INSEAD The Business School for the World®

Working Paper

2021/60/MKT

(Revised version of 2020/58/MKT)

Infinite but Rare: Valuation and Pricing in Marketplaces for Blockchain-Based Nonfungible Tokens

Pavel Kireyev
INSEAD, pavel.kireyev@insead.edu

Ruiqi Lin
Sciences Po, ruiqi.lin@sciencespo.fr

October 2021

Blockchain technologies have enabled the creation of decentralized applications which let users own and transact scarce digital assets called nonfungible tokens or NFTs. Although still in its infancy, the industry has generated over $2.5bn in transaction volume and attracted interest from organizations such as the NBA, several football (soccer) clubs, major brands, and gaming companies to create platforms for trading digital collectibles. A major question faced by NFT platforms is how to help participants value the digital items. We introduce a novel dataset and study how traditional approaches to valuation may exhibit significant biases in this market. We find that while buyers value NFTs much like we would expect them to value physical collectibles, sellers have a tendency to price sub-optimally, which causes traditional hedonic regression approaches to generate inaccurate valuations. We develop a valuation approach based on a structural model of the selling mechanism used in a popular NFT market to highlight these biases and develop a proof-of-concept decision support tool to help participants make more informed decisions.

Keywords: Digital Items; Collectibles; Nonfungible Tokens; Pricing; Valuation; Marketplaces; Blockchain; Structural Model.

We thank participants at the INSEAD Quantitative Marketing and Economics Lab and the 2021 European Marketing Academy Conference for helpful comments and Peipei Cai for helpful research assistance. We thank the Harborne Blockchain Fund for funding this research.

Introduction

The collectibles market, with an estimated market size of over 300 billion USD, consists predominantly of scarce physical assets, such as paintings, trading cards, and sports memorabilia.[1] However, recent innovations in blockchain technology have made it possible to create a sense of scarcity and exclusive ownership around digital assets, such as images and text files. As a result, major organizations are exploring systems to enable their customers to own and trade scarce digital assets, called nonfungible tokens or NFTs. For example, Formula One and the National Basketball Association have licensed their brands to create marketplaces for trading blockchain-based digital collectibles. Major football (soccer) clubs such as Juventus FC, of world-famous player Cristiano Ronaldo, have licensed their brands and player identities to create digital collectable cards. Collectors have spent billions of dollars trading these digital items.[2]

Despite the growing interest in NFTs, limited research exists on how to accurately determine their value. We present a novel dataset from a market for digital collectibles, describe buyer preferences for characteristics of these items, show evidence that sellers may not price items consistently with buyer valuations, and highlight the inaccuracies that may result when using traditional hedonic regression approaches to valuation. We develop a structural model that accounts for the descending auction selling mechanism used in many such markets and compare the optimal prices that it generates to the price predictions of a flexible hedonic regression model estimated with machine learning methods. We use the structural model to develop a proof-of-concept decision support tool that can help both buyers and sellers make more informed decisions. Such a tool can be integrated into NFT marketplaces or offered as a stand-alone application.

[1] https://www.forbes.com/sites/darrenheitner/2016/04/11/playing-ball-in-the-multi-billion-dollar-sports-collectible-market

[2] https://www.reuters.com/technology/nft-sales-volume-surges-25-bln-2021-first-half-2021-07-05/

The literature on pricing and valuation of traditional collectibles, such as art, stamps, and wine, often relies on linear hedonic price regressions (Ashenfelter and Graddy 2003; Burton and Jacobsen 1999; Renneboog and Spaenjers 2013) to uncover the "implicit prices" of individual attributes and infer valuations for items which consist of these attributes. Inferring a valuation amounts to predicting the sale price of an item given its set of attributes. While researchers typically use linear models that regress observed sales prices on item characteristics, recent work has incorporated machine learning to allow for high-dimensional inputs and flexible interactions between the independent variables (Aubry et al. 2019). While it may appear natural to apply these methods to NFT markets as well, they may exhibit significant inaccuracies if they do not take into account details of how these market function. In particular, the market we study uses a descending auction selling mechanism, not all listed auctions result in a sale, and sellers may price sub-optimally. All of these factors can influence the results of hedonic regression approaches to valuation, regardless of the sophistication or flexibility of the underlying model.

We develop a structural model of buyer behavior that accounts for the descending auction selling mechanism, the existence of failed auctions, and corrects for idiosyncrasies in seller pricing decisions. The model amounts to a dynamic sequential game and captures two main competing incentives – on the one hand, buyers wish to wait as they expect the price of the item to fall over time; on the other hand, a competing buyer may purchase the item before the auction expires, limiting the value of waiting. We use optimal prices from the structural model to highlight biases in the traditional hedonic approach to valuation. Our proposed approach of inferring valuations from the structural model amounts to predicting successful sale prices conditional on optimal pricing decisions made by the sellers, thereby eliminating the impact of any seller pricing idiosyncrasies.

First, we identify a *mispricing bias*. We find that hedonic regressions tend to undervalue the NFTs despite the fact that sellers tend to set prices that are too high. This is because sellers also set very wide starting and ending price intervals in the descending auctions which encourages buyers to wait and purchase NFTs at lower prices than they would have had the sellers used more optimal narrower price intervals. Hedonic regressions use data on these lower-than-optimal successful sale prices to infer lower valuations. On the other hand, the structural model recommends narrower optimal price intervals and tends to yield higher predicted sale prices, despite recommending that participants set lower starting and ending prices on average.

Second, we identify a *selection bias*. Hedonic regressions only use data on successful sales to infer valuations. As a result, our hedonic model predicts higher valuations for NFTs that did not sell compared to NFTs that did sell in the data. This is because NFTs with a lower selling rate tend to have higher posted prices. On the occasions that they do sell, they sell for a higher price than most NFTs. Hedonic regressions do not take into account the low selling rate of these items and infer a high valuation. The structural model incorporates information about selling rate and recommends lower optimal prices and hence valuations for items that tend to sell less.

Taken together, these biases can lead to inaccuracies in traditional valuation methods based on hedonic regressions, regardless of their flexibility. We implement the optimal price recommendations of our structural model in a decision support interface that can help increase information transparency in NFT marketplaces.

Background on Nonfungible Tokens

What are NFTs? To fix ideas, consider the digital art industry. Absent blockchain technology, consumers can purchase digital artwork online for personal use. Typically, the artwork is delivered in the form of an image file which the purchaser can easily duplicate and distribute. Firms rely on legal enforcement to prevent duplication, and consumers derive no value from the artwork beyond their personal use. In stark contrast, physical art is difficult to duplicate, and large secondhand markets exist for trading physical art. Some blockchain projects have attempted to bridge these two worlds by limiting the perceived supply of individual pieces through smart contracts. Even if each asset consists of a shareable file, it may have a unique identifier and limited supply as programmed in the smart contract that governs new asset creation.[3] A distributed database stores the information such that no individual entity can alter records of ownership or the code for creating new assets. Although the underlying image file can still be duplicated, it is not possible to duplicate the identifier associated with the image file, or to change the code that governs who owns a particular item or the total number of identifiers that can be associated with a particular image. We investigate a marketplace for such NFTs where participants effectively trade blockchain-based images.

Rare Item Valuation and Blockchains

Research has shown that scarcity and serial numbers can influence prices for consumer products (Parker and Lehmann 2011; Stock and Balachander 2005; Verhallen 1982) and physical collectibles (Koford and Tschoegl 1998), all of which are difficult, if not impossible, to reproduce at scale. In stark contrast, participants can easily download, duplicate, and share digital items such as text files, images, music, and videos, which has led to the widespread adoption of file-sharing (Liebowitz 2006) and the implication that the price of a digital item falls to zero absent expensive legal enforcement efforts. Moreover, research has shown that consumers perceive digital objects as transient, unstable, and incapable of conveying the same level of psychological ownership as equivalent physical objects (Atasoy and Morewedge 2018). As a result, most modern business models based on the consumption of digital objects either do not involve ownership (e.g. Spotify and Netflix) or rely heavily on legal enforcement to prevent redistribution (e.g. stock photos and movie downloads). Given the novelty of blockchain technology, limited research exists on how participants value the “digital scarcity” it enables. In general, research on business applications of blockchains remains largely conceptual or theory-driven (Catalini and Gans 2019), and empirical research mostly focuses on fungible cryptocurrencies like Bitcoin (Halaburda et al. 2019). We aim to introduce a novel application and present one of the first empirical analyses of nonfungible digital items.

[3] In the blockchain context, a smart contract is a type of irreversible code. A smart contract may govern new item creation and specify the supply limit for a number of digital items. In contrast to standard private code repositories, smart contracts cannot be altered by an individual entity without the agreement of members affected by the contract.

On a substantive level, we show that participants value NFTs much like we would expect them to value physical collectibles, with strong preferences for scarce digital attributes and low serial numbers. Practically, we identify the biases that may result from attempting to use standard valuation approaches in these markets, and contribute to literature that identifies the shortcomings of hedonic regressions and proposes structurally-driven adjustments or alternatives (Pakes 2003). Methodologically, we draw on approaches to estimating dynamic structural models (Rust 1987) with the possibility of sequential decisions by multiple agents (Berry 1992, Ellickson and Misra 2011) and use gradient-boosted tree models (Friedman 2001, Chen and Guestrin 2016) with Bayesian hyperparameter optimization (Snoek, Larochelle, and Adams 2012) as our baseline hedonic regression model.

