Biased Auctioneers

Mathieu Aubry, Roman Kräussl, Gustavo Manso, and Christophe Spaenjers*

This version: June 30, 2021

Abstract

We construct a neural network algorithm that generates price predictions for art at auction, relying on both visual and non-visual object characteristics. We find that higher automated valuations relative to auction house pre-sale estimates are associated with substantially higher price-to-estimate ratios and lower buy-in rates, pointing to estimates' informational inefficiency. The relative contribution of machine learning is higher for artists with less dispersed and lower average prices. Furthermore, we show that auctioneers' prediction errors are persistent both at the artist and at the auction house level, and hence directly predictable themselves using information on past errors.

Keywords: art, auctions, experts, asset valuation, biases, machine learning, computer vision.

JEL Codes: C50, D44, G12, Z11.

*Aubry (mathieu.aubry@imagine.enpc.fr) is at École des Ponts ParisTech, Kräussl (roman.kraussl@uni.lu) is at the University of Luxembourg and at Hoover Institution, Stanford University, Manso (manso@haas.berkeley.edu) is at Haas School of Business, University of California at Berkeley, and Spaenjers (spaenjers@hec.fr) is at HEC Paris. The authors thank Simona Abis (discussant), Will Goetzmann, Katy Graddy, Mara Lederman (discussant), Stefano Lovo, Ramana Nanda (discussant), Christophe Pérignon, Luc Renneboog, Donghwa Shin (discussant), Kelly Shue, David Sraer, Lea Stern (discussant), Scott Stern, and conference participants at the 2018 ULB Art Market Workshop, the 2019 HEC Paris Data Day, the 2019 Northeastern University Finance Conference, the 2019 NBER Economics of AI Conference, the 2019 Miami Behavioral Finance Conference, the 2020 GSU-RFS FinTech Conference, and the 2020 WFA for helpful comments. This research is supported by the following grants of the French National Research Agency (ANR): EnHerit (ANR-17-CE23-0008) and "Investissements d'Avenir" (LabEx Ecodec/ANR-11-LABX-0047).

Auction houses do not just match consignors to bidders. They also act as brokers of information. In particular, they publicly communicate market value estimates of the lots for sale. Even though auction theory suggests that "honesty is the best policy" (Milgrom and Weber, 1982), at least for a monopolistic auctioneer, there are good reasons to think that pre-sale estimates may not be unbiased. Art and other collectibles' illiquidity and heterogeneity make the task of valuation far from obvious (Chambers *et al.*, 2020). Moreover, both behavioral frictions and strategic-competitive considerations can impact auction houses' proclaimed valuations.

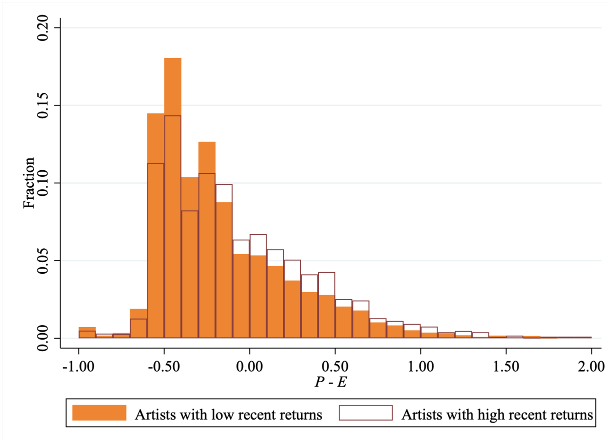
To study whether any individual behavioral or strategic bias systematically skews estimates, one can correlate auctioneers' prediction errors with a proxy for the driver of the hypothesized bias.[1] Let us give a simple example. Prior work suggests that intermediaries in real asset markets are slow to adjust their appraisals, especially downwards. Such behavior can be traced back to cognitive biases, but may also reflect strategic incentives to avoid signaling market downturns to collector-investors (Brown and Matysiak, 2000; Velthuis, 2007; Dimson and Spaenjers, 2011). We can thus hypothesize that recent (adverse) price trends will not always be reflected in auction house estimates. Using the art sales data presented later in this paper, we first compute artist-level average annualized returns over the period 2008–2014, and then compare the distributions of logged price-to-estimate ratios in 2015 for lots by artists in the bottom quartile of recent returns to those by artists in the top quartile. Figure 1 confirms that auctioneers' ex-ante value assessments are more likely to overshoot (undershoot) ex-post transaction prices for artists with low (high) recent returns.

