rules and supplementary diagnostics are reported in Appendix A.

For return construction, each collection week is summarized using the median USD transaction price on a common ISO-week calendar. Weekly collection returns are then computed from consecutive weekly observations without imputing across trading gaps. We also construct weekly market-level series that are used later for regime identification and the baseline factor model. To preserve broad market coverage while maintaining minimum comparability in collection-level estimation, the baseline alpha analysis focuses on collections with at least eight valid weekly observations within the relevant estimation window; stricter sample restrictions are examined only as robustness checks.

We use the term *prefiltered dataset* for the trade-level dataset after basic invalid-trade and outlier filtering, and the term *filtered analysis sample* for the collection-level sample that additionally satisfies the active-weeks restriction used in the main alpha and clustering analyses.

Table 1 reports summary statistics for the prefiltered dataset after basic invalid-trade and outlier filtering, before applying the active-weeks restriction used to construct the filtered analysis sample. The sample contains 36,150,388 trades, 59,730 collections, and 3,200,545 wallet addresses, spanning 2017–10–19 to 2025–03–31. Figure 1 provides a market-level overview of weekly NFT prices, trading activity, and ETH/USD dynamics. Extended distributional diagnostics and concentration measures are deferred to Appendix A.

TABLE 1. Summary statistics of the prefiltered NFT dataset after basic invalid-trade and outlier filtering.

Metric	Value
Total Trades	36,150,388
Number of Collections	59,730
Number of Wallets	3,200,545
Start Date	2017–10–19
End Date	2025–03–31
Median Price (USD)	146.76
Average Price (USD)	1,470.62
Average Weekly Volume (USD)	136,666,867.01

IV. EMPIRICAL DESIGN

This section describes the empirical design used to study regime-dependent cross-sectional alpha heterogeneity in NFT collections. The design has four components: regime segmentation, baseline alpha estimation, structural clustering, and cross-sectional association analysis. Throughout, the empirical objective is descriptive. The baseline factor model is used as a benchmarking device, structural clusters are interpreted as descriptive summaries of observable market structure, and the reported feature–alpha relationships are interpreted as conditional associations rather than causal effects.

A. REGIME SEGMENTATION

To capture major changes in NFT market conditions over time, we construct weekly market indicators summarizing market scale, participation intensity, and price-level dynamics. We then apply multiple change-point detection to these series to identify candidate break dates. Rather than treating algorithmic outputs as mechanically final, we use them as inputs to a transparent baseline partition that emphasizes cross-indicator alignment and interpretability. Detailed implementation choices and sensitivity checks are reported in Appendix B.

The baseline analysis uses three regimes: *pre-boom*, *boom*, and *post-boom*. Specifically, weeks before 2021–01–18 are classified as *pre-boom*, weeks from 2021–01–18 to 2022–05–23 as *boom*, and weeks on or after 2022–05–23 as *post-boom*. This three-regime partition is used as an empirical organizing device rather than as a claim that these boundaries are uniquely determined by the data.

TABLE 2. Market regime boundary dates.

Regime transition	Date
`pre-boom` → `boom`	2021–01–18
`boom` → `post-boom`	2022–05–23

Figure 2 reports the corresponding change-point evidence for key market indicators.

B. BASELINE FACTOR MODEL AND ALPHA ESTIMATION

For each collection i, weekly return $r_{i,t}$ is modeled as

$$r_{i,t} = \alpha_i + \beta_i^{(m)} r_{m,t} + \beta_i^{(fx)} r_{fx,t} + \varepsilon_{i,t}, \tag{1}$$

where $r_{m,t}$ is the NFT market return factor, $r_{fx,t}$ is the ETH/USD return factor, and α_i is the collection-level residual component under this baseline specification.

Collection-level weekly returns are computed as log returns relative to the median weekly USD transaction price, using only consecutive weekly observations and treating trading gaps as missing. This return construction also induces an economically meaningful selection effect. Collections with sparse or intermittent trading contribute fewer return observations and may be underrepresented in the alpha estimation stage. As a result, the estimated alpha distributions may be more reflective of relatively continuously traded collections, while illiquidity, stale pricing, and disappearance from active trading are only partially captured. In this setting, non-trading periods are economically meaningful rather than purely missing observations. The NFT market factor is constructed as a lagged-volume-weighted average of collection-level weekly returns. The factor-estimation sample is restricted to collections with at least eight valid weekly observations within the relevant estimation window.

The purpose of this model is not to provide a definitive asset-pricing specification for NFTs. Instead, it serves as a parsimonious and transparent baseline for partialling out broad NFT-market co-movement and ETH/USD conversion effects before examining residual cross-sectional differences across collections. Accordingly, alpha is interpreted as a descriptive residual benchmark under the baseline factor spec-

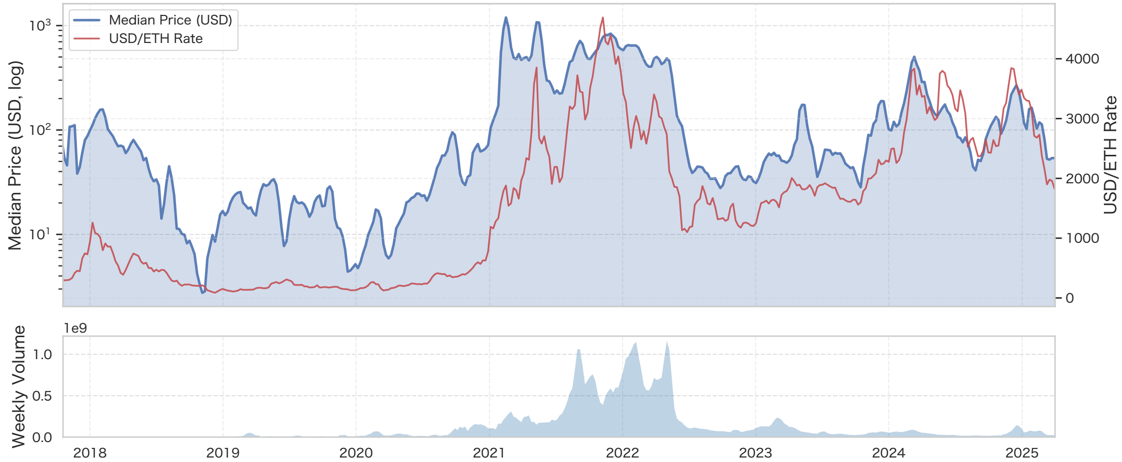

FIGURE 1. **Market overview: weekly NFT price and trading activity with ETH/USD dynamics.**

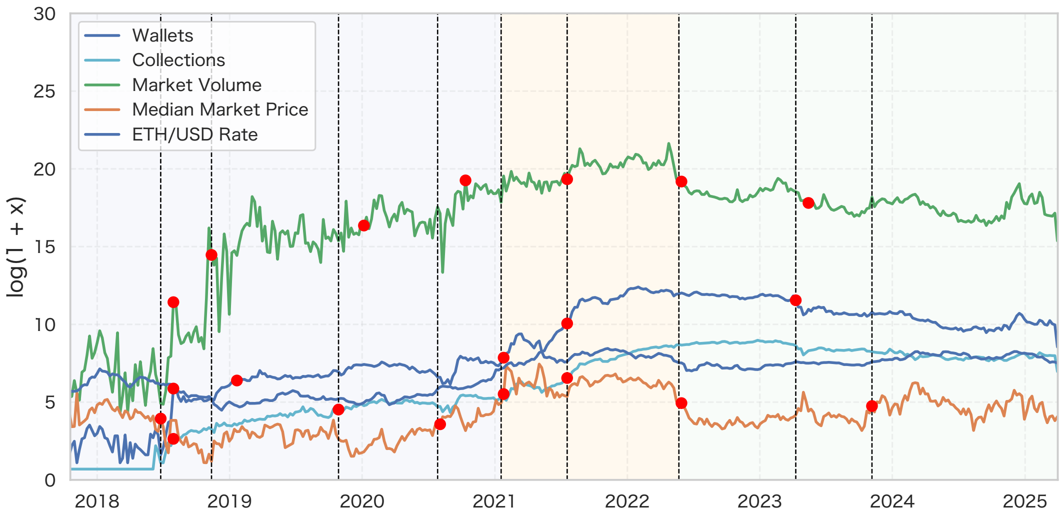

FIGURE 2. **Change-point detection results for key market indicators (candidate break dates).**

ification rather than as a fully purified measure of abnormal performance or collection-specific skill. These residuals may still absorb omitted systematic components, including liquidity conditions, speculative attention, volatility, momentum, platform-level effects, and broader crypto-market fragility. The baseline specification is therefore intended to provide a transparent and comparable residualization framework, not to exhaustively span all systematic sources of NFT return variation. Estimation details and alternative factor specifications are reported in Appendix C and the robustness section.

C. STRUCTURAL FEATURES AND CLUSTERING

To organize cross-sectional heterogeneity beyond broad market exposure, we construct collection-level structural features that summarize four domains: holder structure, trading activity and persistence, price distribution properties, and lightweight project attributes. The feature overview is reported in Table 3. Additional preprocessing details are reported in Appendix D.

After preprocessing and standardization, we apply k-means clustering to obtain descriptive market-structure typologies. The baseline number of clusters is set to $k = 8$ to balance interpretability and avoidance of over-fragmentation, while sensitivity checks for alternative values are reported in Appendix D. Clustering is restricted to collections with sufficiently stable trading histories. These clusters are interpreted as descriptive structural summaries rather than as uniquely identified latent types.

D. CROSS-SECTIONAL ASSOCIATION ANALYSIS

To summarize which structural dimensions are most closely associated with alpha heterogeneity, we combine feature selection with regime-wise cross-sectional regressions. First, we use Lasso as a dimension-reduction device on the structural feature set. Second, we estimate regime-wise regressions in which collection-level alpha is related to cluster indicators

TABLE 3. Overview of structural features used for clustering.

Group	Variable	Description
Holder structure		
unique_holders	Number of unique holders	Number of distinct wallet addresses that have held the collection
unique_holder_ratio	Unique-holder ratio	Ratio of unique holders to total supply
top10_share	Top-10 holder share	Token share held by the top 10 wallets
holder_hhi	Holder concentration (HHI)	HHI-based concentration index computed from holder token shares
buyer_hhi	Buyer concentration (HHI)	HHI-based concentration index across buyer wallets
seller_hhi	Seller concentration (HHI)	HHI-based concentration index across seller wallets
longterm_holder_ratio	Long-term holder ratio	Share of holders who keep holding for at least a specified duration
Trading activity		
trading_days	Trading days	Cumulative number of distinct days with observed trades
active_weeks	Active weeks	Number of weeks with at least one trade
trading_week_span	Trading-week span	Number of weeks between the first and last observed trade
trades	Total trades	Total number of trades during the observation period
daily_trades	Average daily trades	Average number of trades per trading day
buyer_seller_ratio	Buyer-to-seller ratio	Ratio of unique buyers to unique sellers
Price distribution properties		
p25_price	25th percentile price	First quartile of the trade-price distribution
median_price	Median price	Median of the trade-price distribution
p75_price	75th percentile price	Third quartile of the trade-price distribution
stddev_price	Price standard deviation	Standard deviation of trade prices (price dispersion)
diff_price	Price IQR width	Interquartile range width, $p75 - p25$
Project attributes (metadata)		
description_length	Description length	Number of tokens in the collection description
has_project_url	Has project URL	Indicator for an official website URL (0/1)
has_twitter	Has Twitter	Indicator for an official Twitter link (0/1)
has_discord	Has Discord	Indicator for an official Discord link (0/1)
has_instagram	Has Instagram	Indicator for an official Instagram link (0/1)
has_telegram	Has Telegram	Indicator for an official Telegram link (0/1)
is_erc721	ERC-721 token	Indicator for ERC-721 standard (0/1)
is_erc1155	ERC-1155 token	Indicator for ERC-1155 standard (0/1)
has_erc2981	ERC-2981 support	Indicator for ERC-2981 royalty support (0/1)
royalty_fee_percent	Royalty fee rate	Declared royalty fee rate (%)

and selected structural features:

$$\alpha_i = \beta_0 + \sum_{k=1}^{K-1} \gamma_k \, \mathbb{I}(\text{Cluster}_i = k) + \mathbf{X}_i' \boldsymbol{\delta} + \varepsilon_i, \qquad (2)$$

where $\mathbf{X}_i$ denotes the selected structural covariates. Supplementary weighted regressions are used as robustness checks when the precision of alpha estimation varies with the length of the available return history.