Seller Behavior in Marketplaces

We focus on a popular NFT series called CryptoKitties, which generated about 30 million USD in transactions since late 2017 and allows users to own and trade algorithmically-generated digital images of cats (Serada, Sihvonen, and Harviainen 2020). It was one of the first trial applications of NFTs, and attracted media attention for several transactions of individual items in excess of 100,000 USD. Participants who acquire a digital item, either directly from the developer or from another participant, may sell it by posting a descending auction on the marketplace and specifying a starting price, ending price, and duration. While many other NFT marketplaces use fixed price listings, descending auctions are a popular mechanism for pre-sales of NFTs before marketplace launch and can be found in popular NFT projects like Axie Infinity, which has generated over $1bn in transaction volume since last year.[4] Ideally, sellers should price taking into account buyer preferences and the state of the market, but a large number of items may not have a price history, making it difficult for sellers to infer reasonable price ranges. In the market we study, 80% of the items listed have never sold before, forcing sellers to scan the market in an attempt to identify comparable items which have sold before and have a price history. However, this task may be costly given that items

[4] https://sports.yahoo.com/axie-infinity-surpasses-1-billion-161142016.html

differ in their characteristics. Researchers have identified pricing inefficiencies in complex markets like energy and utilities (Doraszelski, Lewis, and Pakes 2018), and retail (Huang, Ellickson, and Lovett 2018). In practice, marketplaces like eBay simply advise sellers to search the marketplace for comparable items to identify the best price for their item.[5] Some marketplaces like Airbnb have released price recommendation tools for their users (e.g. Airbnb's smart pricing),[6] anecdotally suggesting that pricing inefficiencies exist in traditional peer-to-peer marketplaces as well. However, there is limited evidence of the nature of these inefficiencies or the specific design of the tools used to address them. We show evidence of mispricing in NFT marketplaces and connect it directly to biases that may result when using traditional approaches to value NFTs.

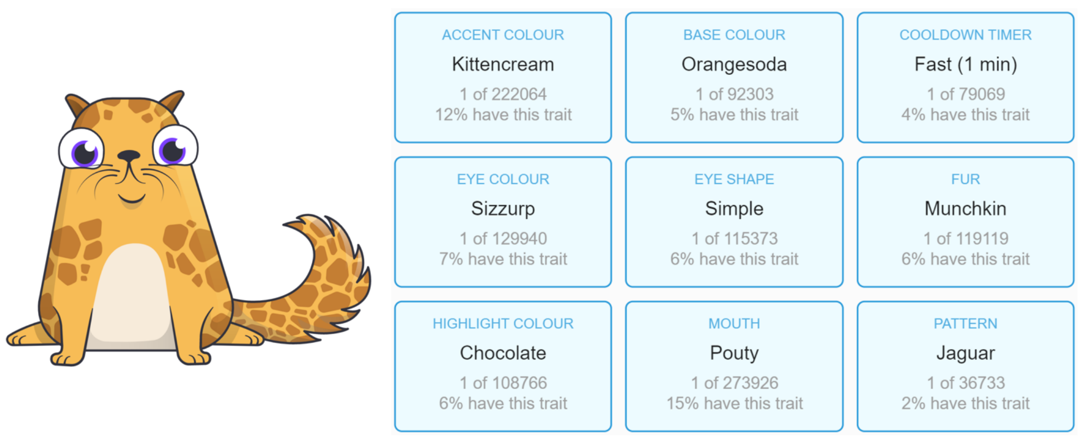
The environment we investigate shares many similarities with earlier research focused on internet auctions conducted through platforms such as eBay (Bajari and Hortaçsu 2004; Lewis 2011). Some of this work studied online auctions for trading physical collectibles such as coins or playing cards, although the research focused on auction design rather than approaches to value the collectibles (Lucking-Reiley 1999; Bajari and Hortaçsu 2003). Relatedly, virtual economies have received research attention because of their surprisingly large size and policy discussions around ownership and secondhand trading of virtual in-game items (Lehdonvirta and Virtanen 2010). Our research contributes to this discussion by highlighting the drivers of consumer valuations and how firms may resolve inefficiencies in these markets with the help of data-driven tools.

[5] https://www.ebay.com/help/selling/selling/pricing-items?id=4133
[6] https://blog.atairbnb.com/smart-pricing/

Data and Setting

We focus on a dataset of CryptoKitties descending auctions in 2019, when the market was fairly mature and stable relative to its early speculative stage in late 2017. We focus on this later period as we expect participants to have more stable preferences relative to earlier periods. CryptoKitties is a game centered around collectible digital cats, characterized by a set of discrete visual attributes, such as their fur, pattern, eyes, and color. Participants can acquire, breed, and trade these cats. By breeding cats, participants create additional cats with attributes that depend on the attributes of the parents and a random component. Beyond appearance and breeding mechanics, cats have limited utility to participants. As a result, one of the primary goals of participants is to acquire and trade cats, much like with traditional collectibles such as fine art or trading cards. Figure 1 shows an example CryptoKitty and some of its attributes. This particular item last sold for 3,254 USD worth of cryptocurrency in August 2019.

Figure 1: Example CryptoKitty with Attributes



Note: Figure presents the image and attributes for item ID 57 in the CryptoKitties ecosystem, taken from the OpenSea marketplace. Each box specifies an attribute category, the specific attribute that the item possesses within the category, and the scarcity of this attribute among all items in existence as of April 2020.

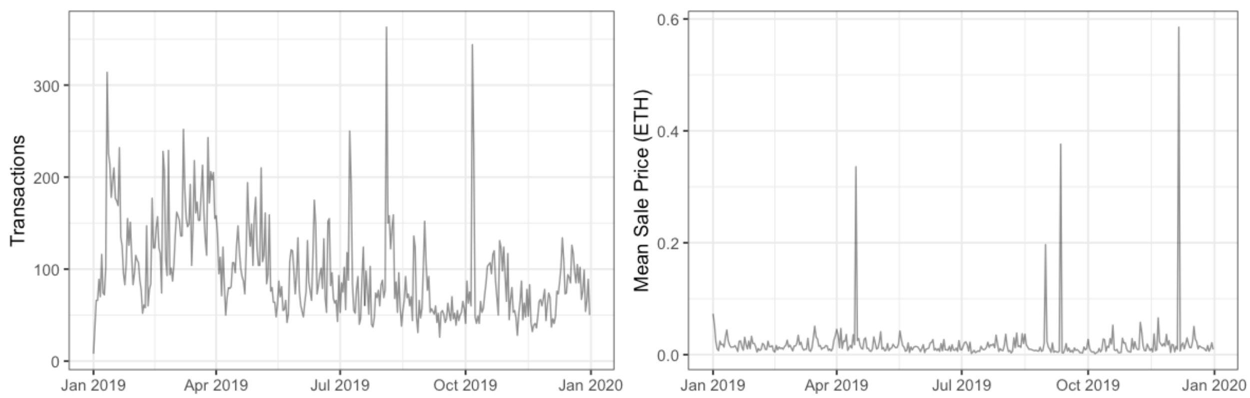
We obtain data from a platform called OpenSea, which offers a marketplace for nonfungible tokens and aggregates CryptoKitties transactions. The transactions occur as follows. First, participants obtain a CryptoKitty directly from the developer, from another participant, or create one with some random attributes by breeding two cats, usually in exchange for a cryptocurrency called ETH.[7] While the developer auctions off an initial set of items, participants tend to host the vast majority of subsequent auctions. Participants who wish to sell an item post a descending auction on the CryptoKitties website or on a trading platform like the OpenSea marketplace. The auction design consists of a starting price, an ending price, and a duration. After the participant posts the auction, the price descends linearly from the starting price to the ending price throughout the duration of the auction until a purchase occurs or the poster cancels the auction. Around 80% of the auctions listed and 95% of the items sold that we observe in our data are for CryptoKitties with no past sales history, suggesting that they were created by participants. This exacerbates the difficulty participants may face in valuing and pricing the items, as they cannot rely on transaction histories for the same item to infer reasonable prices.

Figure 2 shows daily marketplace trends - the number of transactions (left panel) and the average transaction price in ETH (right panel). Participants traded a total of 646 ETH, or $108,244 worth of digital items in our sample. The trends do not appear to change massively over time, confirming our intuition that the market entered a stable phase in this period. Regarding cryptocurrency exchange rates, the USD price of ETH increases in the summer before steadily returning to around its initial value towards the end of the year. ETH does not have as many applications as traditional currencies like USD, and participants may be mentally converting ETH prices to USD prices when they make decisions. Despite the varying exchange rate, transaction prices in ETH exhibit a 0.996 correlation with USD prices for individual items. We model

[7] ETH refers to Ether, a cryptocurrency based on the Ethereum blockchain. At any point in time, a participant can convert between USD and ETH at a time-varying exchange rate using a cryptocurrency exchange.

the prices in ETH in the analysis that follows and find that incorporating the USD equivalent does not affect any of the results.

Figure 2: Daily Marketplace Trends



We use data from OpenSea on all descending auctions posted in 2019. We study a total of 72,299 auctions, consisting of 35,766 successful sales and 36,533 auctions that did not result in a sale. We exclude about 10% of auctions for "fancy" cats from the data as they do not possess standard appearance attributes. We also focus only on auctions that lasted at least one day or longer. The top panel of Table 1 presents summary statistics for the auctions in the sample. Successful auctions have a median starting to ending price range of 0.007 ETH compared to 0.010 ETH for unsuccessful auctions, providing some initial evidence that less successful sellers set wider price ranges, perhaps because of uncertainty about how to value their items. As expected, median starting and ending prices are higher for unsuccessful auctions (0.020 ETH, 0.005 ETH) compared to successful auctions (0.010 ETH, 0.002 ETH). The median unsuccessful auction has a duration 2 days (during which the price decreases), compared to 5 days for the median successful auction. The ranges of these design parameters can vary significantly. Participants have set starting prices in excess of a billion USD and ending prices close to zero. Sale prices can also vary significantly across transactions. The smallest sale involved only 0.0009 ETH or 0.16 USD, whereas the most expensive item sold for 30.24 ETH or 4,848 USD. The average transaction price is 0.018 ETH, whereas the median transaction price is 0.004 ETH, implying that the price distribution is skewed because of occasional expensive transactions.