Exercises like these can help to identify specific drivers of auctioneers' prediction errors. However, the pattern shown in Figure 1 may co-exist with many other—possibly hard to precisely and separately define and measure—sources of systematic variation in the data. Furthermore, estimates may have an idiosyncratic error component as well. Any single cut of the data thus tells us little about the overall informational efficiency of pre-sale estimates. One way to gauge whether pre-sale estimates can be improved upon as predictors of cross-sectional variation in transaction prices in an economically meaningful way, is to come up with reasonable counterfactual estimates

[1] Even though we will explicitly recognize that pre-sale estimates may sometimes be set relatively low or high on purpose for strategic reasons, we will for simplicity speak about "errors" in prediction or valuation.

Figure 1: **Motivating example**

This figure shows two different distributions of $P - E$, winsorized at -1 and $+2$, where P is the log of the hammer price (imputed at 75% of the low estimate in case of a buy-in, where the highest bid remains below the reserve price) and E is the log of the mean pre-sale estimate. The distributions are based on art auctions in 2015. We classify all lots in quartiles based on the artist-level average log return on observed resales over the period 2008–2014. We then compare the distribution for the first quartile ("Artists with low recent returns") to that the fourth quartile ("Artists with high recent returns"). More information on our data can be found in Section 2.



that auctioneers *could have picked* given their information set. Yet, constructing such benchmark valuations is not a straightforward task when relying on standard statistical tools such as linear regression models, in particular when these are estimated based on relatively small data sets. This may explain the conflicting conclusions in prior work on auction house biases (see Ashenfelter and Graddy (2006) for a review of the literature).

Our paper makes progress on this front by creating a novel statistical algorithm that generates automated valuations of artworks based on a large database of past auction outcomes. To do so, we build on recent advances in machine learning and computer vision.[2] A conceptual rather than methodological novelty of our paper is that we will use our technology not just to generate predictions of *prices* that can be compared to auction house estimates, but also to directly generate predictions of *auctioneers' estimation errors*. Like our benchmark analysis, such a statistical approach to studying estimates' informational efficiency does not necessitate any prior hypothesis or knowledge on what biases auction houses are subject to. Of course, auction house prediction errors will

[2]Our paper is the first to directly include images in a statistical model of art pricing. (Pownall and Graddy (2016) and Ma *et al.* (2019) measure specific visual properties, which they use then as inputs in a hedonic model.) Our methods are similar to those of Glaeser *et al.* (2018), who study how house prices are affected by their appearance.

only be predictable out-of-sample if those biases and the resulting errors are sufficiently persistent. To our knowledge, the persistence and predictability of auctioneers' under- and overvaluation patterns has not been examined before.

Our research starts from data on 1.2 million painting auctions from a proprietary database of art sales. The data capture the near-totality of global art auction activity over our sample period 2008–2015. For each lot, the database contains detailed information related to the artist, the artwork, and the auction. It also includes an image of each item. Nearly every artwork in our database is associated with a low and a high pre-sale estimate issued by the auction house; we will work with the mean value in our analysis. If the item sells at auction, we observe its hammer price; otherwise, we know that it has been "bought in". The distribution of art prices has a very long right tail: while the median hammer price in our data set is $3,271, the average equals $61,225. Our initial database includes hundreds of auction houses, but the top three (Christie's, Sotheby's, and Bonhams) account for 22% of all observations—and 70% of aggregate dollar volume.