These regressions are descriptive. They are intended to summarize cross-sectional associations between structural characteristics and collection-level alpha, not to establish causal effects.

E. SUPPLEMENTARY ROLE OF OBSERVED CATEGORIES

Observed categories are treated as a supplementary descriptive attribute rather than as the main organizing principle of the paper. Specifically, we use OpenSea-derived category labels to describe cluster composition and to examine whether category information adds modest incremental explanatory power beyond structural clusters in auxiliary regressions. Supplementary category-based results are reported in the robustness section and appendices.

Observed category coverage is materially higher in the filtered analysis sample than in the broader prefiltered collection universe, which makes the auxiliary category-based checks more informative in the retained sample, even though the category variable remains incomplete and supplementary by design.

F. ROBUSTNESS DESIGN

The robustness analyses address several concerns raised by the empirical design. First, we compare the baseline USD two-factor model with alternative return/factor specifications, including a USD market-only model, an ETH-denominated market-only model, and a USD specification using orthogonalized FX returns. Second, to address possible forward-looking contamination from full-history features, we compare full-history clustering with regime-specific clustering. Third, we examine whether the baseline clustering partition is sensitive to algorithmic choices, including a comparison of Gaussian mixture models reported in Appendix D. Fourth, we examine whether observed categories provide incremental information beyond structural clusters, including labeled-only and weighted specifications. Finally, we assess sensitivity to supplementary suspicious-trade filtering and to a stricter 52-week sample restriction. Detailed results are reported in the robustness section and appendices.

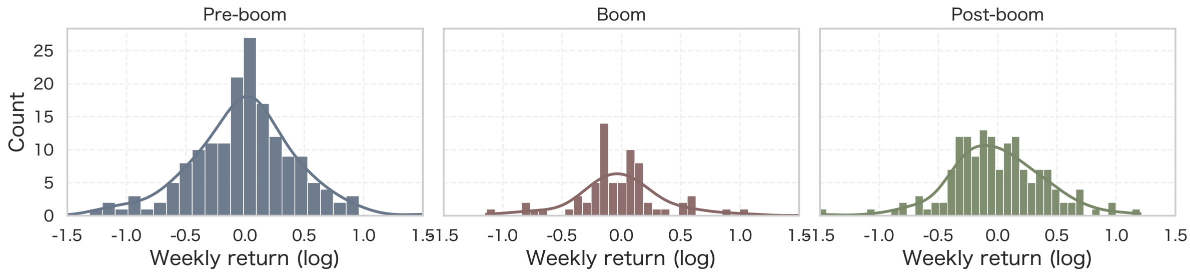

FIGURE 3. **Weekly market return distributions across regimes (*pre-boom, boom, post-boom*).**

V. RESULTS

This section presents four sets of results. We first show that the baseline regime partition captures meaningful time variation in market conditions and in the cross-section of estimated collection-level alpha. We then examine whether structural clusters provide a useful descriptive organization of alpha heterogeneity across regimes. Next, we summarize feature-level associations with estimated alpha using Lasso selection and regime-wise cross-sectional regressions. Finally, we report key robustness evidence, including alternative factor specifications, regime-specific clustering, algorithmic sensitivity of the clustering partition, and supplementary category-based checks.

A. MARKET REGIMES AND CROSS-SECTIONAL ALPHA HETEROGENEITY

Using the change-point evidence summarized in Figure 2, we partition the sample into three baseline regimes: *pre-boom*, *boom*, and *post-boom*. These regimes correspond to distinct combinations of market scale, participation intensity, and price dynamics.

Figure 3 compares weekly market return distributions across regimes. The *boom* regime exhibits substantially wider dispersion and heavier tails than the other periods, whereas the *post-boom* regime shows a more compressed return distribution. These differences support treating the NFT market as regime-dependent rather than as a single stationary environment.

We next estimate the baseline factor model in Eq. (1) within each regime and examine the cross-sectional distributions of collection-level alpha and factor exposures. Figure 4 reports regime-wise kernel densities for α, $\beta^{(m)}$, and $\beta^{(fx)}$.

The cross-sectional distribution of estimated alpha is strongly regime-dependent. Dispersion is widest in the *boom* regime and contracts sharply in the *post-boom* regime. By contrast, changes in factor-exposure distributions are less pronounced. This suggests that an important part of regime dependence appears in the residual cross-section rather than only in common-factor exposure. Because alpha is estimated under a parsimonious benchmark specification, these distributions should be interpreted as residual performance patterns rather than as fully purified abnormal returns.

The *pre-boom* regime is comparatively fragile due to its smaller effective sample size and thinner trading environment. We retain it as an early-market benchmark showing that structural organization is not uniformly informative across all market states, but place the main substantive emphasis on the *boom* and *post-boom* periods.

B. STRUCTURAL CLUSTERS AND REGIME-DEPENDENT ALPHA ORGANIZATION

We next examine whether structural clusters provide a useful descriptive organization of alpha heterogeneity. Using the structural features summarized in Table 3, we cluster collections into market-structure typologies. The clustering is used as an ex post descriptive taxonomy: it organizes collections by observed structural features but is not intended as a real-time classification rule, predictive trading signal, causal mechanism, or unique latent segmentation. Figure 5 shows that these clusters are not confined to a single market regime, indicating that the cluster solution is not merely reproducing the regime partition itself.

Figure 6 reports regime- and cluster-wise alpha distributions. Estimated alpha differs across clusters in both median location and tail thickness. Although the distributions are not perfectly separated, the cluster-level differences are most clearly visible in the *boom* regime and remain detectable, though weaker, in the *post-boom* regime.

This visual evidence is consistent with the supplementary fit comparisons reported in Appendix E. Cluster-based organization is most consistently visible in the *boom* regime, remains detectable but more modest in the *post-boom* regime, and is less stable in the *pre-boom* regime. The magnitude of this organization is limited. In the weighted cluster-only fit comparison, cluster indicators account for $R^2 = 0.0134$ of the cross-sectional variation in estimated alpha in the *boom* regime and $R^2 = 0.0032$ in the *post-boom* regime. The corresponding *pre-boom* value is higher at $R^2 = 0.0614$, but this estimate is based on a much smaller and less stable sample and is therefore interpreted cautiously. These values confirm that the structural taxonomy summarizes only a modest portion of alpha heterogeneity, leaving substantial residual variation within clusters.

The resulting clusters range from thin-liquidity, highly concentrated groups to broad-participation, long-lived market-core collections. A detailed cluster interpretation table and

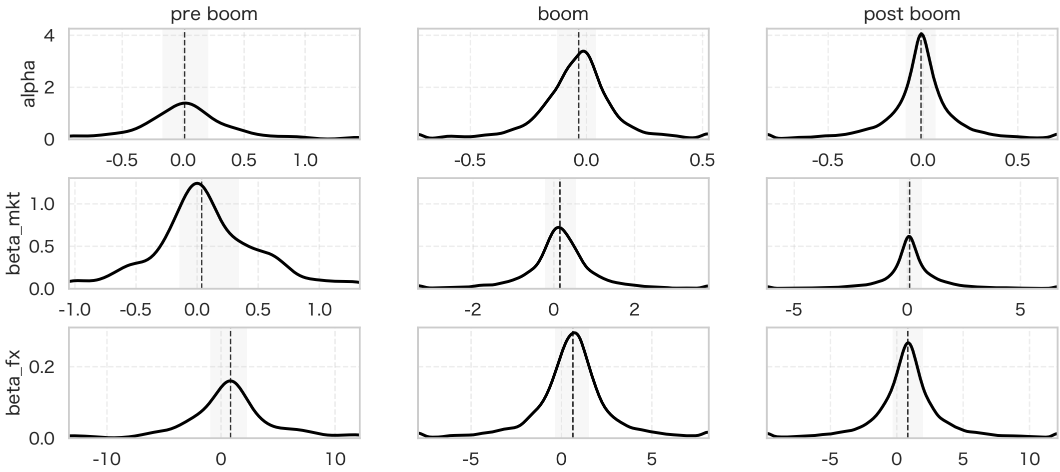

FIGURE 4. Regime-wise KDEs of idiosyncratic alpha α and factor exposures ($\beta^{(m)}$, $\beta^{(fx)}$). Vertical dashed lines indicate medians.

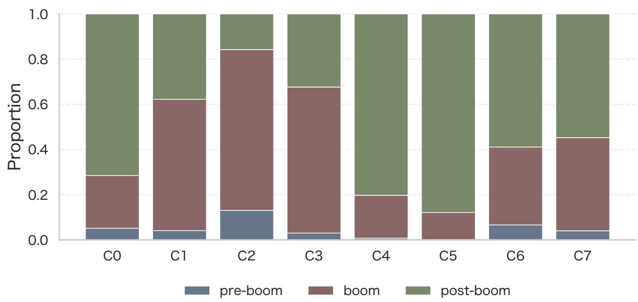

FIGURE 5. Cluster composition by market regime. Each cluster contains collections from multiple regimes, indicating that the cluster solution is not merely a reproduction of the regime partition.

an algorithmic sensitivity check based on Gaussian mixture model clustering are reported in Appendix D. The GMM comparison indicates limited-to-moderate agreement with the baseline k-means partition, reinforcing the interpretation of the clusters as a transparent descriptive taxonomy rather than a uniquely determined segmentation. Observed category composition within clusters is also mixed rather than one-to-one, which suggests that the structural clusters are not simply reproducing broad platform-aligned category labels; a supplementary composition summary is reported in Appendix D.

C. FEATURE-LEVEL ASSOCIATIONS WITH ALPHA

We next examine which structural dimensions are most closely associated with cross-sectional differences in estimated alpha. Figure 7 summarizes the structural features selected by Lasso. The selected set combines a relatively stable core of variables related to trading persistence, concentration, and price-distribution properties, together with additional signals from token-standard, holder-breadth, and some metadata-related variables. Among the selected variables, concentration measures tend to enter with negative coefficients.

Collections with greater trading persistence and wider price dispersion tend, on average, to exhibit higher estimated alpha, whereas more concentrated ownership or trading structures tend to exhibit lower estimated alpha. Additional variables also appear, but these should be interpreted as descriptive correlates rather than causal drivers or out-of-sample predictors.

Regime-wise cross-sectional regressions in Eq. (2) reinforce the same descriptive pattern. Cluster indicators and selected structural features account for a modest but non-negligible share of the estimated alpha cross-section, with explanatory gains most consistently concentrated in the *boom* regime. In the weighted regime-wise comparison, the cluster-only R^2 values are 0.0614 in the *pre-boom* regime, 0.0134 in the *boom* regime, and 0.0032 in the *post-boom* regime. The conditional incremental contribution of clusters given observed categories is positive in all regimes, with $\Delta R^2 = 0.0645$ in the *pre-boom* regime, 0.0109 in the *boom* regime, and 0.0025 in the *post-boom* regime. These values indicate that the structural variables organize part of the cross-sectional dispersion in estimated alpha but do not provide a high-explanatory-power model of collection performance. Full regression tables, including weighted variants, are reported in Appendix E.

Because collection-level alphas are estimated with heterogeneous precision, these cross-sectional analyses should be viewed as descriptive summaries of estimated residual performance rather than inference on latent true alphas. Although precision-weighted variants are reported as robustness checks, fully propagating first-stage estimation uncertainty into clustering and regime-level comparisons remains beyond the scope of this study.