Table 1: Summary of Data

Auctions		Min	Mean	Median	Max
Starting Price (S)		0.001	0.050	0.010	199
Starting Price (U)		0.001	3e+8	0.020	1e+13
Ending Price (S)		2e-4	0.012	0.002	30
Ending Price (U)		1e-08	8.200	0.005	1e+05
Price Range (S)		1e-5	0.041	0.007	199
Price Range (U)		1e-5	3e+8	0.010	1e+13
Duration (S)		1	14.46	5	60
Duration (U)		1	11.54	2	60
Sold		0	0.495	0	1
Sale Price		0.001	0.018	0.004	30
Attribute		**Min**	**Mean**	**Median**	**Max**
ID		45	1,303,561	1,366,634	1,792,385
Generation		0	6.533	5	360
Bred		0	0.605	1	1
Category	**N. Types**	**Most Frequent**	**Sales**	**Least Frequent**	**Sales**
Mouth	32	Pouty	5,094	Walrus	35
Pattern	30	Totesbasic	5,311	Avatar	50
Fur	32	Selkirk	5,130	Kurilian	76
Eyes	32	Thicccbrowz	5,363	Hacker	89
Color 1	32	Greymatter	4,224	Icicle	49
Color 2	32	Swampgreen	4,550	Mertail	23
Color 3	32	Frosting	3,783	Summerbonnet	54
Eye Color	32	Cyan	3,655	Oasis	26
Wild	16	Flapflap	725	Foghornpawhorn	29
Environment	16	Salty	939	Prism & Floorislava	39
Cooldown	14	Snappy (10 min)	3,679	Sluggish (4 days)	925

Note: (S) refers to the subset of successful auctions and (U) refers to the subset of unsuccessful auctions. “N. Types” counts the number of unique types of a particular attribute.

The middle panel of Table 1 summarizes a small set of numerical and appearance-related attributes of the individual items across all auctions. Each item has an ID which describes the order of item creation. There existed a total of about 1.8 million different CryptoKitties by the end of 2019. In the sample, we mostly observe trades involving later-stage items with ID numbers over 1 million, although the lowest ID item traded was 45. Participants may prefer low ID numbers as they appear rarer than IDs in the millions, or in some way closer to the “origin” or the creator of the items (Smith, Newman, and Dhar 2016). Similarly, generation describes the relative age of an item. First generation cats were not created through breeding but were rather introduced to the market by the developer. The generation of the remaining cats is determined as one plus the maximum generation of its parent cats. As a result, generation provides an additional measure of "closeness to the origin" or scarcity as low generation numbers are more unique than high generation numbers. The attribute "bred" simply describes whether a cat has any children.

The bottom panel of Table 1 describes several appearance-related CryptoKitty attributes. There are 10 visual attributes which exclusively define the appearance of an item. Each attribute can take on between 14 and 32 different values. We list the values that occur most and least frequently in the transactions data for each attribute. The most frequent attribute value occurs 5,363 times and the least frequent value occurs 23 times. The attribute labelled "Cooldown" describes the time participants must wait after breeding a cat before breeding it again. This attribute is related to the cat's number of children because the developers implemented a mechanic that increases the cooldown for cats with more children in a deterministic fashion. The developers adopted this feature to limit the creation of new items and thereby the supply of items over time. As a result, we are unable to separate the functional benefit of the cooldown from the preferences participants may have for cats with multiple children, and simply consider cooldown as another control attribute.

Descriptive Analysis

Buyer Preferences

We first investigate buyer preferences, focusing on the subset of data corresponding to successful transactions. We begin with a linear hedonic regression specification that is common in the collectibles literature (Renneboog and Spaenjers 2013) and also used to study price formation in auctions (Lewis 2011). At this stage, we do not consider the potential biases that may be induced by this approach, but rather use it to demonstrate that the systematic differences in sale prices across different types of NFTs are primarily driven by NFT characteristics and not time or other control variables. Our regression model is of the form

$$\log P_{jt} = \alpha W_j + \sum_k \beta_k X_{kjt} + f(t) + \gamma Z_{jt} + \epsilon_{jt} \qquad \text{Equation 1}$$

where P_{jt} denotes the ETH sale price of item j at time t, W_j is a set of continuous attributes of the item with associated coefficients α, X_{kjt} is an indicator for the presence of discrete attribute k in item j at time t with β_k as the associated coefficient, $f(t)$ includes time-specific fixed effects, Z_{jt} is a set of control variables with associated coefficient γ, and ϵ_{jt} is an error term. The coefficients α and β_k capture the implicit prices of different attributes.

Our setting has several advantages over other data commonly used to study the valuation of collectibles. One advantage is that the appearance of each item is uniquely described by discrete attributes which we observe in the data, allowing us to fully control for the appearance of an item. We can further account for possible omitted variables by incorporating textual features of the items, such as their name and algorithmically-generated description, as control variables. In principle, because of the digital nature of the items, the information available to the researcher is similar to the information available to purchasers.

Note that while our regressions focus on serial numbers and scarce attributes, participants may value digital items for other reasons as well. For example, purchasers of digital items may act as early investors in an ecosystem. Similar to how crowdfunding allows fans of an idea to support its growth (Agrawal, Catalini, and Goldfarb 2014), early purchasers of digital items may support the development of the associated ecosystem while possibly benefiting from an increase in item prices later as additional users join a more advanced environment. Participants may also obtain direct utility from digital items if they can use them in derivative applications such as games. We focus on the CryptoKitties market in 2019, when the market is more mature and stable relative to its highly-speculative early days, to limit the potential influence of factors other than scarcity or serial numbers. In particular, users derive limited practical utility from owning CryptoKitties, which qualifies the item as a pure collectible.

The full regression model includes a total of 272 identified coefficients, excluding time-specific fixed effects. The large number of coefficients stems from the inclusion of indicators for a large set of possible discrete attributes. However, given the relatively large sample size, we are able to estimate all identified coefficients. For clarity, we present estimates only for the effects of token ID and generation, even though all coefficients are estimated. Unless otherwise specified, we obtain standard errors through a bootstrap procedure (detailed in Web Appendix A) which allows for clustering of residuals at the weekly level (Cameron, Gelbach, and Miller 2008).

The first column of Table 2 presents estimates from a regression that includes weekly fixed effects. We find significant negative effects for the numerical attributes. Given our log-log specification, the estimates suggest that a percent increase in the ID number of the item corresponds to a 0.16 percent decrease in transaction price, whereas a percent increase in generation corresponds to about a percent decrease in transaction price. Note that these estimates control for the visual attributes of an item, implying that participants value blockchain-based identifiers of "older" digital items, even if the items look similar to those created more recently. In terms of the economic magnitude of the impact of IDs, the 45th item released (corresponding to the lowest ID observed in the sample), attracts a sale price about 1.5 times higher than the 1000th item, and about 5 times higher than the millionth item, controlling for all other attributes. The estimates also imply that high ID numbers and high generation numbers lead to sale prices near zero, as we would expect for easily-duplicated image files with no scarce attributes. These findings are largely similar to findings in the work on physical collectibles.

Table 2: Price Regressions

	i	ii	iii	iv	v
log(ID)	-0.161***	-0.176***	-0.183***	-0.153***	-0.082*
	(0.045)	(0.044)	(0.042)	(0.046)	(0.044)
log(Generation)	-1.007***	-0.969***	-0.990***	-0.996***	-0.717***
	(0.042)	(0.040)	(0.041)	(0.041)	(0.033)
Attributes	Y	Y	Y	Y	Y
Text Variables				Y	
Seller FE					Y
Week FE	Y			Y	Y
Day FE			Y		
R^2	0.427	0.384	0.460	0.430	0.296
Observations	35,766	35,766	35,766	35,766	35,228

Note: *: p<0.1, **: p<0.05, ***: p<0.01. The R^2 in column v excludes the variation explained by seller fixed effects as these are differenced out.

We explore the extent to which time and additional controls can explain the variation in transaction prices. As we model transaction prices in ETH, the varying ETH/USD exchange rate may contribute to the price patterns we observe, and time-specific fixed effects capture this variation. The second column of Table 2 shows that removing time fixed effects does not significantly alter the parameter estimates and only slightly worsens the fit of the model based on the R^2. The third column replaces weekly fixed effects with more granular daily fixed effects and demonstrates that the additional granularity does not lead to significant changes. The fourth column shows that including text variables based on the item name and description as additional controls have a small contribution compared to the attributes of the item (see Web Appendix B for details). The fifth column similarly confirms that the inclusion of seller fixed effects does not qualitatively change the results, although the estimates do become slightly smaller in magnitude. We explore the possibility of sellers attracting heterogeneous sets of buyers or using different advertising strategies to promote their NFTs and the potential impact on our findings in Web Appendix C. The R^2 of the regression in column v, which captures variation explained by the attributes within sellers, does not fall significantly. As a result, variation in attributes across items primarily explains the price variation across transactions.