We use the price data for the period 2008–2014 to generate price predictions for art objects auctioned in 2015 using machine learning. We train a neural network, which can be seen as a method to define very large parametric models, in which the parameters are learned from the observations in an iterative and stochastic manner. Because we want to use the picture of each artwork, we estimate a type of model—a "convolutional" neural network—that is often used for image-recognition tasks. Next to the image, our benchmark prediction relies on independent variables derived from the textual and numerical data in the database (e.g., artist, artwork materials, auction house). When constructing and estimating the model, we build in a number of features that help to avoid overfitting, in particular with respect to the artist dummies.

In order to be useful as counterfactual pre-sale valuations, our machine-learning price predictions need to be sufficiently accurate. We verify this in an out-of-sample test set of auctions that took place in 2015, where we first filtered out artists and auction houses that are economically of only minor importance. We find that our automated valuations explain nearly 75% of the transaction price variation in the test sample. We then let the neural network make new predictions after dropping different (sets of) variables, and study by how much the R-squared goes down. We find that artist-related information is much more relevant for price predictions than artwork properties. The incremental explanatory power of images is relatively limited, at least once conditioning on artist and artwork characteristics.

As a benchmark for our novel machine-learning valuations, we also use a standard linear hedonic model to generate price predictions for all test set lots. When relating price outcomes to the hedonic valuations, we find an R-squared of 67.6%, compared to 74.2% for the predictions coming out of our neural network. Because of the lack of interaction effects in a standard hedonic model, machine-learning valuations are much more likely to be accurate than hedonic valuations for works by high-volume and high-dispersion artists—which also tend to be more expensive on average.

Not surprisingly, even our most sophisticated automated valuations are a worse "predictor" of transaction prices than the auctioneers' pre-sale estimates (R-squared above 90%). Auction house experts have access to more qualitative information about the artwork (e.g., condition, provenance) and about the artist's place in art history than our algorithms. Also, if potential buyers anchor their valuations on publicly available auction house estimates, then the estimates may be endogenously correlated with price outcomes. We thus cannot conclude from our ex-post comparison of predictive power that humans beat machines in the task of art price prediction; it is in any case not a goal of this paper to set up such a "horse race".

We then study whether the relative level of our novel machine-generated price predictions (compared to auction house estimates) predicts relative price outcomes. We show that, after orthogonalizing with respect to the pre-sale estimates, our machine-learning valuations have economically and statistically significant explanatory power for price-to-estimate ratios. We can also expect an effect on buy-in rates, as consignors typically set their reserve at a level slightly below the low estimate provided by the auction house, implying a strong correlation between the two. Indeed, when our automated valuation is high relative to the auction house estimate, the buy-in probability is only about 25%, while this probability exceeds 45% when the prediction generated by the machine-learning algorithm is low relative to the auctioneer's valuation.

We dig deeper into the drivers of the *relative* usefulness of machine-based valuations, which is determined jointly by the accuracy of our automated predictions and that of the auction house estimates. We find that our machine-learning predictions—but not human experts—become predictably more accurate for objects by artists with a narrow range of price levels and for artworks that based on their characteristics can be expected to be easier to value in an automated fashion. Our neural network price predictions are also more likely to be more accurate than pre-sale estimates for less expensive artists, artists that are associated with high prediction errors by auctioneers historically, high-volume artists, and more recent artworks.[3]

If auctioneers are affected by systematic biases, then it might be possible to predict ex ante situations in which auctioneers' valuations are likely to be too optimistic or too pessimistic. We show that prediction errors are indeed persistent both at the artist and at the auction house level. We then use our statistical framework to directly predict the prediction errors of auctioneers. More precisely, we generate an ex-ante prediction of the price-to-estimate ratio for each lot, using exactly the same inputs as before (including the image), and also the pre-sale estimate. We find a substantial correlation between predicted and actual deviations of transaction prices from estimates. Also the buy-in rate decreases sharply in the *predicted* price-to-estimate ratio. This predictability exists even if the network does not have access to the auctioneer's estimate. Both the identity of the artist and the identity of the auctioneer are important drivers of this predictability, suggesting that both behavioral biases and strategic considerations may play a role.