D. SUPPLEMENTARY CATEGORY AND ROBUSTNESS EVIDENCE

Observed categories are used as supplementary descriptive information rather than as the central organizing variable. Auxiliary analyses show that they retain some descriptive

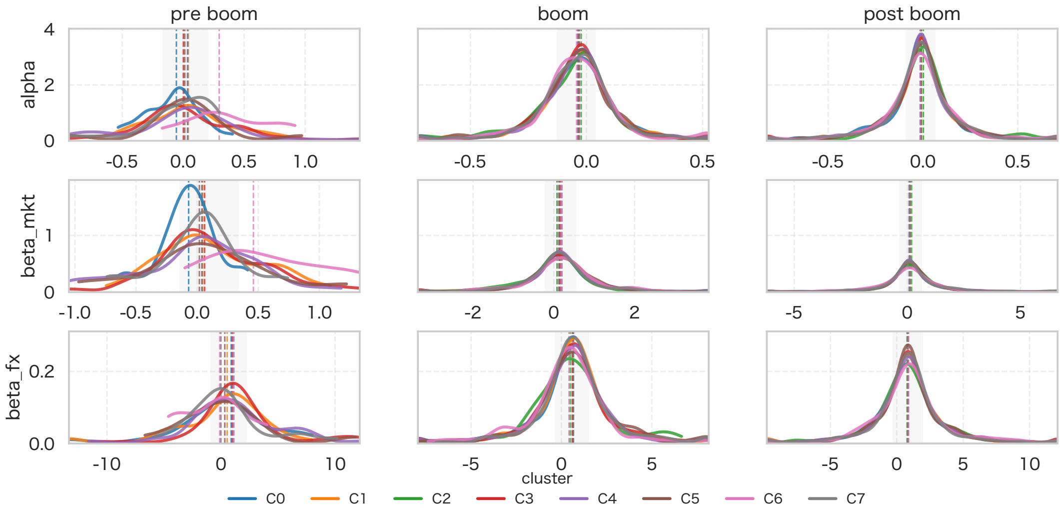

FIGURE 6. Regime- and cluster-wise KDEs of idiosyncratic alpha α and factor exposures ($\beta^{(m)}$, $\beta^{(fx)}$). Vertical dashed lines indicate medians.

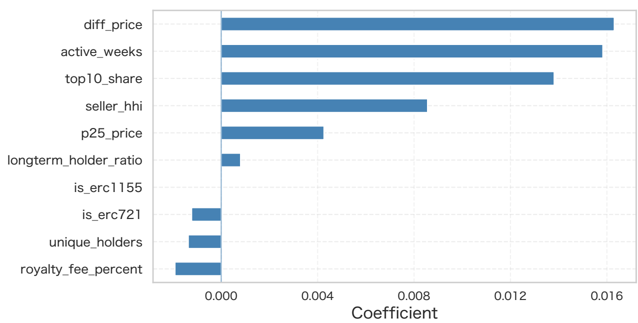

FIGURE 7. Structural features selected by Lasso (non-zero coefficients). All candidate features were standardized to z-scores prior to estimation, so coefficient magnitudes are comparable across predictors.

information about alpha heterogeneity, but their incremental contribution beyond structural clusters is modest. In pooled weighted specifications, the regime-only baseline has $R^2 = 0.0035$, while adding structural clusters increases the fit to $R^2 = 0.0066$ and adding both categories and clusters increases it to $R^2 = 0.0090$. The conditional incremental contribution of clusters given categories is $\Delta R^2 = 0.0026$, which is positive but small. Coverage of observed categories is also substantially higher in the filtered analysis sample than in the broader prefiltered collection universe, so the supplementary category-based checks are not driven solely by extreme sparsity.

The main regime-dependent cluster-based pattern remains broadly similar across alternative return/factor specifications. Cluster-based organization is most consistently supported in the *boom* regime, remains visible but weaker in the *post-boom* regime, and is more sensitive in the *pre-boom* regime. The low correlation between the NFT market factor and the FX factor suggests that the two-factor specification is not driven by simple double-counting of exchange-rate movements. Regime-specific clustering and labeled-only checks yield the same broad interpretation, while observed categories retain some additional incremental signal, especially within the labeled-only subsample.

As a further robustness check, we re-estimated the category-based comparisons after excluding the sizeable `none` group and restricting attention to labeled collections only. The pooled and *boom* regime conclusions remained qualitatively similar, while the *pre-boom* results were more sensitive to sample definition. This reinforces our interpretation that observed categories should be treated as supplementary descriptive information rather than as the central organizing variable of the analysis. Detailed results are reported in Appendix E and Appendix F.

E. SUMMARY OF MAIN FINDINGS

The results support four main conclusions. First, the NFT market is strongly non-stationary and can be organized into broad regimes with distinct return distributions and estimated alpha dispersion. Second, collection-level alpha heterogeneity is most visible in the *boom* regime and remains detectable, though weaker, in the *post-boom* regime. Third, structural clusters provide a useful descriptive organization of this heterogeneity, especially during the *boom* period, but the explanatory power remains modest and the exact partition is algorithm-dependent. Fourth, feature-level evidence suggests that persistence and concentration help organize cross-sectional differences in estimated alpha, although these associations should not be interpreted as causal effects or out-of-sample predictive relationships.

VI. DISCUSSION

Interpretation of the main result.

The main result is that cross-sectional alpha heterogeneity in NFT markets is regime-dependent and can be usefully organized by structural clusters and structural features, most clearly in the *boom* regime and more weakly in the *post-boom* regime. The broad market state and the collection-level market structure should therefore be considered jointly when interpreting cross-sectional differences in NFT outcomes. However, this organization should be interpreted descriptively. The clusters summarize recurring combinations of structural features and residual performance patterns; they should not be read as real-time predictors, causal mechanisms, trading signals, or unique latent classes of NFT collections. A stricter real-time design would require regime-specific or rolling clustering based only on information available up to each regime or date. We partially assess this concern through regime-specific clustering checks, but leave a fully real-time clustering framework to future work.

Descriptive role of structural features.

Feature-level analyses indicate that persistence is, on average, positively associated with estimated alpha, whereas concentration measures are negatively associated with estimated alpha. Descriptively, collections with broader and more durable participation appear to occupy a different market position from collections whose activity depends on a narrower and more fragile trading base. These associations should be interpreted as conditional descriptive relationships rather than causal drivers or out-of-sample predictors of collection performance.

Interpretation of the pre-boom regime.

The *pre-boom* regime differs from later market states, but the evidence there is substantially more fragile because the sample is smaller, the market is thinner, and cluster-based comparisons are less stable. Although some weighted fit measures are larger in the *pre-boom* regime, this should not be interpreted as stronger substantive evidence because the effective sample is much smaller and the early market structure is less mature. We therefore treat the *pre-boom* results as exploratory and mainly useful for showing that the descriptive relevance of structural organization is not uniform across market conditions.

Interpretation of observed categories.

Coarse platform-aligned categories are useful for describing broad market composition, but they do not replace internal market-structure measures for organizing collection-level alpha heterogeneity. Observed categories remain incomplete and coarse, and they do not map one-to-one onto the structural clusters. They are therefore best interpreted as supplementary descriptive attributes rather than as a ground-truth taxonomy of NFT collections.

Interpretation of the post-boom regime.

The *post-boom* regime spans a relatively long interval and likely contains substantial internal heterogeneity. The adopted three-regime partition should therefore be read as a transparent baseline rather than as a claim of internal homogeneity. More granular segmentation of the post-boom period may reveal additional sub-regimes or changes in the relationship between structural features and estimated alpha.

Implications.

These findings suggest that coarse collection labels alone are unlikely to sufficiently summarize cross-sectional heterogeneity in NFT markets. Structural signals such as participation breadth, persistence, and concentration provide a more informative descriptive lens within the same market environment. At the same time, the explanatory power of these structural variables remains modest. The results therefore support the use of structural features as an organizing framework, not as a high-powered predictive model of collection-level performance.

Limitations.

This study has several limitations. (i) *Scope*: the analysis focuses on Ethereum-based collections, so generalization to other chains or market environments may be limited. (ii) *Observability*: on-chain and metadata-based variables do not fully capture off-chain drivers such as attention, community quality, reputation, marketing activity, or broader social dynamics. (iii) *Non-causality*: all reported relationships are descriptive associations rather than causal effects. In particular, reverse causality cannot be ruled out: collections with high realized alpha may subsequently attract broader participation, longer trading persistence, or lower concentration. (iv) *Alpha interpretation and factor specification*: alpha is estimated under an intentionally parsimonious two-factor benchmark and should be interpreted as residual performance relative to that benchmark rather than as a fully purified measure of abnormal returns or collection-specific skill. Additional systematic components may matter, including liquidity conditions, speculative attention, volatility, momentum, platform-level effects, and broader crypto-market fragility. Omitted components may therefore remain in the residual benchmark even under the supplementary robustness checks reported in the appendix. In practice, this limitation also reflects data comparability constraints, as richer liquidity-, attention-, or broader crypto-market-related proxies are difficult to measure in a standardized and reproducible manner across the full weekly collection-level panel analyzed here. (v) *Estimation uncertainty*: alpha is estimated with heterogeneous precision across collections. The weighted robustness checks mitigate this concern, but they do not fully propagate first-stage estimation uncertainty through all downstream clustering, group comparisons, and regime-level analyses. (vi) *Illiquidity and non-trading weeks*: returns are computed only for consecutive trading weeks, with gaps treated as missing. This construction may underrepresent highly illiquid or intermittently traded collections in the alpha-estimation stage. Consequently, estimated alpha distributions may be more reflective of relatively continuously traded collections, while non-trading, stale pricing, and disappearance from active trading are only partially captured. In NFT markets, such non-trading periods are economically meaningful rather than purely technical missing values. (vii) *Clustering design*: several structural features are measured using full-sample information, so the baseline clustering should be interpreted as an ex post descriptive taxonomy rather than as a real-time classification system. Regime-specific clustering checks partially address this concern, but a fully real-time or rolling clustering design is left for future work. In addition, the Gaussian mixture model comparison reported in Appendix D shows limited-to-moderate agreement with the baseline k-means partition, indicating that exact cluster assignments are algorithm-dependent and should not be interpreted as invariant latent classes. (viii) *Category measurement*: the observed category variable used in this study is based on broad platform-aligned labels, but it remains an incomplete and coarse descriptor rather than a ground-truth taxonomy of NFT collections. (ix) *Effect size*: although many between-group differences are statistically meaningful, the corresponding explanatory power remains modest. In the weighted cluster-only comparisons, the cluster-based R^2 is 0.0134 in the *boom* regime and 0.0032 in the *post-boom* regime, while the higher *pre-boom* value of 0.0614 is based on a much smaller and less stable sample. The reported variables should therefore be viewed as organizing part of the estimated alpha cross-section rather than providing a high-powered predictive model. (x) *Regime definition*: the adopted three-regime partition is a transparent baseline, but alternative segmentations are possible, especially within the long *post-boom* period. (xi) *Suspicious trading*: although we report supplementary sensitivity checks based on simple suspicious-trade rules, these should be viewed as lightweight sensitivity screens rather than a full wash-trade or manipulation-detection system. NFT transaction data may still contain strategically generated volume, self-dealing, coordinated trading, or manipulation-related measurement error that is difficult to remove without richer wallet-level behavioral or transaction-graph modeling.

Future research.

Potential extensions include replication across additional blockchains, the development of richer common-factor specifications that incorporate liquidity, attention, trading activity, volatility, momentum, platform-level conditions, and broader crypto-market-related proxies, integration of off-chain signals, more granular segmentation of the *post-boom* period, and fully real-time or rolling clustering designs that avoid full-history feature construction.

VII. CONCLUSION

This paper examined regime-dependent cross-sectional alpha heterogeneity in NFT markets. Using a descriptive empirical framework, we showed that the NFT market is strongly nonstationary and that collection-level estimated alpha heterogeneity is most visible in the *boom* regime, while remaining detectable but weaker in the *post-boom* regime. The *pre-boom* evidence is comparatively fragile and should be interpreted more cautiously.

The main contribution of the paper is to show that structural clusters and structural features provide a useful descriptive framework for organizing regime-dependent heterogeneity in estimated residual performance. In particular, participation dispersion, trading persistence, and concentration help summarize cross-sectional differences across collections. Observed categories retain some supplementary information, but their incremental contribution beyond structural clusters is modest in the core parts of the sample.

Overall, the results suggest that collection-level residual performance in NFT markets is associated with both broad market conditions and internal market structure, in ways that vary across major market regimes. The paper should therefore be read as a regime-aware descriptive framework rather than as a causal account of alpha generation, a real-time prediction system, a definitive segmentation of NFT collections, or a complete asset-pricing model for NFT returns.