Seller Pricing Behavior

Having demonstrated systematic differences in sales prices across different types of NFTs that are driven mostly by NFT characteristics, we now demonstrate that sellers may not be entirely certain about how to value and price their items. As previously described in Table 1, sellers tend to set very wide price ranges. We conduct additional regressions to study how much of the variation in transaction prices can be explained by the pricing decisions themselves.

Note that hedonic price regressions do not require that sellers set prices independently of item attributes. On the contrary, the research on hedonic price regressions often assumes that the pricing data result from some equilibrium behavior where competing sellers set prices to maximize their profits (Pakes 2003). If a seller expects scarce attributes to command a higher price, she may naturally set a higher price for items with such attributes. This argument is often used as support for hedonic regressions as tools for item valuation. However, if sellers are not actually pricing optimally, the hedonic regressions estimated on observed sale prices may suffer from idiosyncrasies in seller pricing decisions.

We run a two-stage regression in which we regress the log of transaction prices on the auction design parameters in the first stage, and the resulting residuals on item attributes in the second stage. This procedure enables us to uncover how much of the variation in the transaction prices can be explained by item characteristics after removing the variation explained by auction design parameters like starting price and ending price. If sellers are confident about how to value their items, then auction design parameters should explain a significant portion of the variation in transaction prices. However, if sellers show high levels of uncertainty, then item characteristics should still explain some of the residual variation, suggesting that within the price ranges that sellers set, different types of items tend to sell for different prices. This would provide some evidence that sellers are not pricing optimally but rather setting very wide ranges in hopes of attracting a buyer at some price point.

Table 3: Seller Uncertainty About Item Valuations

	i	ii
log(ID)		0.051**
		(0.023)
log(Generation)		-0.118***
		(0.017)
log(Starting Price)	0.199***	
	(0.022)	
log(Ending Price$^+$)	0.776***	
	(0.020)	
log(Duration)	0.050***	
	(0.010)	
Attributes	N	Y
Week FE	Y	N
R^2	0.776	0.117
Observations	35,766	35,766

Note: *: p<0.1, **: p<0.05, ***: p<0.01.

The first column of Table 3 presents estimates from the first-stage regression. Auction design parameters and week fixed effects alone yield an R^2 statistic of 0.776. The second column shows the estimates from a regression of the residuals on item characteristics. We find that the coefficients remain statistically significant, suggesting that seller pricing decisions cannot fully explain the variation in the resulting transaction prices for items which vary along these dimensions. The coefficient on token ID reverses its sign, implying that conditional on prices set by the sellers, higher Token IDs actually sell for more. This may imply that sellers tend to overvalue the extent to which buyers may prefer low IDs. If a seller sets an excessively high price for low ID items, it may appear that conditional on these prices, higher IDs sell for more. Namely, participants are more likely to buy high ID items closer to the top of their price range, whereas they wait until the price falls closer to the bottom of the price range for low ID items. The estimate on the effect of generation shrinks relative to the regressions in Table 2 as the auction design parameters explain a significant amount of price variation, implying that sellers do tend to set higher starting and ending prices for items with low generation numbers. However, the fact that the coefficient remains significant and negative implies that sellers also tend to set wide price ranges which overlap for different types of items and cannot pinpoint the best price to set for the particular generation of their item.

Our descriptive results imply that sellers, while conscious of differences in buyer preferences across items, may be setting price ranges that are too wide and that may not accurately reflect optimal pricing decisions. An alternative explanation could be that it is actually optimal for sellers to set very wide price ranges given the descending auction mechanism – a hypothesis we later refute with our structural model. In addition to this evidence, there is abundant anecdotal evidence that sellers may not know how to optimally price. Namely, active forums dedicated to valuation exist in CryptoKitties social media communities such as Reddit[8] and Discord[9] where participants frequently ask other members for pricing help.

These pricing inaccuracies may limit our ability to recover accurate valuations using hedonic regression approaches. We explore this further by developing a structural model of descending auctions in this market. We use the model to infer optimal prices, compare them to actual seller pricing decisions, and identify the biases that may be present in hedonic regression valuation approaches that only use observed sale price data. We use a machine learning-based hedonic regression approach to demonstrate that the biases persist even with more flexible models.

Structural Model

We develop a stylized dynamic structural model of buyer behavior in this market. The model will capture two main incentives. First, buyers have a *dynamic* incentive to wait when making their purchase decisions as they know that given the descending auction mechanism the price of the item they are considering will fall over time. Second, buyers may not wish to wait too long because the item they are considering may be purchased by another buyer – a *competitive* incentive. These counteracting incentives will interact to

[8] Several threads on https://www.reddit.com/r/CryptoKitties/ ask participants to value a user's CryptoKitties.
[9] The channel labelled "kitty-appraisals" on https://discord.com/invite/3GvT66U is dedicated entirely to peer discussions about CryptoKitty valuation.

determine the buyer's choice of when to purchase an NFT and at what price. Buyers may also choose not to make a purchase. Our framework can be viewed as an explicit modelling of daily decisions made by potential buyers in a descending auction that is sensitive to auction duration. Most structural auctions models do not consider duration or model daily decisions but derive optimal bids given the auction mechanism and match them to observed data (Lafont and Vuong 1996). We can go a level deeper with our data and explicitly model daily decisions to wait or purchase.

However, given data limitations and modelling complexity, we must make several simplifying assumptions. First, we focus on each descending auction in isolation and do not consider interactions or competition for buyers between auctions. This is a typical assumption in empirical research on auctions and alternative selling mechanisms (Boudreau, Lakhani and Menietti 2016, Kireyev 2020, Yao and Mela 2008, Yoganarasimhan 2013). Second, we do not have access to data on consideration sets and model each auction as being considered by two identical, representative buyers. This is the minimum number of buyers to capture the *competitive* incentive. While this assumption is clearly a simplification, it allows us to capture the main incentives faced by buyers in the market. We test this assumption in Web Appendix C. Third, we assume that the two buyers make their decisions in sequence, with the first buyer always making the first move. Sequential decision reduces the possibility of multiple equilibria and streamlines estimation.

The resulting model is a two-agent sequential finite horizon dynamic game for each auction. Without loss of generality, we assume that agent 1 always moves first. More formally, the utility of agent $i = 1,2$ for purchasing item j at time period $t \in \{1, \dots, T_j\}$ in the auction is given by

$$u_{ijt}^{B} = \alpha + \beta p_{jt} + \gamma X_{jt} + \epsilon_{ijt}^{B}$$

where p_{jt} is the price of the item at time t, X_{jt} includes an intercept for the first day of the auction (to capture possible promotion of new auctions on the marketplace website), a set of characteristics of the item, including token ID and generation, as well as a measure for the "attractiveness" of the item, created by summing the fixed effects corresponding to all of the possible categorical features of the item from a first-stage logistic regression of sale incidence on item characteristics. We recover attractiveness in this fashion, as opposed to directly including the hundreds of attribute indicator variables in the model, to reduce dimensionality and accelerate estimation. Based on our descriptive analysis of buyer preferences, omitted variables such as text descriptions of the items or time fixed effects do not appear incredibly relevant and are excluded from the model. Furthermore, the fact that the appearance of a CryptoKitty can be explained entirely by its attributes limits concerns about omitted attributes. The remaining terms correspond to the coefficients of the model – α, β and the vector γ – and the error term ϵ_{ijt}^{B} which we assume follows a T1EV distribution. The utility for choosing not to purchase the item, but rather to wait in the given time period is

$$u_{ijt}^{W} = \epsilon_{ijt}^{W} + \theta_{ijt}\mathbb{E}_{\epsilon}V_{t+1}(p_{jt+1}, X_{jt+1})$$

where ϵ_{ijt}^{W} is a T1EV error associated with waiting, $\mathbb{E}_{\epsilon}V(p_{jt+1}, X_{jt+1})$ is the expectation (taken over the error terms) of the value function associated with waiting, which will take as arguments the price and characteristics of the item and auction in the following period. For periods $t = 1, \dots, T_j - 1$, where T_j is the final period of the auction, this expected value function takes the recursive form

$$\mathbb{E}_{\epsilon}V_t(p_{jt}, X_{jt}) = \log\left(e^{\alpha+\beta p_{jt}+\gamma X_{jt}} + e^{\theta_{ijt+1}\mathbb{E}_{\epsilon}V_{t+1}(p_{jt+1}, X_{jt+1})}\right).$$

In the final period, it takes on the form

$$\mathbb{E}_{\epsilon}V_{T_j}\left(p_{jT_j}, X_{jT_j}\right) = \log\left(e^{\alpha+\beta p_{jT_j}+\gamma X_{jT_j}} + 1\right)$$

as the deterministic utility obtained from waiting in the final period is set to zero.

The remaining term θ_{ijt} captures the probability that the game will not end from the perspective of agent i at time t. The game may end if the other player purchases the item. From the perspective of agent 1, who always moves first, θ_{1jt} is the probability that agent 2 does not purchase the item in period t. From the perspective of agent 2, who always moves second, θ_{2jt} is the probability that agent 1 does not purchase the item at the start of period $t + 1$. Given the very slight difference in timing and the fact that the two agents have identical preferences, θ_{1jt} is approximately equal to θ_{2jt}, a fact that we make use of in estimation as we recover these probabilities directly from the data using a flexible machine learning model before estimating the parameters of the structural model. Note that our model excludes a traditional discount factor term. This is because the auctions we consider tend to be very short.