Finally, we highlight that non-fundamental variation in auction house pre-sale estimates has real economic effects as it drives heterogeneity in art market participants' investment outcomes. Because consignors' reserve prices are highly correlated to auction house estimates, buy-in rates are higher if auctioneers are more optimistic. But also bidders may anchor on auction house estimates. We can therefore expect that artworks associated with more aggressive estimates—relative to our automated valuations—will have lower returns going forward.[4] We find suggestive evidence in

[3] As such, although the current paper does not set up a prediction horse race between humans and algorithms, we contribute to the discussion about the relative strengths and weaknesses of "men" vs. "machines" in financial-economic decision-making (e.g., Abis, 2020; Coleman *et al.*, 2020; Fuster *et al.*, 2020; Erel *et al.*, 2021) and that about the implications of machine learning for job occupations (e.g., Autor, 2015; Acemoglu and Restrepo, 2018; Agrawal *et al.*, 2018; Brynjolfsson *et al.*, 2018; Grennan and Michaely, 2020).

[4] Mei and Moses (2005) also study this hypothesis of "credulous" art buyers, and show that works with higher estimates have higher realized returns and lower future returns. However, as they do not control for the pricing of artwork characteristics, their results are also consistent with variation in estimates correctly anticipating patterns in bidders' willingness-to-pay.

support of this hypothesis using data on artwork *resales* over the period 2016–2018 for which the purchase is part of our year-2015 test data set.

While the empirical setting studied in this paper is the art auction market—for which large amounts of historical data on both prices and human experts' valuations are available—investigating the predictability both of prices and of biases in intermediaries' information provision clearly carries relevance for other real asset markets as well. Consider, in particular, real estate. While automated valuation models are increasingly common in the housing market, as illustrated by the rise of "iBuyers" (Buchak *et al.*, 2020), not much is known about their predictive power for transaction values—and about the heterogeneity therein. Also, it is clear that behavioral frictions and incentive issues can affect housing appraisals (Salzman and Zwinkels, 2017), and that real estate market participants may anchor their valuations on those of human experts.

The remainder of this paper is organized as follows. Section 1 provides more information on the empirical setting. Section 2 presents the data. Section 3 introduces our machine-learning algorithm, while 4 assesses its predictive power. Section 5 presents evidence that auction house estimates are informationally inefficient, and Section 6 shows that auctioneers' prediction errors are themselves predictable. Section 7 discusses some key take-aways and implications of our findings. Section 8 concludes.

1 Art Auctions and Pre-Sale Estimates

Art auctions are typically organized as "English" (i.e., ascending-bid, open-outcry) auctions. Each consignor sets a reserve price, which is the lowest price she is willing to accept, in agreement with the auction house. If the highest bid at the auction meets or exceeds the reserve price, the object will be sold at this price—the "hammer price".[5] If the highest bid remains below the reserve price, the item is said to be "bought in"; it does not sell and instead returns to the consignor.

Prior to most auctions, auction house experts publicly share a "low" and a "high" estimate for each lot. Artworks' market values are difficult to determine because every object is unique and

[5]The auction house will charge a "buyer's premium" on top of the hammer price. Moreover, the consignor has to pay a "seller's commission". We do not consider transaction costs here.

only trades infrequently. Nonetheless, thousands of objects are sold publicly at auction every year. Art auction sales databases thus contain a lot of information about collectors' willingness-to-pay, and auctioneers do consider recent auction prices for similar objects.

Auction house estimates are said to be representing auction house experts' opinion "about the range in which the lot might sell at auction". And, indeed, pre-sale estimates seem to be relatively accurate *on average*, at least once taking into account buy-ins (McAndrew *et al.*, 2012). Yet, at the same time, auction houses will typically argue that their estimates only serve as "an approximate guide to current market value and should not be interpreted as a representation or prediction of actual selling prices" (both quotes in this paragraph taken from Sotheby's website). The latter statement of course points to the fact that auction houses may sometimes strategically choose to be relatively aggressive or conservative in their estimates. On the one hand, higher estimates may be useful to lure consignors away from competitors (Gammon, 2019), or to increase bids by "credulous" investors on certain lots (Mei and Moses, 2005). On the other hand, lower estimates might steer consignors to lower reserves (thereby increasing sale rates), or to attract bidders to the auction. The upshot is that pre-sale estimates do not necessarily represent auctioneers' honest assessment of what they think the hammer price will be.