DATA AVAILABILITY

A subset of the processed research dataset underlying this study is publicly archived on Zenodo (DOI: `10.5281/zenodo.19062204`). The public archive includes the analysis-ready tables required to reproduce the reported results, along with a variable dictionary that describes the released tables and variables.

CODE AVAILABILITY

The code used for data extraction, preprocessing, feature construction, regime detection, clustering, regression analysis, and figure/table generation is publicly available at https://github.com/skondo/nft-trading-dataset. The repository includes environment specifications, implementation details, and step-by-step instructions for reproducing the empirical pipeline.

REPRODUCIBILITY NOTE

The underlying blockchain records originate from public Ethereum data accessed programmatically via the Infura API and related processing pipelines. Raw third-party metadata and other externally sourced data subject to redistribution restrictions are not redistributed; instead, the public repository documents the corresponding acquisition, preprocessing, and integration procedures so that authorized users can re-obtain those inputs directly from the respective providers under their own credentials and applicable terms.

APPENDIX A DATA CONSTRUCTION AND PREPROCESSING DETAILS

This appendix documents the construction of the NFT transaction dataset and the preprocessing steps used in the empirical analyses. The main text focuses on the sample scope, the unit of analysis, and key summary statistics. Here, we provide the implementation details required for reproducibility, along with a limited set of diagnostics confirming that the sample's main stylized features are not artifacts of filtering.

A. DATA SOURCES AND UNIT OF ANALYSIS

We construct secondary-market trade records on the Ethereum mainnet using on-chain event logs accessed programmatically via the Infura API [25]. We reconstruct trades from protocol-level events rather than relying on pre-aggregated indices. The primary analysis unit is the *collection-week*. Trades are aggregated weekly using block timestamps (UTC) and ISO weeks (weeks that start on Monday).

For reproducibility, the empirical analysis reported in this paper is based on a versioned processed dataset constructed from these underlying blockchain records. Rather than requiring readers to reconstruct the entire extraction workflow from raw chain data, we provide the processed research dataset together with documentation of the extraction, filtering, and aggregation pipeline. The public archival location of the processed dataset and the associated analysis code are described in the Data Availability and Code Availability statements in the main text.

B. TARGET PROTOCOLS AND CONTRACT COVERAGE

We target major secondary-market protocols that cover dominant flow over the sample period, including CryptoPunks, OpenSea Wyvern (v1/v2), OpenSea Seaport (v1.1–v1.6), and Blur (v1/v2).

C. ETH/USD SERIES AND TIME ALIGNMENT

Trades are observed in ETH. We convert ETH amounts to USD using daily ETH/USD data from Etherscan [24]. For each trade, we apply the rate for the trade date (UTC) and aggregate USD values to the collection-week level using the same ISO-week calendar. Weekly ETH/USD returns are constructed using the same week calendar as the FX factor.

D. FILTERING INVALID AND ANOMALOUS TRADES

We apply a minimal, deterministic preprocessing policy designed as a basic filter rather than a comprehensive wash-trade detection system. The steps below outline the basic trade-level filtering used to construct the prefiltered dataset before weekly aggregation.

1) Baseline filters (trade definition)

- Restrict payment tokens to ETH and WETH.
- Remove trades with missing, zero, or negative prices.
- Remove self-trades (buyer address equals seller address).
- Remove trades routed through OpenSea's *Shared Storefront* contract (to avoid ambiguous collection attribution).

2) Outlier filtering (extreme trade prices)

For each collection, we compute the 99.9th percentile of USD trade prices and remove trades above this local threshold. Additionally, we apply a global upper bound and remove trades with USD price above \$5,000,000. Table 4 summarizes the impact.

TABLE 4. Impact of basic invalid-trade and outlier filtering on the trade distribution used to construct the prefiltered dataset.

Metric	Origin	Filtered	Changed
Trades	36,184,632	36,150,388	−0.09%
Collections	59,740	59,730	−0.02%
Total Volume (USD)	9.73×10^{13}	5.32×10^{10}	−99.95%
Min Price (USD)	0.00	0.00	0.00%
Max Price (USD)	2.71×10^{12}	5.00×10^{6}	−99.99%
Median Price (USD)	143.02	142.71	−0.22%
Unique Traders	3,202,720	3,200,545	−0.07%

E. SUPPLEMENTARY SENSITIVITY CHECKS FOR SUSPICIOUS TRADING

Our baseline preprocessing is intentionally conservative and should be viewed as a basic screening procedure rather than a comprehensive wash-trade or manipulation-detection system. NFT transaction data may still contain strategically generated volume, self-dealing, coordinated trading patterns, or other forms of manipulation that are difficult to identify without wallet-level behavioral modeling, transaction-graph analysis, or visual analytics-oriented detection frameworks [26]. We do not attempt to implement such a full detection framework here. Instead, the supplementary screen below is used only to assess whether the paper's main regime-dependent findings are mechanically driven by a small subset of potentially anomalous trades.

To this end, we apply two supplementary trade-level flagging rules. First, we define a *7-day round-trip proxy* that flags a trade when the same token is subsequently traded back between the same wallet pair within a 7-day window. Second, we define a *rapid-resale-plus-price-jump proxy* that flags trades involving unusually rapid resale followed by a large price jump, using a quantile-based threshold calibrated to the sample. These rules are intended only as lightweight sensitivity checks and should not be interpreted as definitive identification of wash trading or market manipulation.

Table 5 summarizes the number of trades flagged by each rule and by their union. In total, the union of the two rules flags 98,326 trades, corresponding to 0.28% of the prefiltered trade-level dataset.

We then re-ran the key group-based comparisons after excluding these flagged trades. The qualitative conclusions remained unchanged. In particular, the main regime-dependent pattern in the structural-cluster analyses was preserved, and the supplementary category-based results were not materially altered. These results indicate that the paper's central conclusions are not driven solely by the flagged suspicious-trade subset. They should not, however, be interpreted as evidence that the dataset is free from manipulation-related measurement error or that all wash-trading activity has been removed.

F. TRADING PERSISTENCE MEASURES

A *trading week* is a week with at least one observed trade for a collection. We define two persistence measures: (i)

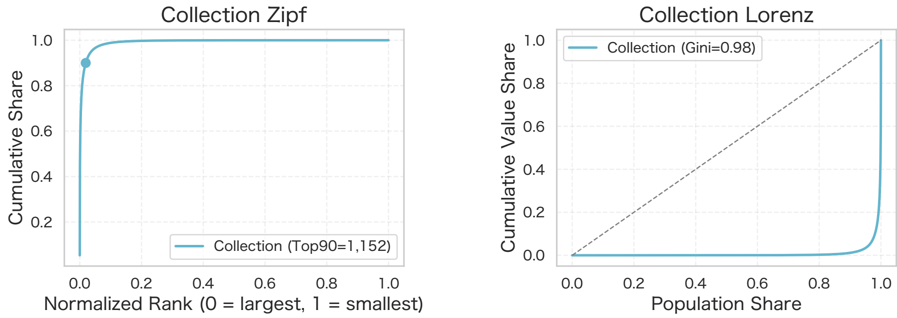

FIGURE 8. **Collection-level concentration diagnostics (Zipf and Lorenz). Trading volume is extremely concentrated across collections (Gini = 0.98; Top90 = 1,152).**

TABLE 5. **Counts of trades flagged by the supplementary suspicious-trade rules.**

Rule	Flagged trades	Share (%)
Rule A: 7-day round-trip proxy	40,774	0.12
Rule B: rapid resale + price-jump proxy	58,849	0.17
Union (Rule A or Rule B)	98,326	0.28

`active_weeks`, the number of distinct trading weeks over the sample period, and (ii) `max_consecutive_weeks`, the maximum run length of consecutive trading weeks. Table 6 reports summary statistics.

TABLE 6. **Summary statistics of collection-level trading-week persistence measures.**

Week duration	Collections	%	volume (USD)	%
1–3 weeks	31,663	53.00	436,150,620.12	0.80
4–7 weeks	9,012	15.10	409,378,564.57	0.80
8–15 weeks	6,859	11.50	1,553,451,611.08	2.90
16–31 weeks	5,027	8.40	2,084,098,453.51	3.90
32–51 weeks	2,773	4.60	2,984,514,138.83	5.60
52+ weeks	4,396	7.40	45,695,817,880.40	86.00

G. MINIMAL CONCENTRATION DIAGNOSTICS

To summarize concentration, we report a collection-level Zipf/Lorenz diagnostic. Figure 8 shows that trading volume is extremely concentrated across collections (Gini = 0.98; Top90 = 1,152). Participant-level concentration is similarly high and is omitted here only for space reasons.

APPENDIX B SUPPLEMENTARY METHODOLOGY FOR MARKET REGIME IDENTIFICATION

This appendix documents the change-point pipeline used in Section IV-A. The purpose of this appendix is reproducibility. The adopted three-regime partition is treated as a transparent baseline empirical segmentation rather than as a uniquely correct partition of NFT market history.

A. INPUTS AND PREPROCESSING

We apply change-point detection to multiple weekly market indicators aligned to the ISO week calendar (UTC). These indicators summarize market scale, participation intensity, and price-level dynamics. Before detection, each indicator is transformed using $\log(1+x)$. Missing values are forward-filled, and leading values are set to zero to form a continuous weekly series.

B. LOCAL DETECTION (PELT)

Detection is implemented in Python using `ruptures` with PELT [27]. We use the RBF cost, which is sensitive to changes in mean and variance or, more broadly, distributional shape, and set `pen=6` in the baseline implementation. This value is used as a practical baseline that avoids excessive short-lived breaks while preserving an interpretable three-regime partition. Sensitivity checks for alternative penalty choices are reported below.

C. SENSITIVITY TO THE PENALTY CHOICE

Table 7 reports a sensitivity check for the PELT penalty parameter (`pen`) under the same preprocessing and the RBF cost. The *post-boom* boundary date is stable across the tested values (`pen` $\in$ $\{4, 6, 8, 10\}$), consistently selecting 2022–05–23. By contrast, the boundary marking the start of the *boom* regime is sensitive only at the lowest penalty. With the penalty parameter set to 4, the algorithm yields an additional early-2021 candidate. It selects 2021–03–29 as the first 2021 breakpoint, whereas for penalty values of 6 or higher the boundary is stable at 2021–01–18. Lower penalties also increase the number of global change points, indicating greater segmentation of short-lived fluctuations.

We therefore use `pen=6` as the baseline setting in the main analysis. The important point for the paper is not that this value is uniquely optimal, but that the adopted boundary dates remain unchanged across a moderate range of penalty choices once over-segmentation is avoided.

D. GLOBAL CANDIDATES AND BOUNDARY MAPPING

We pool local detections across indicators and temporally merge close dates with `tol_days=84` (12 weeks), using

TABLE 7. Penalty sensitivity check for regime boundary dates (PELT, RBF).

pen	Boom start	Post-boom start	# global CPs
4	2021–03–29	2022–05–23	11
6	2021–01–18	2022–05–23	9
8	2021–01–18	2022–05–23	9
10	2021–01–18	2022–05–23	9

the earliest date in each merged group as the candidate representative. A candidate receives one vote per indicator if the indicator has any local change point within ± 84 days of the candidate (`radius_days=84`). Candidates are then selected greedily by vote count (descending) and date (ascending), subject to a minimum separation of `tol_days`. We keep only candidates dated on or after `2018-01-01`.

Final regime boundaries are mapped deterministically from the selected candidate set:

- **Boom start:** earliest selected candidate in calendar year 2021.
- **Post-boom start:** earliest selected candidate in calendar year 2022.

This mapping rule is rule-based and reproducible, but it should not be interpreted as implying that the resulting boundaries are uniquely determined by the data. Rather, it defines the baseline segmentation used consistently throughout the empirical analysis.

Table 8 provides an audit trail of detections, merged candidates, and vote counts.

APPENDIX C SUPPLEMENTARY METHODOLOGY FOR THE FACTOR MODEL AND ALPHA ESTIMATION

This appendix records the implementation choices for the baseline factor model used in Section IV-B. We retain only the settings required to reproduce the collection-level estimates of α and β. For reproducibility, we restate here the estimation equation introduced in the main text; Eq. (3) is identical in form to Eq. (1).