Estimation

We estimate the model in two stages. The first stage involves recovering certain inputs directly from the data. First, to recover an "attractiveness" measure we run a logistic regression of sale incidence for item j in period t on the item's price in period t, its token ID, generation number, and a large set of categorical fixed characteristics that we used in earlier regressions. We recover the item's attractiveness as the log-odds prediction from this model after removing the intercept and the effects of price, token ID, and generation. We model attractiveness as a single variable derived from this simplified model as it is infeasible to include hundreds of indicator variables in the structural model without making the estimation time indefinite. Second, we estimate the probabilities θ_{ijt} that the game will continue into the subsequent stage (in the style of Hotz and Miller 1993). As $\theta_{1jt} \approx \theta_{2jt}$ we simply aim to estimate one quantity θ_{jt} for both agents. We do this by using a gradient-boosted trees model (Chen and Guestrin 2016) to predict the probability of a non-sale for item j at time t as a function of the same variables that enter the structural model, including the item's price, its token ID, generation, and attractiveness. We also include time-related variables in the machine learning model to proxy for the dynamic incentives. These variables include time to end of auction and the duration of the auction. We generate cross-validation predictions from this model, ensuring that the data on which we generate predictions is not the same as the data used to train the model and inform its hyperparameters.[10] The resulting prediction gives us the nonparametrically estimated probability θ^*_{jt} that a sale will not occur for item j at time t. This is the probability that both agents do not purchase the item at that time. Given $\theta_{1jt} \approx \theta_{2jt} = \theta_{jt}$, the probability that neither agent purchases the

[10] The machine learning model has a cross-validated AUC of 92.9%.

item at time t is given by $\sqrt{\theta_{jt}^{*}} = \theta_{jt}$. We use this quantity for both agents in the structural model as the probability that the game does not end after their turn.

In the second stage, we estimate the structural model using maximum likelihood methods. For each auction we observe either no sale or the day on which the sale has occurred. In the case of a sale, we do not know which agent made the purchase but, given their symmetry, can nevertheless derive a likelihood function which can be maximized to obtain parameter estimates. Let $\hat{u}_{ijt}^{W} = u_{ijt}^{W} - \epsilon_{ijt}^{W}$ and $\hat{u}_{ijt}^{B} = u_{ijt}^{B} - \epsilon_{ijt}^{B}$, and drop the i subscript given the symmetry of the deterministic utility component of the two agents. Let NS_{jt} denote the event that no sale occurs and S_{jt} denote the event that a sale does occur for item j at time t. Then the probability of a sale not occurring for item j at time t is

$$\Pr\{NS_{jt}\} = \left(\frac{e^{\hat{u}_{jt}^{W}}}{e^{\hat{u}_{jt}^{B}} + e^{\hat{u}_{jt}^{W}}}\right)^{2}$$

which corresponds to the probability that both agent 1 and agent 2 choose to wait. The probability that a sale does occur for item j at time t is

$$\Pr\{S_{jt}\} = \left(\frac{e^{\hat{u}_{jt}^{B}}}{e^{\hat{u}_{jt}^{B}} + e^{\hat{u}_{jt}^{W}}}\right) + \left(\frac{e^{\hat{u}_{jt}^{W}}}{e^{\hat{u}_{jt}^{B}} + e^{\hat{u}_{jt}^{W}}}\right)\left(\frac{e^{\hat{u}_{jt}^{B}}}{e^{\hat{u}_{jt}^{B}} + e^{\hat{u}_{jt}^{W}}}\right)$$

where the first term is the probability that agent 1 purchases the item (in this case agent 2 does not have an opportunity to purchase it as this agent moves second) and the second term is the probability that agent 1 decides to wait but agent 2 decides to purchase the item. Let T_j^{*} denote the time period when a sale occurs in the data, with $T_j^{*} = T_j$ when no sale occurs. Then, the probability of observing a sequence of decisions for an auction/item j is given by

$$\prod_{t=1}^{T_j^{*}} \Pr\{NS_{jt}\}^{NS_{jt}} \Pr\{S_{jt}\}^{S_{jt}}.$$

In other words, we multiply the probabilities of no sale occurring for each day that the sale does not occur. If a sale does occur, we multiply all preceding non-sale probabilities by the probability of observing a sale on that day. We do not consider any subsequent time periods as the item is no longer available on the market and no decisions can be made. If no sale ever occurs for the item, we simply multiply the probabilities of no sale occurring until the expiration of the auction.

The likelihood of the data given a parameter vector can be obtained by multiplying the expression above for all auctions in the data. We take logs to transform the likelihood into an additive form and use Nelder-Mead optimization to find the parameters that maximize the log-likelihood. We obtain standard errors for the parameter estimates by bootstrapping auctions 100 times and taking the standard deviations of the resulting estimates across the bootstrap iterations.

Discussion and Limitations

The most obvious limitation is data-related. We do not observe consideration sets and cannot model decisions by the multiple agents who may visit the auction at different times. This limitation is fairly standard given that the only way to obtain consideration sets is through browsing data which is usually unavailable. This may lead to errors in our estimates if the extent to which different auctions are considered is correlated with the characteristics of the item or the auction. We test the extent to which our results may differ by assuming that lower priced auctions receive more consideration, whereas the higher-priced auctions that do not sell do not receive any consideration. We find that our main results remain unchanged based on the analysis in Web Appendix C.

We are also unable to accurately model buyer heterogeneity as a result of this lack of data. Heterogeneity can affect the parameter estimates and imply different optimal pricing decisions. To investigate the potential impact of buyer heterogeneity, we allow for buyer preference parameters to differ across sellers. Different sellers may use different strategies to attract buyers to their auctions, which may result in buyer preference heterogeneity across sellers. We make use of the fact that a small number of sellers account for almost 50%

of the auctions in the data and estimate the structural model separately for each one of these sellers to recover a nonparametric heterogeneity distribution. We assume that all remaining sellers share the same preference parameter. Our main findings do not change. This analysis is available in Web Appendix C.

Results

Table 4 shows the parameter estimates from the structural model. The coefficients qualitatively resemble the estimates obtained in the descriptive hedonic regressions in Table 2. We find negative effects for both token ID and generation, and a positive effect for attractiveness. However, the hedonic regressions implied that generation was significantly more important than ID in determining buyer preferences, with a coefficient of around -1 for generation, roughly 6 times greater than -0.16, the coefficient for token ID. In contrast, the estimates from the structural model imply less of a difference between the two numerical traits, with the coefficient for generation only 2 times as large as the coefficient for token ID. These differences may result in different implications from the models when they are used to value NFTs. In addition, we find a strong negative effect of price on the purchase utility, as expected. We also find a positive effect of the first day intercept on the purchase utility, corresponding to the increased visibility and promotion that new auctions may receive on the platform.

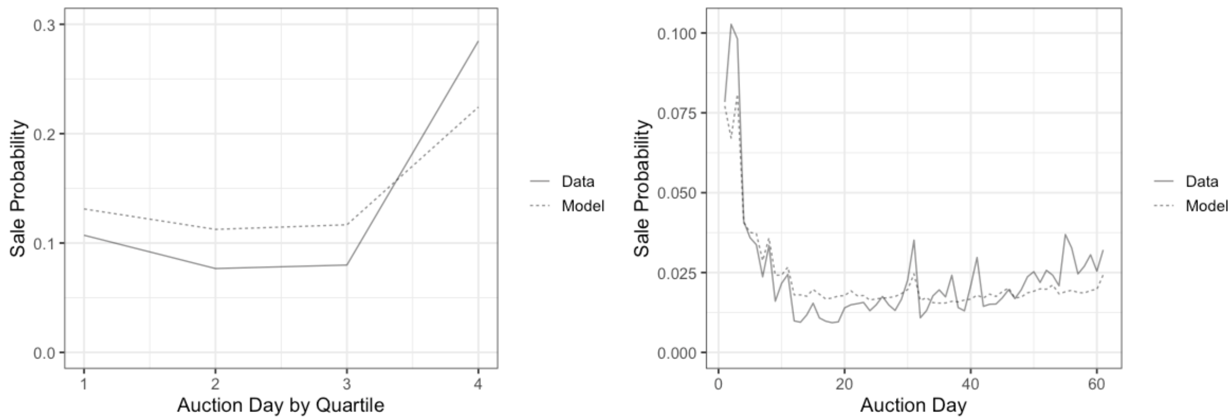
Table 4: Structural Model Estimates

	i
log(ID)	-0.324***
	(0.009)
log(Generation)	-0.636***
	(0.022)
log(Daily Price)	-1.360***
	(0.016)
Auction Day = 1	0.560***
	(0.021)
Attractiveness	1.149***
	(0.027)
Auction Days	840,282

Note: *: p<0.1, **: p<0.05, ***: p<0.01.

To evaluate the fit of the model, we perform several simulations. First, we obtain the probabilities of a successful sale as a function of the number of days that an auction has been active. The left panel of Figure 3 shows the average sale probability on the vertical axis and the day of the auction (as a quartile of total auction duration) on the horizontal axis. We split each auction duration into four quartiles to be able to compare across auctions of different durations. The left panel shows that our model is able to predict the increase in the sale probability that occurs when an auction approaches its end. The value of waiting drops significantly towards the end of the auctions which results in an increased purchase probability in the final quartile. The right panel of Figure 3 shows the probability of a successful sale as a function of the number of days since the start of the auction. The plot aggregates data across all auctions of different durations. The model is able to fit the data well and captures the phenomenon that most sales tend to occur early on, primarily driven by the fact that most auctions are short and benefit from a combination of the deadline effect highlighted earlier and the promotion of new auctions by the selling platform.