A. ESTIMATION SPECIFICATION

For each collection i, we estimate

$$r_{i,t} = \alpha_i + \beta_i^{(m)} r_{m,t} + \beta_i^{(fx)} r_{fx,t} + \varepsilon_{i,t}. \tag{3}$$

At the collection-week level, trades in the prefiltered dataset are first aggregated on the common ISO-week calendar. For each collection i and week t, we construct the weekly price measure $P_{i,t}$ as the median USD transaction price observed for that collection-week. Collection-level weekly returns are then computed as

$$r_{i,t} = \log\left(\frac{P_{i,t}}{P_{i,t-1}}\right), \tag{4}$$

but only when the previous observation for the same collection occurs exactly one week earlier. When trading gaps occur, the corresponding non-trading weeks are treated as missing, and no return is imputed across the gap.

The NFT market factor $r_{m,t}$ is constructed from the same weekly collection-return panel. Let $V_{i,t-1}$ denote the total USD trading volume of collection i in week $t-1$. For week t, the market factor is defined as the lagged-volume-weighted average of collection-level returns:

$$r_{m,t} = \frac{\sum_i r_{i,t} V_{i,t-1}}{\sum_i V_{i,t-1}}, \tag{5}$$

where the summation is taken over collections with a valid weekly return at week t and positive lagged trading volume. Thus, the market factor is aligned with the same weekly calendar as the collection-level return series and serves as a broad benchmark for NFT market movements.

The FX factor $r_{fx,t}$ is the weekly log return of the ETH/USD exchange rate constructed on the same weekly calendar.

Return winsorization.

To reduce the influence of extreme outcomes, we cap the return series at the 1st and 99th percentiles after removing missing values.

Inference and standard errors.

We compute Newey–West HAC standard errors [28] with `maxlags=8` at the weekly frequency.

Exclusion rules.

We exclude (i) samples with fewer than 8 valid weekly observations and (ii) degenerate cases with $R^2 > 0.999$.

Interpretation of alpha.

Throughout the paper, α_i is interpreted as a collection-level residual return component conditional on the baseline factor specification in Eq. (3). It is used as a descriptive residual benchmark rather than as a model-free measure of abnormal performance. The purpose of the specification is to provide a transparent and consistently estimable baseline decomposition for downstream cross-sectional comparison.

APPENDIX D SUPPLEMENTARY METHODOLOGY FOR CLUSTERING AND STRUCTURAL FEATURES

In this appendix, the term "baseline clustering sample" refers to the filtered analysis sample used for clustering, namely collections with at least eight active trading weeks, consistent with the main-text clustering definition in Section IV-C. This appendix documents the clustering pipeline with a minimal set of reproducibility elements: the feature set, preprocessing, cluster-size diagnostics, cluster interpretation, algorithmic sensitivity, and K-selection evidence.

A. STRUCTURAL FEATURES USED FOR CLUSTERING

The clustering analysis uses only structural features and excludes downstream outcomes such as returns, alpha, and factor exposures. The feature set is based on a fixed set of included variables designed to capture four broad dimensions of internal market structure: (i) holder structure and concentration, (ii) trading activity and persistence, (iii)

TABLE 8. Summary of detected change points across major market indicators.

Changepoint Date	Wallets	Collections	Market Volume	Median Market Price	ETH/USD	Global
2018-06-25	○	○	○	✓	○	⋆
2018-07-30	✓	✓	✓			
2018-08-27					✓	
2018-11-12	○		✓		○	⋆
2019-01-21	✓					
2019-10-28		✓	○			⋆
2020-01-06			✓			
2020-07-27			○	○	✓	⋆
2020-08-03				✓		
2020-10-12			✓			
2021-01-18	○			○	✓	⋆
2021-01-25	✓			✓		
2021-07-19	✓	✓	✓			⋆
2022-05-23			○	○	✓	⋆
2022-05-30			✓	✓		
2023-04-10	✓		○			⋆
2023-05-15			✓			
2023-11-06				✓	○	⋆
2023-12-04					✓	

price-distribution properties, and (iv) lightweight project-level metadata attributes. The full variable list is summarized in Table 3 in the main text.

Because several structural features are measured using full-sample information, the resulting clusters should be interpreted as an ex post descriptive taxonomy rather than as an implementable real-time classification rule. The clustering is therefore used to organize recurring combinations of participation breadth, persistence, concentration, and price structure, not to predict within-regime alpha outcomes in real time or to identify unique latent classes of NFT collections. It should also not be interpreted as an out-of-sample validation exercise for the predictive content of structural features.

B. FEATURE PREPROCESSING

We impute missing numeric feature values using the column median, and set them to zero if values remain missing after conversion. $\pm\infty$ values are treated as missing and imputed identically. To reduce the dominance of extreme values in Euclidean distances, we winsorize each feature at the 1st and 99th quantiles after transformations. For strongly right-skewed variables, we apply $\log(1+x)$ before standardization. Finally, all features are standardized to z-scores before clustering.

This preprocessing is intended to ensure that clustering reflects relative structural differences across collections rather than being driven mechanically by a small number of extreme-scale variables. Nevertheless, the resulting partition may still depend on feature scaling, transformations, and algorithmic choices. The baseline k-means solution is therefore interpreted as a transparent and reproducible descriptive partition, not as a uniquely determined segmentation of the NFT collection universe.

C. CHOOSING THE NUMBER OF CLUSTERS AND VALIDATION DIAGNOSTICS

We evaluate $K = 2, \ldots, 12$ using within-cluster sum of squares (WCSS) and the silhouette coefficient, and select $K = 8$ as a balanced specification that preserves interpretability while avoiding over-fragmentation. Figure 9 reports these diagnostics for the baseline clustering sample, that is, the filtered analysis sample of collections with at least eight active trading weeks. This figure therefore justifies the cluster choice used in the baseline analysis reported in the main text and Appendix D, rather than the stricter 52-week restricted sample considered later in Appendix F.

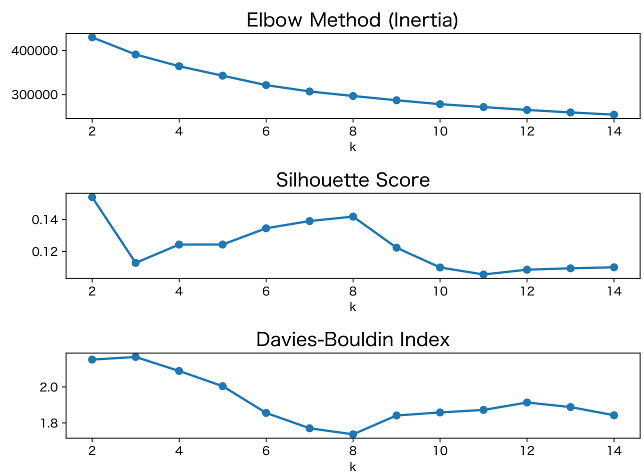

FIGURE 9. Diagnostics for selecting the number of clusters in the baseline clustering sample (collections with at least 8 active trading weeks).

Because clustering results need not correspond to a uniquely correct partition, we interpret cluster validation in a cautious and multi-criterion way. The WCSS and silhouette diagnostics are used to select an interpretable baseline value of K, but they do not imply that the resulting k-means partition is the only valid representation of structural heterogeneity.

D. ALGORITHMIC SENSITIVITY OF THE CLUSTERING PARTITION

We additionally evaluate whether the baseline k-means taxonomy is sensitive to the clustering algorithm. Specifically, we estimate a Gaussian mixture model (GMM) with full covariance matrices, k-means initialization, five random initializations, and the same number of components as the baseline k-means specification.

The agreement between the two partitions is limited to moderate. The adjusted Rand index is 0.237, and the normalized mutual information is 0.365. These values indicate that the detailed cluster assignments are not invariant to the clustering algorithm. The GMM comparison therefore should not be interpreted as confirming a uniquely robust latent classification. Rather, it supports a more cautious interpretation: the baseline k-means clusters provide a transparent and interpretable ex post taxonomy, while the exact partition remains algorithm-dependent.

TABLE 9. Algorithmic sensitivity of the baseline clustering partition.

Alternative method	Setting	Agreement with baseline
Gaussian mixture model	Full covariance; same number of components as baseline k-means; k-means initialization; $n_{\text{init}} = 5$	$ARI = 0.237$; $NMI = 0.365$

Taken together, the K-selection diagnostics and the GMM comparison indicate that the baseline partition is useful for descriptive organization but should not be read as a definitive segmentation. Alternative clustering approaches, including feature-weighted clustering, unsupervised feature selection, spectral clustering, or rolling and dynamic clustering procedures, may emphasize different aspects of the feature space and could yield different partitions.

E. CLUSTER SIZE AND REGIME COMPOSITION

Table 10 reports the distribution of structural clusters across the three market regimes together with each cluster's total size and overall sample share. The composition is clearly uneven across both clusters and regimes. C1, C3, and C4 are the largest clusters, together accounting for most collections in the filtered analysis sample, whereas C2, C5, and C6 remain comparatively small. The regime breakdown also shows that several clusters are disproportionately concentrated in particular market phases. In particular, C4 is strongly *post-boom* heavy, while C1 and C3 contain especially large components from the *boom* period. This pattern is consistent with the detailed interpretation below, where C3 is characterized as a broad market-core cluster, C4 as an active low-price cluster with strong *post-boom* representation, and C1 as a broad intermediate cluster. By contrast, the *pre-boom* regime remains small across all clusters, reinforcing the view that *pre-boom* evidence should be interpreted more cautiously. Overall, the regime composition suggests that the structural partition captures heterogeneous market profiles whose prevalence shifts over time rather than remaining constant across regimes.

TABLE 10. Cross-tabulation of clusters and market regimes in the baseline clustering sample (collections with 8+ active trading weeks).

Cluster	Pre-boom	Boom	Post-boom	Total	Share (%)
C0	121	543	1666	2330	12.2
C1	180	2539	1652	4371	22.9
C2	53	289	64	406	2.1
C3	129	2760	1382	4271	22.4
C4	32	776	3286	4094	21.5
C5	1	119	871	991	5.2
C6	45	235	401	681	3.6
C7	77	788	1046	1911	10.0

F. DETAILED CLUSTER INTERPRETATION

Tables 11 and 12 summarize the baseline $k = 8$ clustering solution and the structural characteristics underlying each cluster. The partition is uneven in size, with three large clusters (C1, C3, and C4) accounting for most collections, one very small but distinctive cluster (C2), and several smaller specialized groups. The selected median features clarify that the clusters differ not only in size but also in economically meaningful market structure. In particular, C3 combines the longest trading history, the broadest participation, the highest activity, and the lowest concentration, supporting its interpretation as a "market core" cluster. C4 is also active but is located in the low-price segment, while C1 forms a broad intermediate cluster with moderate persistence and participation. By contrast, C2 is a small, high-price, high-volatility niche cluster, C6 is distinguished by extreme seller concentration and a strongly imbalanced trading structure, and C0 represents a thin-liquidity, highly concentrated profile. Taken together, these tables indicate that the structural partition captures distinct market-structure types rather than merely splitting collections by scale alone. However, in light of the algorithmic sensitivity documented above, these types should be interpreted as an interpretable descriptive typology rather than as invariant latent classes.

G. OBSERVED CATEGORY COMPOSITION WITHIN STRUCTURAL CLUSTERS

Table 13 reports a compact summary of observed category composition within each structural cluster. The table shows that all clusters contain multiple observed categories and that category concentration remains moderate rather than near-exclusive. Although several clusters are dominated by broad OpenSea-derived labels such as PFPs or Art, the dominant-category shares are generally well below complete concentration, and some clusters are instead dominated by the `none` category. Category entropy also remains substantial across clusters. Taken together, these patterns indicate that the structural clusters do not map one-to-one onto the observed cat-

TABLE 11. Interpretation of market-structure clusters in the baseline clustering sample (8+ active trading weeks).

Cluster	N	Structural profile	Interpretive summary
C0	2,330	Very low persistence, very small participant base, low trading activity, high buyer concentration and high top-holder concentration, mid-price segment	Thin-liquidity, highly concentrated cluster with limited market depth
C1	4,371	Moderate persistence, moderate participant base, low-to-moderate concentration, moderate price level, relatively balanced trading profile	Broad intermediate cluster with moderate liquidity and participation
C2	406	Longer persistence, moderate breadth, low concentration, extremely high price level, and very high price dispersion	Small premium-price cluster with extreme valuation and volatility
C3	4,271	Highest persistence, largest participant base, thick trading activity, lowest concentration, long trading span	"Market core" cluster with broad participation and long-lived trading history
C4	4,094	Short-to-moderate persistence, active trading, mid-sized holder base, relatively low concentration, lowest-price segment	Active low-price trading cluster with moderate breadth
C5	991	Shorter persistence, modest activity, moderate concentration, lower-mid price segment, unusually low royalty rate / ERC2981-heavy profile	Smaller operational cluster with moderate liquidity and distinctive metadata/royalty structure
C6	681	Very low persistence, very small participant base, low activity, extreme seller concentration, and highly imbalanced buyer-seller structure	Seller-dominated, highly concentrated niche cluster
C7	1,911	Moderate persistence, moderate activity, moderate holder base, relatively dispersed participation, mid-to-high price dispersion	Persistent middle cluster with moderate liquidity and relatively dispersed participation

TABLE 12. Selected median structural features by cluster in the baseline clustering sample (8+ active trading weeks).

Feature	C0	C1	C2	C3	C4	C5	C6	C7
active_weeks	11.0	20.0	33.0	81.0	17.0	18.0	11.0	21.0
buyer_hhi	0.134	0.035	0.022	0.004	0.014	0.034	0.054	0.030
seller_hhi	0.086	0.029	0.026	0.003	0.008	0.029	0.963	0.036
top10_share	0.862	0.479	0.373	0.172	0.308	0.467	0.556	0.434
trades	39.0	98.0	117.0	2789.0	725.5	157.0	37.0	117.0
trading_days	318.0	527.0	1024.5	1047.0	425.5	471.0	278.0	591.0
unique_holders	15.0	48.0	60.0	766.0	209.0	60.0	26.0	56.0
p75_price	60.970	340.108	8583.712	421.685	29.353	80.029	102.247	161.221

TABLE 13. Compact summary of observed category composition within each structural cluster.

Clus.	No.	Dom. cat.	Dom. share (%)	Entropy
C0	10	none	45.19	1.414
C1	10	PFPs	36.77	1.526
C2	10	Art	38.18	1.685
C3	10	PFPs	52.38	1.439
C4	10	PFPs	51.49	1.286
C5	8	Art	33.91	1.533
C6	8	none	33.48	1.529
C7	10	Art	39.09	1.591

egory labels. Rather, the observed categories provide only a coarse external description of market composition, whereas the structural clustering captures a different dimension of within-market heterogeneity.

H. CLUSTER-WISE FEATURE DISTRIBUTIONS

To support interpretation of the market-structure typologies, Figure 10 visualizes how representative structural features differ across clusters. The systematic shifts in feature distributions are consistent with recurring market-structure types, but they should not be interpreted as evidence of a unique or algorithm-invariant classification.

Taken together, the cluster-size table, interpretation table, compact category-composition table, and feature-distribution plot suggest that the clustering provides a descriptive summary of recurring combinations of participation breadth, persistence, concentration, and price structure.

APPENDIX E SUPPLEMENTARY ANALYSES AND ROBUSTNESS CHECKS

This appendix reports the main robustness evidence referenced in the main text while avoiding large suites of figures or tables. We focus on six issues: (i) stability of feature selection, (ii) regime-wise nonparametric group comparisons, (iii) pooled and regime-wise fit comparisons, (iv) weighted robustness checks that account for heterogeneous alpha-estimation precision across collections, (v) supplementary evidence on observed categories, and (vi) robustness to alternative factor specifications and clustering designs.

A. LASSO SELECTION STABILITY

To assess the sensitivity of the Lasso-selected feature set to sample variation, we conducted a subsample-based stability

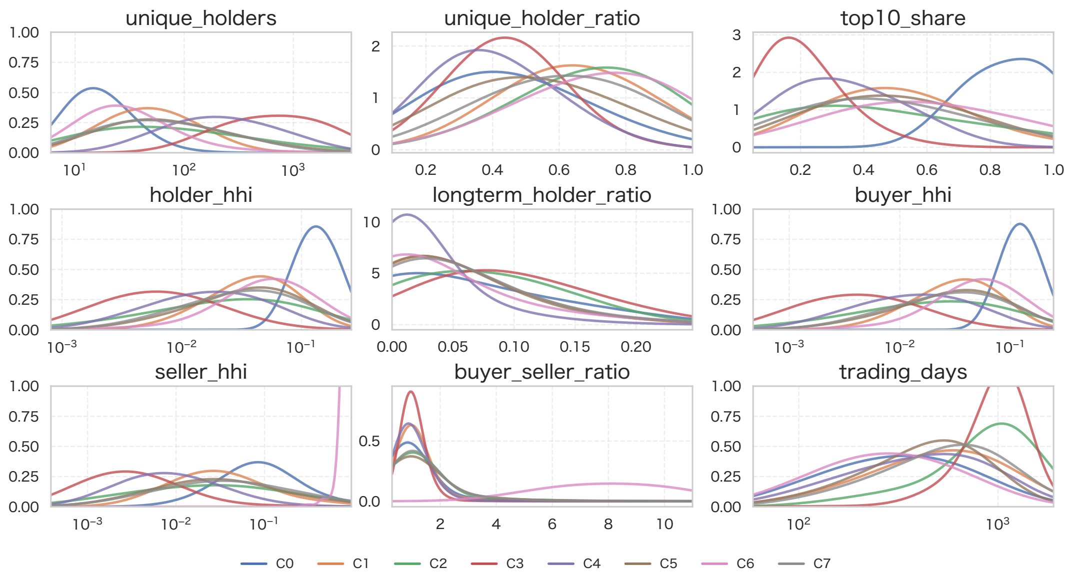

FIGURE 10. Cluster-wise distributions of representative structural features: holder structure/concentration and participation-balance related variables.

check. Specifically, we repeated the Lasso estimation 200 times using random subsamples containing 80% of the observations (drawn without replacement), while keeping the same preprocessing and cross-validation procedure as in the baseline specification. Table 14 reports how often each feature was selected with a non-zero coefficient.

The results indicate that several core features are consistently selected across repetitions. The most stable variables are price-dispersion, concentration, and trading-persistence measures, including `log1p_diff_price`, `log1p_top10_share`, `log1p_active_weeks`, and `log1p_seller_hhi`. A second tier of variables appears with moderate frequency and includes price-level, royalty, token-standard, holder-breadth, and long-term-holder measures. Lower-frequency variables should be interpreted with greater caution, as their inclusion depends more strongly on the specific subsample and on partially overlapping predictor blocks.

Overall, the stability check suggests that the Lasso results are moderately to strongly stable for a core subset of persistence, concentration, and price-distribution variables, but not uniquely determined for the full feature set. This is consistent with the appearance of some additional token-standard, holder-breadth, and metadata-related variables in the selected set, whose inclusion appears to depend more on sample variation and correlation structure.

TABLE 14. Subsample-based Lasso selection frequencies across 200 repetitions.

Feature	Count	Freq. (%)
log1p_diff_price	200	100.0
log1p_top10_share	199	99.5
log1p_active_weeks	199	99.5
log1p_seller_hhi	198	99.0
log1p_p25_price	166	83.0
log1p_royalty_fee_percent	160	80.0
log1p_is_erc721	134	67.0
log1p_unique_holders	131	65.5
log1p_is_erc1155	123	61.5
log1p_longterm_holder_ratio	105	52.5
log1p_buyer_hhi	92	46.0
log1p_has_discord	86	43.0
log1p_stddev_price	55	27.5
log1p_description_length	54	27.0
log1p_unique_holder_ratio	52	26.0

B. REGIME-WISE NONPARAMETRIC GROUP COMPARISONS

We use the Kruskal–Wallis test to evaluate whether alpha distributions differ across groups within each regime. Table 15 reports the test statistics and p-values for both observed categories and structural clusters.

The results indicate that cross-sectional separation is regime-dependent under both grouping schemes. In the *boom* regime, both observed categories and structural clusters show clear between-group differences, with strong statistical significance under each partition. In the *post-boom* regime, the same pattern remains visible, although the structural-cluster signal is stronger than the category-based signal. In the *pre-boom* regime, by contrast, the evidence is weaker and does not reach conventional significance levels under either grouping scheme. Overall, these nonparametric comparisons support the view that group-based alpha heterogeneity is most clearly detectable in the *boom* regime, remains present in the *post-boom* regime, and is comparatively fragile in the *pre-boom* regime.

TABLE 15. Kruskal–Wallis test results for category- and cluster-based differences in alpha within each regime.

	Category			Cluster		
Regime	H	p	Grp.	H	p	Clus.
pre-boom	12.348	5.46e-02	6	5.869	3.19e-01	6
boom	73.528	1.79e-13	10	68.958	2.42e-12	8
post-boom	32.754	1.48e-04	10	56.713	6.34e-10	8

TABLE 16. Kruskal–Wallis test results for category- and cluster-based differences in alpha within each regime after suspicious-trade filtering.

	Category			Cluster		
Regime	H	p	Grp.	H	p	Clus.
pre-boom	17.306	4.41e-02	6	6.272	5.08e-01	6
boom	11.728	2.29e-01	10	86.833	5.52e-16	8
post-boom	23.692	4.82e-03	10	63.786	2.64e-11	8

C. SENSITIVITY TO SUPPLEMENTARY SUSPICIOUS-TRADE FILTERING

To assess whether the main cross-sectional results are sensitive to suspicious trading, we re-ran the key regime-wise nonparametric comparisons after excluding trades flagged by the supplementary suspicious-trade rules defined in the suspicious-trading sensitivity subsection of Appendix A.

Table 16 shows that the qualitative regime-dependent pattern remains unchanged after excluding the flagged trades. In particular, cluster-based differences remain weak in the *pre-boom* regime and detectable in the *boom* and *post-boom* regimes. Category-based differences remain more visible in the *pre-boom* and *post-boom* regimes than in the *boom* regime, and under suspicious-trade filtering, the category signal in the *boom* regime becomes notably weaker. Thus, the main pattern is not driven solely by trades flagged by the supplementary suspicious-trade rules.

D. POOLED INCREMENTAL CONTRIBUTION OF STRUCTURAL CLUSTERS AND OBSERVED CATEGORIES

We compare how much observed categories and structural clusters explain cross-sectional alpha heterogeneity using pooled weighted regressions and goodness-of-fit measures (R^2, adjusted R^2), where the weights are given by the number of weeks used to estimate alpha. This weighting scheme provides a simple sensitivity check for the fact that alpha-estimation precision differs across collections. The alpha-analysis sample is smaller than the full structural sample because it additionally requires valid factor-model estimation under the baseline eight-week rule.

Table 17 shows that both observed categories and structural clusters contribute explanatory power in pooled weighted specifications, although the absolute fit remains modest. The improvement over the regime-only baseline is positive for both additions, and the gains from adding categories and clusters are similar in magnitude. Table 18 further shows that both grouping schemes retain positive conditional incremental contribution, with the pooled incremental contributions again being close in magnitude. Taken together, these pooled results suggest that observed categories and structural clusters capture partly distinct aspects of cross-sectional alpha heterogeneity rather than indicating a clear pooled dominance of either grouping scheme.

TABLE 17. Pooled regression fit comparison (R^2): WLS (weighted by estimation weeks).

Model	$\mathbf{R^2}$	Adj. $\mathbf{R^2}$	N	$\Delta\mathbf{R^2}$
Baseline (regime only)	0.0035	0.0034	15,079	
+ Category	0.0065	0.0058	15,079	0.0030
+ Cluster	0.0066	0.0060	15,079	0.0032
+ Category + Cluster	0.0090	0.0079	15,079	

TABLE 18. Conditional incremental contribution ($\acute{}R^2$): WLS (weighted by estimation weeks).