Figure 3: Structural Model Fit



As an additional test of model fit, we generate sale prices that would be predicted by the model for the auctions that did actually result in a sale in the data. We find a 98.7% correlation between the observed sale prices for the successful auctions and the predicted sale prices for these auctions, assuming they are successful. Overall, we find the model is able to predict sales incidence and sale prices conditional on a successful sale reasonably well.

Counterfactuals

We use the estimated model to generate counterfactual optimal prices. In Table 3, we showed evidence that sellers may not know how to price optimally in this market. As a result, it is unlikely that the prices observed in the data reflect optimal pricing decisions. We use the structural model to solve for optimal prices. We assume that as the cost of acquiring the item is a sunk cost, participants should set prices to maximize their expected revenues, not taking into account the cost of acquiring the item. In addition, we assume that participants obtain no value from holding the item if it does not sell. This assumption may result in slightly lower optimal prices than if we had assigned a positive future value to holding an unsold item, but it would not affect our main findings regarding the biases induced by hedonic regression valuation methods. Furthermore, participants need to incur a fee for cancelling and posting a new auction (referred to as a gas fee on the Ethereum blockchain) which reduces the future value they can expect from holding an unsold item, especially in the case of relatively low-priced NFTs like CryptoKitties. Incorporating these elements and simulating optimal prices with different optimization objectives is nevertheless feasible given our structural framework.

Comparing Optimal to Observed Prices

First, we compare the optimal prices generated from the structural model to the prices observed in the data in Table 5. We find that although the two differ, they are positively correlated. Both starting and ending prices, as well as the interval between the two prices are positively correlated across all auctions in the sample. However, the correlation between the optimal and observed price intervals is low and even negative for short auctions. Participants in the market appear to be pricing in the direction of the optimal prices, as we identified before with the descriptive regressions in Table 3, but the observed price intervals appear to be less consistent with optimal price intervals.

Table 5: Correlation between Actual and Optimal Prices

Duration	Starting Price	Ending Price	Price Interval
All	0.256	0.147	0.088
[1, 7]	0.276	0.200	-0.019
[8, 14]	0.265	0.229	-0.016
[15, 60]	0.275	0.381	0.127
Observations	72,299		

Table 6 displays the average starting and ending prices for the data compared to the model, as well as the actual average sales price for the successful auctions in the data. We remove the auctions with the highest 5% starting prices before generating this table to avoid outliers where participants set excessively high prices for their auctions (most of which, of course, do not result in a sale). The difference between starting and ending prices in the data is much greater than the optimal price intervals generated from the structural model, again consistent with the descriptive findings in Table 3 that suggested that observed price intervals may be too wide. Additionally, we find that the observed starting and ending prices are higher than the optimal prices on average across all auctions. Naturally, the observed prices for the successful auctions tend to be lower, whereas the structural model suggests similar average optimal prices for both (observed) successful and unsuccessful auctions. The average actual observed sale price falls within the bounds of the optimal starting and ending prices.

Table 6: Average Prices

	Starting Prices	Ending Prices	Actual Sale Prices
Data | All Auctions	0.036	0.011	
Model | All Auctions	0.018	0.009	
Data | Successful Auctions	0.026	0.007	0.012
Model | Successful Auctions	0.020	0.010	
Observations	68,693		

Biases of Hedonic Regression Approaches to Valuation

One approach to "value" an item is to predict its expected sale price, or the price at which it will sell conditional on selling, which is consistent with how valuations are established using traditional hedonic approaches. We use the counterfactual optimal prices generated from our structural model to highlight two biases induced by traditional valuation approaches that do not take into account the selling mechanism and the possibility of unsuccessful auctions. To quantify these biases, we require baseline valuations based on hedonic regressions.

Machine Learning Approach to Value Items

Research on collectibles has traditionally used hedonic price regressions to generate valuations and only recently have researchers begun to investigate the potential of using more sophisticated machine learning models instead (Aubry et al. 2019). We build a machine learning model which takes as input item characteristics and outputs a predicted transaction price of the item which we interpret as its valuation. This approach builds on the findings of our hedonic price regression analysis of buyer preferences in Table 2 but allows for more flexible interactions and the possibility of nonlinearities in the relationship between the input variables and the outcome.[11]

We estimate a gradient-boosted trees model (Friedman 2001) using the extreme gradient-boosting implementation (Chen and Guestrin 2016). This model combines the predictions of several decision trees built in sequence, such that each subsequent decision tree attempts to minimize errors for observations not well explained by a model consisting of all prior decision trees. The effectiveness of gradient-boosted trees models has been demonstrated across a variety of economic contexts (Chalfin et al. 2016; Kleinberg et al. 2018) making them a natural candidate for our application. To evaluate the effectiveness of our approach, we estimate the model on 50% of the transactions data and use the remaining 50% to evaluate the accuracy of the predictions. We use Bayesian hyperparameter optimization (Snoek, Larochelle, and Adams 2012) to identify the best parameters for maximum tree depth, subsampling rate, and minimum child weight based on our training-test data split, yielding an optimal maximum tree depth of 7, a subsampling rate of 100%, and a minimum child weight of 1, suggesting that substantial levels of interaction may be present between

[11] Qualitatively similar results hold if we use linear hedonic regression models as the baseline as well.

the input variables but no subsampling or adjustment to the weight of each tree leaf is required to maximize performance.[12]

Table 7 compares the out-of-sample performance of our machine learning model with the linear hedonic price regression estimated in column i of Table 2 as a benchmark. Unlike the machine learning model, we estimate the hedonic regression on the entire dataset and incorporate weekly fixed effects. As a result, the hedonic regression has an advantage as it uses data from the validation sample to obtain parameter estimates. Nevertheless, we find that the gradient-boosted trees model outperforms the hedonic regression model on the validation set based on both root-mean-squared error (RMSE) and out-of-sample R^2 suggesting that non-linearities and interactions may contribute to predictive accuracy.

Table 7: Out-of-Sample Model Performance

Model	RMSE	R^2
Gradient-Boosted Trees	0.868	0.511
Hedonic Linear Regression	0.952	0.412

Note: Table shows out-of-sample performance of our proposed model compared to a linear hedonic regression estimated on the entire dataset. The response variable is the log of the sales price. Both models are estimated only on data from successful auctions.

Next, we turn to comparing the biases induced by valuations generated from the hedonic machine learning model relative to the structural model.

Mispricing Bias

The first form of bias can result from the fact that observed sale prices on which any hedonic regression model is estimated may be the result of a suboptimal pricing decisions used by the seller. In particular, if sellers do not set optimal prices, the successful sale prices that we observe in the data would reflect these

[12] Minimum child weight is a measure of the minimum number of observations necessary to justify a split at each node in each decision tree and is an additional parameter used to control the complexity of the model.

seller idiosyncrasies as opposed to the true value of the item if it were sold under optimal conditions. To identify this *mispricing bias*, it is important to appropriately model the dynamic incentives of the buyers given the descending auctions mechanism used by this market.

This bias manifests in the systematic under-valuation of assets by the hedonic machine learning model. The sale price predicted by the hedonic model is lower than the optimal starting price 82.8% of the time and lower than the optimal ending price for 64.8% of the auctions in our sample. This occurs because most of the successful auctions in the data are the ones with low ending prices (recall Table 6 which shows that successful auctions in the data have an average ending price of 0.007 ETH, lower than the average optimal ending price of 0.009 ETH and the average observed ending price of 0.011 ETH across all auctions excluding outliers). The wide price intervals in the data also encourage buyers to wait longer before purchasing, resulting in successful sales at low prices. If the participants were pricing optimally, they would have set prices that are overall lower than the observed prices (recall Table 6), but the narrower optimal price ranges would have led participants to actually purchase at higher prices, increasing the implied valuations of the NFTs. Taking into account the dynamic incentives and potential mispricing in the market is crucial to avoid this bias in hedonic regression models.

Selection Bias

The second form of bias relates to the fact that hedonic models only use successful auctions and do not consider the possibility that some items did not sell in the data. We refer to this as a *selection bias* as it results from focusing on a selected sample of “successful” auctions in the hedonic regression approach. To study the effects of this bias, Table 8 splits the data into different subgroups and illustrates the relationship between the rate of sales and the tendency of the hedonic model to overvalue assets.

Table 8: Overpricing in the Hedonic Model

Subgroup	Observations	Selling Rate (%)	Overpriced by Hedonic Model (%)
Overpriced by Hedonic Model	17,536	40.20	100
Underpriced by Hedonic Model	54,763	52.44	0
High ID & High Generation	54,123	51.49	23.57
Low ID & High Generation	15,579	45.25	14.12
Low ID & Low Generation	2,466	32.97	99.27
High ID & Low Generation	131	28.24	100

The first two rows of Table 8 split the data based on whether or not the machine learning model over- or underprices the assets compared to the midpoint of the optimal price range. The "Selling Rate" column shows that items overpriced by the machine learning model tend to have a lower selling rate in the data. In other words, items with a lower selling rate tend to have higher prices, so when they do sell, they sell for a higher price. The hedonic model considers only successful sale prices and gives a higher valuation to these assets. The structural model takes into account the failures to sell and actually assigns lower optimal prices and hence a lower valuation to these items. More precisely, the machine learning model predicts an average price of 0.025 ETH for items that did not sell in the data and a price of 0.013 ETH for items that did sell in the data. The structural model generates average optimal starting and ending prices of 0.016 and 0.009 ETH, respectively, for items that did not sell and a slightly higher 0.020 - 0.010 ETH for items that did sell. The hedonic approach actually yields higher valuations for items from unsuccessful auctions compared to items from successful auctions, whereas the structural model does the opposite as it considers data from unsuccessful auctions.