Metric	Value
Conditional ΔR^2: Cluster (given Category)	0.0026
Conditional ΔR^2: Category (given Cluster)	0.0024

E. REGIME-WISE FIT COMPARISONS

This section compares the explanatory power of observed categories and structural clusters for cross-sectional alpha heterogeneity within each regime using weighted regressions. We report R^2 values from regressions that include either category dummies only or cluster dummies only, with the same regime-specific sample, together with the conditional incremental contribution of clusters given categories, $\Delta R^2(\mathrm{Cluster} \mid \mathrm{Category})$.

Table 19 shows that the relative contribution of categories and clusters is regime-dependent, but structural clusters retain a positive conditional contribution in all three regimes. In the *boom* regime, cluster-based specifications fit better than category-based specifications, and the conditional contribution of clusters remains positive. In the *pre-boom* regime, the cluster-only fit is also higher than the category-only fit under this weighted specification, and the conditional contribution of clusters is largest in magnitude, although this result should be interpreted cautiously because the *pre-boom* sample remains much smaller and empirically less stable than the later regimes. In the *post-boom* regime, the standalone fit difference between categories and clusters is very small, but the conditional contribution of clusters remains positive. Overall, these weighted regime-wise comparisons are consistent with the view that structural clusters provide supplementary explanatory power beyond broad observed categories. Because the *pre-boom* sample is much smaller and less stable, we place greater substantive weight on the evidence from the *boom* regime and interpret the *post-boom* contribution as more modest.

TABLE 19. Regime-wise R^2 comparison (category vs. cluster): WLS (weighted by estimation weeks).

Regime	N	Cat. R^2	Clu. R^2	Cond. ΔR^2
pre–boom	206	0.0312	0.0614	0.0645
boom	4,628	0.0100	0.0134	0.0109
post–boom	10,245	0.0034	0.0032	0.0025

F. WEIGHTED ROBUSTNESS SUMMARY

Table 20 reports weighted cluster-dummy fit measures by regime as a sensitivity check for heterogeneous alpha-estimation precision. Under this specification, the weighted cluster-based fit is highest in the *pre-boom* regime, remains clearly positive in the *boom* regime, and is smaller but still positive in the *post-boom* regime. However, because the *pre-boom* sample is much smaller and empirically less stable, this stronger fit should be interpreted cautiously. Accordingly, the weighted results do not overturn the broader interpretation adopted in the main text, which places the main substantive emphasis on the *boom* regime and, more modestly, on the *post-boom* regime.

TABLE 20. Fit summary for weighted cluster-dummy regressions.

Regime	N	R^2	Adj. R^2
pre–boom	206	0.0614	0.0282
boom	4,628	0.0134	0.0119
post–boom	10,245	0.0032	0.0025

G. COVERAGE AND DISTRIBUTION OF OBSERVED CATEGORIES

Because observed categories are used only as supplementary descriptive information, it is useful to document their empirical coverage in the retained analysis sample. Table 21 compares the coverage of observed categories between the broader prefiltered collection universe and the filtered analysis sample used in the structural analyses.

The comparison shows that observed-category coverage improves substantially after filtering. In the prefiltered collection universe, 46.35% of collections have an observed category, whereas 53.65% remain unlabeled. By contrast, in the filtered analysis sample, 77.45% of collections have an observed category and 22.55% remain unlabeled. Thus, although the category variable remains incomplete, the retained sample used in the supplementary category-based analyses has substantially better coverage than the broader collection universe.

Table 22 further reports the distribution of observed categories in the filtered analysis sample. The labeled subset is concentrated in a small number of broad platform-aligned categories, especially PFPs and Art, while a nontrivial `none` group remains. This reinforces the interpretation adopted in the main text: observed categories provide a coarse descriptive summary of market composition, but do not constitute a complete taxonomy of NFT collections.

TABLE 21. Coverage of observed categories in the filtered analysis sample and the prefiltered collection universe.

Sample	Total	Labeled	Unlabeled	Labeled (%)	Unlabeled (%)
Filtered	19,055	14,759	4,296	77.45	22.55
Prefiltered	59,730	27,682	32,048	46.35	53.65

TABLE 22. Distribution of observed categories in the filtered analysis sample.

Category	Collections	Share (%)
PFPs	7,233	37.96
Art	4,605	24.17
Gaming	1,338	7.02
Memberships	909	4.77
Photography	277	1.45
Virtual worlds	190	1.00
Music	125	0.66
Sports collectibles	59	0.31
Domain names	23	0.12
none	4,296	22.55

TABLE 23. Collection-level factor estimates by specification and regime.

Spec.	Regime	N	Avg. obs.
usd_2f	pre-boom	206	31.09
usd_2f	boom	4,628	19.62
usd_2f	post-boom	10,245	38.14
usd_1f	pre-boom	206	31.09
usd_1f	boom	4,628	19.62
usd_1f	post-boom	10,245	38.14
eth_1f	pre-boom	205	31.20
eth_1f	boom	4,620	19.64
eth_1f	post-boom	10,228	38.18
usd_2f_orth	pre-boom	206	31.09
usd_2f_orth	boom	4,628	19.62
usd_2f_orth	post-boom	10,245	38.14

H. ALTERNATIVE FACTOR SPECIFICATIONS AND FACTOR-CORRELATION DIAGNOSTICS

We compare four specifications: the baseline USD two-factor model (`usd_2f`), a USD market-only model (`usd_1f`), an ETH-denominated market-only model (`eth_1f`), and a USD model using orthogonalized FX returns (`usd_2f_orth`).

Table 23 shows that usable estimation counts are very similar across specifications. The correlation between the baseline NFT market factor and the FX factor is low (0.0467 in the baseline USD specification and 0.0130 in the ETH-denominated formulation), suggesting that the two-factor benchmark is not driven by simple double-counting of exchange-rate movements. Table 24 shows that the qualitative conclusions are broadly preserved across specifications: cluster-based organization is most clearly supported in the *boom* regime, remains detectable but modest in the *post-boom* regime, and is more fragile and sample-sensitive in the *pre-boom* regime.

TABLE 24. Supplementary regression comparison across alternative factor specifications under WLS (weighted by estimation weeks): cluster-only vs. cluster-plus-category models.

Spec.	Regime	R^2 Clu.	Adj. R^2 Clu.	R^2 Clu. + Cat.	Adj. R^2 Clu. + Cat.	Δ R^2 Cat.
usd_2f	pre-boom	0.0283	-0.0025	0.0348	-0.0174	0.0065
usd_2f	boom	0.0093	0.0078	0.0129	0.0097	0.0036
usd_2f	post-boom	0.0009	0.0002	0.0024	0.0008	0.0015
usd_1f	pre-boom	0.0639	0.0342	0.0732	0.0231	0.0093
usd_1f	boom	0.0110	0.0095	0.0143	0.0111	0.0033
usd_1f	post-boom	0.0010	0.0003	0.0029	0.0013	0.0019
eth_1f	pre-boom	0.0628	0.0329	0.0721	0.0217	0.0094
eth_1f	boom	0.0090	0.0075	0.0115	0.0082	0.0024
eth_1f	post-boom	0.0010	0.0003	0.0028	0.0012	0.0018
usd_2f_orth	pre-boom	0.0327	0.0020	0.0398	-0.0121	0.0071
usd_2f_orth	boom	0.0095	0.0080	0.0132	0.0100	0.0037
usd_2f_orth	post-boom	0.0009	0.0003	0.0024	0.0008	0.0015

TABLE 25. Comparison of full-history and regime-specific clustering: cluster-only fit under WLS (weighted by estimation weeks).

Design	Regime	N	R^2 Clus. only
Full-history clustering	boom	4,624	0.0088
Regime-specific clustering	boom	4,624	0.0223
Full-history clustering	post-boom	10,245	0.0012
Regime-specific clustering	post-boom	10,245	0.0012

TABLE 26. Weighted fit comparison on the labeled-only subsample. Here, K denotes structural clusters and C denotes observed categories.

Regime	N	R^2_K	Adj. R^2_K	R^2_{K+C}	Adj. R^2_{K+C}	ΔR^2_C
pre-boom	170	0.0856	0.0460	0.1325	0.0541	0.0469
boom	4,308	0.0105	0.0089	0.0189	0.0155	0.0084
post-boom	8,437	0.0030	0.0022	0.0053	0.0036	0.0023

I. REGIME-SPECIFIC CLUSTERING AS A CHECK AGAINST LOOK-AHEAD BIAS

To examine whether the main findings depend materially on full-history feature construction, we compare the baseline clustering design with regime-specific clustering exercises in which clusters are re-estimated separately within each regime. This comparison is intended as a practical check against possible forward-looking contamination. We do not report a separate *pre-boom* regime-specific clustering result because the effective *pre-boom* sample is too small to support stable cluster re-estimation and comparison.

Table 25 reports the cluster-only fit comparison for the *boom* and *post-boom* regimes under WLS. In the *boom* regime, regime-specific clustering yields a clearly larger cluster-based fit than the full-history design, whereas in the *post-boom* regime the corresponding fit is essentially unchanged. This suggests that the main finding in the *boom* regime is not driven by forward-looking contamination from full-history cluster assignments.

Accordingly, full-history clustering should be interpreted as an ex post descriptive organization device rather than as a strictly real-time classification rule, and a fully real-time clustering framework is left for future work.

J. LABELED-ONLY CATEGORY ROBUSTNESS

Because observed categories remain incomplete, we re-estimate the supplementary category-versus-cluster comparison on the labeled-only subsample. Table 26 reports the regime-wise weighted fit results.

The qualitative interpretation is broadly unchanged. In the *boom* and *post-boom* regimes, structural clusters retain positive explanatory power, while observed categories add some incremental information within the labeled-only subsample. The incremental category contribution is larger in the *pre-boom* regime, but this result should be interpreted cautiously because the *pre-boom* labeled sample is small. Thus, restricting attention to labeled collections does not overturn the main interpretation that observed categories and structural clusters capture partly distinct aspects of cross-sectional alpha heterogeneity, with the central regime-dependent pattern remaining concentrated in the *boom* regime and, more modestly, in the *post-boom* regime.

APPENDIX F ROBUSTNESS CHECK USING A 52-WEEK SAMPLE RESTRICTION

This appendix reports robustness checks based on a stricter sample definition that retains only collections observed for at least 52 weeks. Because the NFT market contains many short-lived collections, this restriction serves as a sensitivity check for longer-lived projects rather than as a replacement for the broader baseline clustering sample used in the main text. The restriction helps assess whether the main findings persist among collections with longer observed trading histories, but it also changes the sample composition and should not be interpreted as eliminating illiquidity, survivorship, or selection concerns.

For the 52-week restricted sample, we re-ran the clustering procedure independently using the same preprocessing and k-means workflow described in Appendix D. Because the restricted sample differs materially from the baseline in size and composition, we re-evaluated the number of clusters rather than mechanically retaining the baseline $k = 8$ partition. Figure 11 shows that the restricted long-history sample supports $k = 6$. Accordingly, all cluster-based analyses in this appendix use the re-estimated $k = 6$ solution.

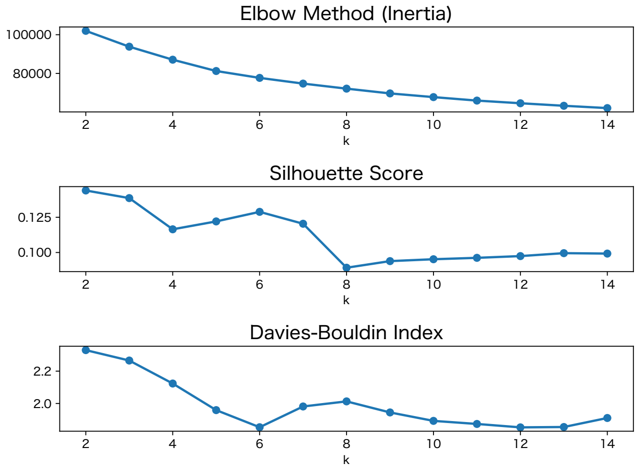

FIGURE 11. Cluster-selection diagnostics for the 52-week restricted sample. Using the same procedure as in the baseline analysis, the restricted long-history sample supports a six-cluster solution.

A. COMPARISON WITH THE BASELINE CLUSTERING SAMPLE

Compared with the baseline results reported in Appendix E, the 52-week restriction reduces the sample size substantially, especially in the *pre-boom* regime, and changes explanatory power in a regime-dependent way. The main qualitative conclusions remain similar: the *pre-boom* regime remains unstable, the cluster signal in the *boom* regime becomes substantially stronger, and the *post-boom* evidence remains weaker than in the *boom* regime. The restricted-sample results are summarized in Tables 27 and 28.

TABLE 27. Regime-wise R^2 comparison under the 52-week sample restriction: WLS (weighted by estimation weeks).

Regime	N	Cat. R^2	Clu. R^2	Cond. ΔR^2
pre–boom	126	0.0509	0.0353	0.0362
boom	2,245	0.0144	0.0478	0.0397
post–boom	4,281	0.0033	0.0028	0.0021

TABLE 28. Kruskal–Wallis test results for category- and cluster-based differences in alpha within each regime under the 52-week sample restriction.

	Category			Cluster		
Regime	H	p	Grp.	H	p	Clus.
pre-boom	9.029	2.51e-01	8	5.051	4.10e-01	6
boom	43.868	1.49e-06	10	135.565	1.57e-27	6
post-boom	17.426	4.24e-02	10	22.129	4.95e-04	6

B. REGIME-WISE FIT COMPARISON UNDER THE 52-WEEK RESTRICTION

Table 27 reports regime-wise R^2 comparisons under the 52-week restriction. In the *boom* regime, structural clusters explain substantially more cross-sectional variation in α than observed categories and retain a positive conditional contribution. In the *pre-boom* regime, category-only fit remains higher, although interpretation is fragile because of the small sample. In the *post-boom* regime, the difference in standalone fit is small, but the conditional contribution of clusters remains positive.

C. NONPARAMETRIC COMPARISON UNDER THE 52-WEEK RESTRICTION

Table 28 reports Kruskal–Wallis test results under the 52-week restriction. In the restricted sample, the *pre-boom* regime no longer exhibits stable between-group differences under either grouping scheme. By contrast, both the *boom* and *post-boom* regimes continue to show statistically significant between-group differences, with structural clusters yielding stronger separation than observed categories. Thus, the restricted-sample evidence remains consistent with the main descriptive interpretation that structural clusters provide the sharper organization of cross-sectional alpha heterogeneity in the *boom* regime and, more modestly, in the *post-boom* regime.

D. DISTRIBUTIONAL COMPARISON

Figure 12 shows regime- and cluster-wise KDEs under the 52-week restriction. Relative to the baseline clustering sample, the distributions are more concentrated, indicating lower overall dispersion. Even so, cross-cluster heterogeneity remains visible, particularly in the *boom* regime and, more modestly, in the *post-boom* regime.

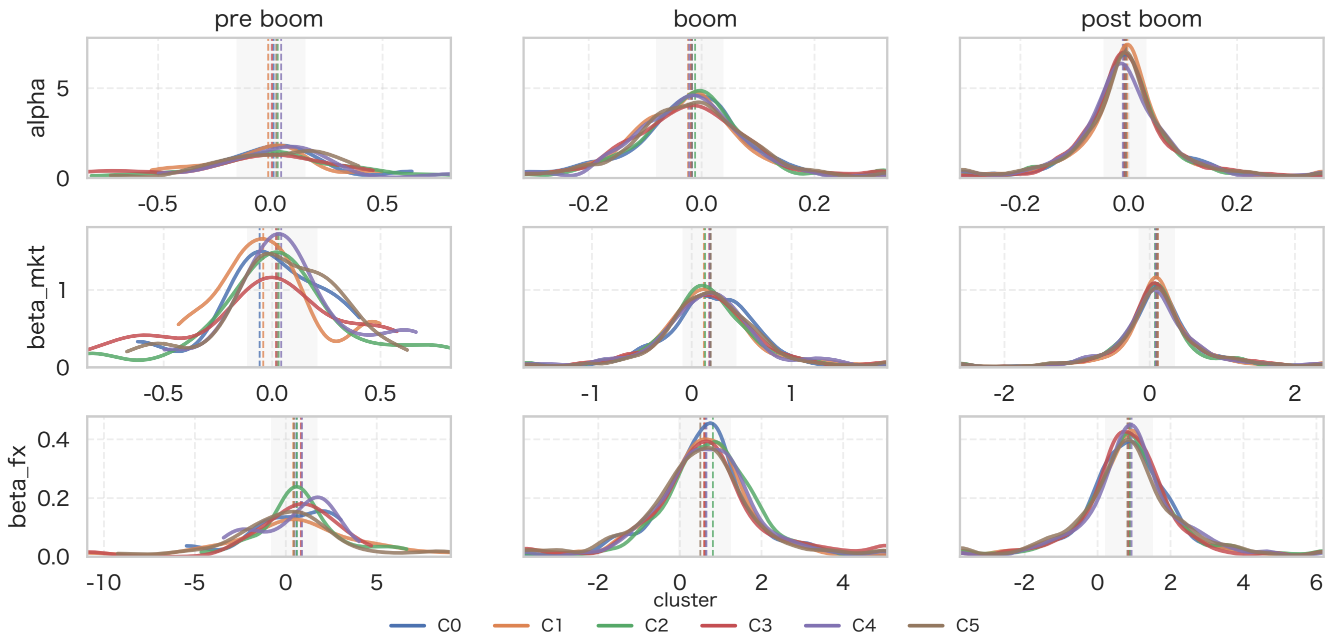

FIGURE 12. Regime- and cluster-wise KDEs of collection-level alpha α and factor exposures ($\beta^{(m)}$, $\beta^{(fx)}$) under the 52-week sample restriction. Compared with the baseline clustering sample, the distributions are more concentrated, although cross-cluster separation remains visible in the *boom* and *post-boom* regimes. Clustering is re-estimated for the restricted sample and uses the $k = 6$ solution from Figure 11.

ACKNOWLEDGMENT

The authors used an artificial intelligence system (ChatGPT, OpenAI) to assist with English-language refinement and limited editorial support during manuscript preparation. All AI-assisted outputs were reviewed and revised by the authors, who take full responsibility for the final content.

REFERENCES

[1] V. Buterin, "Ethereum: A next-generation smart contract and decentralized application platform," https://ethereum.org/en/whitepaper/, 2014, [Online]. Accessed: 2025-01-15.
[2] Ethereum Foundation, "Ethereum documentation," https://ethereum.org/en/developers/docs/, 2024, [Online]. Accessed: 2025-01-15.
[3] Ethereum Improvement Proposals, "ERC-721 non-fungible token standard," https://eips.ethereum.org/EIPS/eip-721, 2018, [Online]. Accessed: 2025-01-15.
[4] ——, "ERC-1155 multi token standard," https://eips.ethereum.org/EIPS/eip-1155, 2018, [Online]. Accessed: 2025-01-15.
[5] OpenSea, Inc., "OpenSea," https://opensea.io/, 2017, [Online]. Accessed: 2025-01-15.
[6] Larva Labs, "CryptoPunks," https://www.larvalabs.com/cryptopunks, 2017, [Online]. Accessed: 2025-01-15.
[7] Yuga Labs, "Bored Ape Yacht Club," https://boredapeyachtclub.com/, 2021, [Online]. Accessed: 2025-01-15.
[8] M. Nadini, L. Alessandretti, F. Di Giacinto, M. Martino, L. M. Aiello, and A. Baronchelli, "Mapping the NFT revolution: market trends, trade networks, and visual features," *Scientific Reports*, vol. 11, no. 1, p. 20902, 2021.
[9] L. Ante, "The Non-Fungible Token (NFT) Market and Its Relationship with Bitcoin and Ethereum," *FinTech*, vol. 1, no. 3, pp. 216–224, 2022.
[10] M. Dowling, "Is non-fungible token pricing driven by cryptocurrencies?" *Finance Research Letters*, vol. 44, p. 102097, 2022.
[11] F. Horky, C. Rachel, and J. Fidrmuc, "Price determinants of non-fungible tokens in the digital art market," *Finance Research Letters*, vol. 48, p. 103007, 2022.
[12] Y. Wang, "Volatility spillovers across NFTs, news attention, and financial markets," *International Review of Financial Analysis*, vol. 83, p. 102313, 2022.
[13] D. Shannon, M. Dowling, M. Zhaf, and B. Sheehan, "Dutch auction dynamics in non-fungible token (NFT) markets," *Economic Modelling*, vol. 141, p. 106882, 2024.
[14] R. Kräussl and A. Tugnetti, "Non-fungible tokens (nfts): A review of pricing determinants, applications and opportunities," *Journal of Economic Surveys*, vol. 38, no. 2, pp. 555–574, 2024.
[15] M. Dowling, "Fertile land: Pricing non-fungible tokens," *Finance Research Letters*, vol. 44, p. 102096, 2022.
[16] S. Wilkoff and S. Yildiz, "The behavior and determinants of illiquidity in the non-fungible tokens (NFTs) market," *Global Finance Journal*, vol. 55, p. 100782, 2023, metadata corrected (the DOI year is 2022 in the journal record).
[17] I. Yousaf and L. Yarovaya, "The relationship between trading volume, volatility and returns of non-fungible tokens: A quantile approach," *Finance Research Letters*, vol. 50, p. 103175, 2022.
[18] G. Fridgen, R. Kräussl, O. Papageorgiou, and A. Tugnetti, "Pricing dynamics and herding behaviour of NFTs," *European Financial Management*, vol. 31, no. 2, pp. 670–710, 2025.
[19] T. Panagiotidis and G. Papapanagiotou, "A note on the determinants of non-fungible tokens returns," *International Journal of Finance & Economics*, vol. 30, no. 3, pp. 3201–3211, 2025.
[20] A. Alsultan, A. Kourtis, and R. N. Markellos, "Can we price beauty? aesthetics and digital art markets," *Economics Letters*, p. 111572, 2024.
[21] F. Ali and N. Ülkü, "Quest for a parsimonious factor model in the wake of quality-minus-junk, misvaluation and fama-french-six factors," *Finance Research Letters*, vol. 41, p. 101847, 2021.
[22] J. D. Hamilton, "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, vol. 57, no. 2, pp. 357–384, 1989.
[23] O. Dedousi, A. P. Fassas, and D. Philippas, "Investor behavior in the NFTs market: A bibliometric and systematic literature review," *Finance Research Letters*, vol. 80, p. 107398, 2025.
[24] Etherscan, "Etherscan: Ethereum blockchain explorer," https://etherscan.io/, 2025, [Online]. Accessed: 2025-01-15.
[25] Infura, "Infura: Ethereum api and ipfs infrastructure," https://www.infura.io/, 2025, [Online]. Accessed: 2025-01-15.
[26] X. Wen, Y. Wang, X. Yue, F. Zhu, and M. Zhu, "NFTDisk: Visual Detection of Wash Trading in NFT Markets," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23. Association for Computing Machinery, 2023, pp. 215:1–215:15.
[27] R. Killick, P. Fearnhead, and I. A. Eckley, "Optimal detection of changepoints with a linear computational cost," *Journal of the American Statistical Association*, vol. 107, no. 500, pp. 1590–1598, 2012.
[28] W. K. Newey and K. D. West, "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, vol. 55, no. 3, pp. 703–708, 1987.

SATOSHI KONDO received the master's degree from the Graduate School of Social Data Science, Hitotsubashi University, Tokyo, Japan, and the M.B.A. degree from the Graduate School of Management, GLOBIS University, Tokyo, Japan. He has industry experience in data platforms, analytics, and AI-driven business applications. His research interests include NFT markets, digital assets, market structure, and applied data science.

RYOHEI BANNO received the B.E. degree and the M.S. degree in Information Science and Technology from Hokkaido University in 2010 and 2012, respectively, and the Ph.D. degree in Science from Tokyo Institute of Technology in 2018. From 2012 to 2018, he was a researcher at NTT Network Innovation Laboratories. From 2018 to 2020, he was a researcher at Tokyo Institute of Technology. From 2020 to 2022, he was an Assistant Professor at Kogakuin University, and from 2022 to 2024, he was an Associate Professor. Since 2024, he has been an Associate Professor at Hitotsubashi University. His research interests include distributed systems and IoT systems. He is a member of ACM, IEEE, IEICE, and IPSJ.