We reinforce this result in the final four rows of Table 8. We split auctions based on the token ID and generation of the asset. We find that assets with a low generation tend to have the lowest selling rates and are also most often overpriced by the hedonic model relative to the structural model. The structural model takes selection into account and assigns a relatively lower valuation to items that frequently fail to sell, whereas the hedonic model only considers successful sales and may overvalue such items.

Conclusion and Practical Implementation

We provide one of the first analyses of pricing and valuation in a marketplace for nonfungible tokens. We show evidence that buyers value digital collectibles much like physical collectibles, but sellers may be uncertain of the value of their items as they tend to set sub-optimally wide price ranges in the descending auctions they organize. Our findings point to an inefficiency in the market that can create biases in hedonic regression approaches to valuation. We resolve this inefficiency by developing a structural model that accounts for the selling mechanism and characteristics of the market. The structural model yields optimal prices and helps us identify a *mispricing bias* and a *selection bias* that plague hedonic approaches to valuation.

We propose a proof-of-concept decision support interface based on our structural model which can be used by sellers to identify the optimal price ranges and valuations of their items and reduce the risk of mispricing. We incorporate uncertainty in optimal pricing by estimating 100 models on different bootstrapped versions of the data, and generating a distribution of optimal prices, sale probabilities, and predicted sale prices from these models.

Figure 4: Proof-of-Concept Decision Support Tool

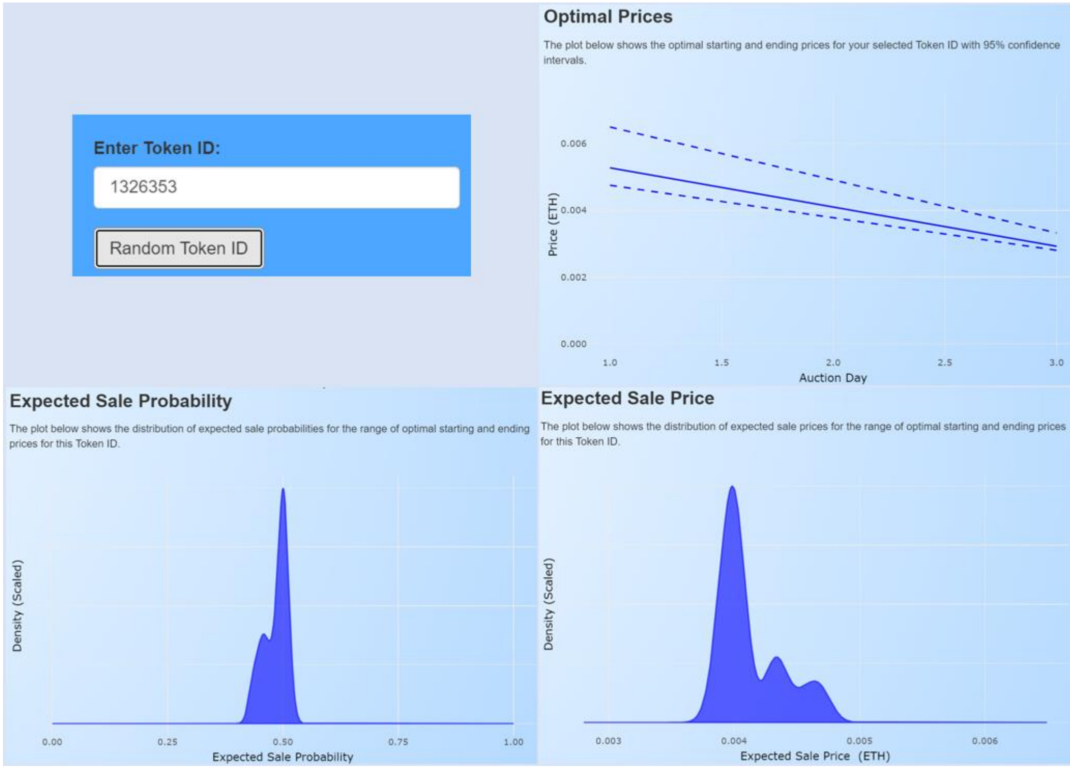


Figure 4 shows a screenshot from the decision support interface, and an interactive version is available at https://digital-assets.shinyapps.io/NFT_Pricing_and_Valuation_Interface for a random sample of 995 CryptoKitty tokens. Sellers enter the ID of the NFT they wish to price. In the background, bootstrapped structural models are used to obtain optimal starting and ending prices for a fixed auction duration of 3 days (which is the most common auction duration in the data). The output is an uncertainty distribution of the optimal starting and ending prices, sale probabilities, and expected sale prices conditional on a successful sale, which can be interpreted as valuations. The tool can be extended to cover a larger fraction of tokens and allow for variable auction durations. It can also be “productionized” to update model estimates as new data comes in from the market. We leave these engineering efforts to future research. Such a tool can help reduce inefficiencies in the market and can be offered to sellers as a stand-alone application or integrated with new marketplaces for nonfungible tokens or other types of digital collectibles. We hope this research can make NFT platforms aware of the potential biases induced by hedonic approaches to valuation and can provide a starting point for the development of valuation and pricing decision support tools in these new markets.

References

Agrawal A, Catalini C, Goldfarb A (2014) Some simple economics of crowdfunding. *Innov. Policy Econ.* 14(1):63-97.

Ashenfelter O, Graddy K (2003) Auctions and the price of art. *J. Econ. Lit.* 41(3):763-787.

Atasoy O, Morewedge CK (2018) Digital goods are valued less than physical goods. *J. Consum. Res.* 44(6):1343-1357.

Aubry M, Kräussl R, Manso G, Spaenjers C (2019) Machine learning, human experts, and the valuation of real assets. *Working Paper*.

Bajari P, Hortaçsu A (2003) The winner's curse, reserve prices, and endogenous entry: Empirical insights from eBay auctions. *RAND J. Econ.* 34(2):329-355.

Bajari P, Hortaçsu A (2004) Economic insights from internet auctions. *J. Econ. Lit.* 42(2):457-486.

Berry ST (1992) Estimation of a model of entry in the airline industry. *Econometrica*. 60(4):889-917.

Boudreau KJ, Lakhani KR, Menietti M (2016) Performance responses to competition across skill levels in rank-order tournaments: Field evidence and implications for tournament design. *RAND J. Econ.* 47(1):140-165.

Burton BJ, Jacobsen JP (1999) Measuring returns on investments in collectibles. *J. Econ. Perspect.* 13(4):193-212.

Cameron AC, Gelbach JB, Miller DL (2008) Bootstrap-based improvements for inference with clustered errors. *Rev. Econ. Stat.* 90(3):414-427.

Catalini C, Gans JS (2019) Some simple economics of the blockchain. *Commun. ACM.* Forthcoming.

Chalfin A, Danieli O, Hillis A, Jelveh Z, Luca M, Ludwig J, Mullainathan S (2016) Productivity and selection of human capital with machine learning. *Am. Econ. Rev.* 106(5):124-127.

Chen T, Guestrin C (2016) "Xgboost: A scalable tree boosting system. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* 785-794.

Doraszelski U, Lewis G, Pakes A (2018) Just starting out: Learning and equilibrium in a new market. *Am. Econ. Rev.* 108(3):565-615.

Ellickson PB, Misra S (2011) Estimating discrete games. *Market. Sci.* 30(6):997-1010.

Friedman JH (2001) Greedy function approximation: A gradient boosting machine. *Ann. Stat.* 29(5):1189-1232.

Halaburda H, Haeringer G (2019) Bitcoin and blockchain: What we know and what questions are still open. Working Paper, NYU Stern School of Business, New York.

Hotz JV, Miller RA (1993) Conditional choice probabilities and the estimation of dynamic models. *Rev. Econ. Stud.* 60(3):497-529.

Huang Y, Ellickson PB, Lovett MJ (2018) Learning to set prices. *Working Paper*.

Kireyev P (2020) Markets for ideas: Prize structure, entry limits, and the design of ideation contests. *RAND J. Econ.* 51(2):563-588.

Koford K, Tschoegl AE (1998) The market value of rarity. *J. Econ. Behav. Organ.* 34(3):445-457.

Kleinberg J, Lakkaraju H, Leskovec J, Ludwig J, Mullainathan S (2018) Human decisions and machine predictions. *Q. J. Econ.* 133(1):237-293.

Laffont JJ, Vuong Q (1996) Structural analysis of auction data. *Am. Econ. Rev.* 86(2):414-420.

Lehdonvirta V, Virtanen P (2010) A new frontier in digital content policy: Case studies in the regulation of virtual goods and artificial scarcity. *Policy Internet.* 2(3):7-29.

Lewis G (2011) Asymmetric information, adverse selection and online disclosure: The case of eBay motors. *Am. Econ. Rev.* 101(4):1535-1546.

Liebowitz SJ (2006) File sharing: creative destruction or just plain destruction? *J. Law Econ.* 49(1):1-28.

Lucking-Reiley D (1999) Using field experiments to test equivalence between auction formats: Magic on the internet. *Am. Econ. Rev.* 89(5):1063-1080.

Pakes A (2003) A reconsideration of hedonic price indexes with an application to PC's. *Am. Econ. Rev.* 93(5):1578-1596.

Parker JR, Lehmann DR (2011) When shelf-based scarcity impacts consumer preferences. *J. Retailing.* 87(2):142-155.

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

Rust J (1987) Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica.* 55(5):999-1033.

Serada A, Sihvonen T, Harviainen JT (2020) CryptoKitties and the new ludic economy: How blockchain introduces value, ownership, and scarcity in digital gaming. *Games Cult*. 16(4):457-480

Smith RK, Newman GE, Dhar R (2016) Closer to the creator: Temporal contagion explains the preference for earlier serial numbers. *J. Consum. Res.* 42(5):653-668.

Snoek J, Larochelle H, Adams RP (2012) Practical Bayesian optimization of machine learning algorithms. *Adv. Neur. In.* 25:2951-2959.

Stock A, Balachander S (2005) The making of a “hot product”: A signaling explanation of marketers’ scarcity strategy. *Manage. Sci.* 51(8):1181-1192.

Verhallen TM (1982) Scarcity and consumer choice behavior. *J. Econ. Psychol.* 2(4):299-322.

Yao S, Mela CF (2008) Online auction demand. *Market. Sci.* 27(5):861-885.

Yoganarasimhan H (2013) The value of reputation in an online freelance marketplace. *Market. Sci.* 32(6):860-891.

Web Appendix

A. Bootstrap Procedures

We use a bootstrap procedure based on Cameron, Gelbach, and Miller (2008). The procedure obtains standard errors for the first-stage hedonic price regressions (Equation 1) accounting for clustered errors at the weekly level. We apply this procedure to obtain the estimates in all cases unless otherwise specified in the main text. It involves the following steps:

1) Resample each week t with replacement such that the total number of weeks is the same as in the original dataset. Assign a unique identifier to each resampled week.
2) On the resampled dataset, estimate the specified hedonic price regression.
3) Repeat the first two steps 400 times and store the resulting estimates for each resampled dataset.
4) For each parameter, report its average value across all resampled datasets. Obtain standard errors as the standard deviation of each parameter across all resampled dataset.

In the case of the structural model, we resample auctions to obtain standard errors. We perform the resampling 100 times and use the bootstrapped datasets to obtain optimal prices and to construct the statistics for our decision support interface.

B. Attribute Endogeneity and Seller Advertising

In this section, we consider the possibility that items with scarce attributes or low numerical identifiers may possess correlated but omitted characteristics that influence sales prices. As discussed before, the set of visual characteristics uniquely describes the appearance of the NFTs we consider. However, additional unstructured text-based characteristics or information about the seller may affect participant decisions. We elaborate on the estimates in column iv of Table 2 in the main text which shows that including additional text-based characteristics does not significantly affect the estimates of the key parameters.

Each item has a name, which can be modified by its owner, and an algorithmically-generated description, which cannot be modified. The owner may change the name of the item to promote some of its characteristics, which may influence sales and result in an overestimate of the effect of any correlated attribute. Similarly, items with certain characteristics may have more unique or interesting algorithmically-generated descriptions, which may affect buyer and seller behavior.

Table A1 shows the results of a regression which incorporates a number of text and sentiment variables for both the name and the description (Mohammad and Turney 2013). We do not attempt to assign a causal interpretation to the parameters for the text variables but rather treat them as a set of controls. We add an indicator for whether or not an item's name includes the word "Gen" as some sellers promote their item by emphasizing its low generation, but find no significant effect here. However, we do find that items with an emoji in their description tend to sell for more. We include additional sentiment variables but do not find that they influence our findings in any meaningful way although some of the coefficients are statistically significant. Overall, textual variables do not change the main results in Table 2. They also suggest that seller's attempts to promote items by changing their name may not be extremely effective, reducing concerns about seller advertising confounding the effects of other attributes of the item.

Table A1: Text Characteristics

		Text Variable	Name	Description
No Name	0.013	Emoji	0.045	0.002**
	(0.118)		(0.030)	(0.001)
“Gen” in Name	0.090	log(1+Unknown Words)	-0.109***	0.016
	(0.060)		(0.032)	(0.018)
		log(1+Characters)	0.055*	0.030
			(0.031)	(0.022)
		Anger	-0.150	-0.004
			(0.109)	(0.013)
		Anticipation	-0.278**	0.001
			(0.114)	(0.007)
		Disgust	0.084	-0.020**
			(0.090)	(0.010)
		Fear	-0.013	0.005
			(0.113)	(0.010)
		Joy	0.280**	-0.013
			(0.117)	(0.008)
		Sadness	0.033	0.020**
			(0.052)	(0.010)
		Surprise	0.039	-0.016*
			(0.075)	(0.009)
		Trust	-0.042	-0.002
			(0.038)	(0.008)
		Negative	-0.078**	0.005
			(0.037)	(0.007)
		Positive	-0.017	0.012**
			(0.032)	(0.006)
Attributes	Y			
Week FE	Y			
R^2	0.464			
Observations	35,766			

Note: *: p<0.1, **: p<0.05, ***: p<0.01.

C. Robustness to Varying Buyer Consideration Sets and Seller Heterogeneity

We test the robustness of our structural model. First, we examine the assumption that each auction is considered by two buyers. It may be the case that auctions that have a high price are less likely to be considered by potential buyers. Then, our price estimate may be too negative if we misclassify auctions that were not actually considered by many buyers as auctions where the buyers explicitly chose not to make a purchase. To test the impact of this variation in considered auctions, we estimate the structural model assuming that high-priced auctions that did not result in a sale were not actually considered by any buyers.

This amounts to excluding the 25% highest-priced unsuccessful auctions (based on their ending price) from estimation. Table A2 shows that the parameter estimates obtained from this subset of data do not differ significantly from the estimates we obtained on the full dataset, implying that the consideration set assumption is unlikely to significantly affect our findings.

Table A2: Structural Model Estimates Excluding High-Priced Unsuccessful Auctions

	i
log(ID)	-0.339***
	(0.018)
log(Generation)	-0.683***
	(0.030)
log(Daily Price)	-1.256***
	(0.023)
Auction Day = 1	0.315***
	(0.022)
Attractiveness	1.288***
	(0.036)
Auction Days	650,891

Note: *: p<0.1, **: p<0.05, ***: p<0.01.

Next, we turn to investigate heterogeneity. We take advantage of the fact that a small number of sellers account for a significant proportion of auctions in the data and estimate separate parameters for each one of these top sellers. We assume that all of the remaining sellers share the same set of parameters. These parameters can capture differences in buyer preferences across auctions posted by different sellers. In total, there are 29 sellers who form the top 0.8% of all sellers and account for 47% of all of the auctions in our data. The first two columns of Table A3 report the means and standard deviations of the estimates obtained from the structural model applied to each seller individually, corresponding to a non-parametric heterogeneity distribution across these top sellers. While the mean estimates appear to correspond to the estimates obtained in the homogenous model, the standard deviations indicate that sellers can differ in the preference parameters of their buyers. These differences could be driven by the approaches used by sellers to attract buyers to their auctions.

Table A3: Heterogeneous Seller Model Estimates

	(i) Mean	(ii) Standard Deviation	(iii) Remaining Sellers
log(ID)	-0.338	0.444	-0.212
log(Generation)	-0.755	1.016	-0.892
log(Daily Price)	-1.453	1.058	-1.235
Auction Day = 1	0.103	1.275	0.317
Attractiveness	0.877	0.560	1.083
Auction Days	535,483		304,799

The third column of Table A3 shows the estimates obtained for the remaining 3,596 sellers assuming that they all share the same set of parameters. These estimates largely correspond to the estimates obtained from the homogenous model estimated on all sellers.

The heterogeneous estimates can be used to test the extent to which our main findings may differ if we allow for differences in buyer types across auctions posted by different sellers. First, we check how the optimal prices obtained from the heterogeneous model compare to the prices predicted from our hedonic machine learning model.[13] The sale price predicted by the hedonic model is lower than the optimal starting price 89.5% of the time and lower than the optimal ending price for 76.5% of all auctions, consistent with our findings in the main analysis. The *mispricing bias* persists – the hedonic model tends to underestimate item valuations because it does not consider the selling mechanism used. Second, we verify if items that have a lower selling rate in the data tend to be overvalued by the hedonic model. Table A4 replicates the findings of Table 8 in the main text and confirms that *selection bias* also persists when we allow for heterogeneity. Namely, as the hedonic model considers only data from successful sales it is more likely to overvalue items that have a low selling rate in the data.

[13] For 9 of the 29 heterogeneous sellers, demand does not appear very sensitive to price, which may be caused by statistical error in the coefficient or because some sellers do not actually sell to the market but may be "wash-trading" or transferring items between accounts that they own. In other words, the coefficient on prices for these sellers is greater than -1, which implies that they can set infinitely high prices as they are on an inelastic portion of the demand curve. As it is unlikely that infinite prices are optimal, we fix the price coefficient for these sellers to the highest estimate below -1 obtained for the remaining sellers. Removing these sellers does not affect our results either.

Table A4: Overpricing in the Hedonic Model Compared to Heterogeneous Dynamic Model

Subgroup	Observations	Selling Rate (%)	Overpriced by Hedonic Model (%)
Overpriced by Hedonic Model	10,611	42.89	100
Underpriced by Hedonic Model	61,687	50.60	0
High ID & High Generation	54,123	51.49	14.06
Low ID & High Generation	15,579	45.25	6.55
Low ID & Low Generation	2,465	32.94	76.63
High ID & Low Generation	131	28.24	68.70

Web Appendix References

Cameron AC, Gelbach JB, Miller DL (2008) Bootstrap-based improvements for inference with clustered errors. *Rev. Econ. Stat.* 90(3):414-427.

Mohammad SM, Turney PD (2013) Crowdsourcing a word–emotion association lexicon. *Comput. Intell.* 29(3):436-465.