Next to (conscious) strategic biases, also (unconscious) behavioral biases in expectations formation may affect auctioneers' choices of pre-sale estimates. Prior research on illiquid real asset markets has shown that investors and intermediaries may suffer from biases related to extrapolation (Glaeser and Nathanson, 2017), reference dependence (Genesove and Mayer, 2001; Andersen *et al.*, 2021), anchoring (Beggs and Graddy, 2009), and confirmation bias (Eriksen *et al.*, 2020), among others.

Different strategic and behavioral biases can of course aggregate and interact in complicated ways. Auctioneers may moreover make idiosyncratic errors.[6] If the cumulative impact of these factors on auctioneers' estimates is sufficiently large, then it may be possible to improve upon these estimates as predictors of (cross-sectional variation in) transaction prices. Moreover, if auctioneers'

[6]More formally, suppose that it is an auctioneer's job to come up with an estimate V so that auction price $P = V + \epsilon$, where ϵ is a zero-mean noise term. Even in the absence of systematic biases, the estimate E may sometimes deviate in an idiosyncratic fashion from V, meaning that $E = V + \eta$.

biases and the resultant prediction errors are persistent, then it should be possible to pick up predictability in the distribution of deviations of ex-post transaction prices from ex-ante auction house estimates. In our empirical analysis, we will study both of these hypotheses.

2 Data

The analysis in this paper relies on proprietary data coming from the Blouin Art Sales Index, which tracks auction sales at hundreds of auction houses worldwide, including at Christie's and Sotheby's. (The database has been used before by Korteweg *et al.* (2016).) We use data on paintings offered at auctions over the period 2008–2015. In total, our data set contains information on 1.2 million lots at hundreds of auction houses of works by about 130,000 individual artists.

For each lot, the database contains information related to the artist (artist name, nationality, birth and death year, style), the artwork (year of creation, size, materials, title, markings such as signature or date), and the auction (auction house, auction location, auction date, pre-sale low and high estimates, a buy-in indicator, hammer price if sold). All price data are converted to U.S. dollars using the spot rate at the time of the sale. Uniquely, we also have access to a high-quality image of each painting.

About two thirds of these auction lots have been sold, while the remaining one third were bought in because the highest bid remained below the consignor's reserve price. Table 1 shows some information on the distribution of hammer prices for the overall data set and for the top-three most frequent artists, auction houses, and auction locations in our data set. The average hammer price is $61,225, while the median is $3,271, indicating a long right tail of very expensive paintings. The two top-selling artists over our sample period, Pablo Picasso and Andy Warhol, both have a mean hammer price exceeding one million dollars. Prices are also clearly higher-than-average at Christie's and Sotheby's, which are the two auction houses with the highest number of sales, followed by Bonhams. While Paris is the most frequently-observed auction location, prices are substantially lower there than in New York and London. Taken together, the top-three auction houses and auction locations account for 22% and 28% of observations, and 70% and 73% of aggregate dollar volume, respectively.

Table 1: **Descriptive statistics for hammer prices and price-to-estimate ratios**

This table shows descriptive statistics (mean, median (P50), first quartile (P25), and third quartile (P75)) for hammer prices in U.S. dollars and price-to-estimate ratios based on the lots that sold successfully. The first row shows statistics starting from the full data set, which covers auctions worldwide over the years 2008–2015. The next sets of rows show statistics for the top-three most frequently observed artists, auction houses, and auction locations separately.

			Hammer price ($)				*Price / estimate*			
	N	% sold	Mean	P25	P50	P75	Mean	P25	P50	P75
All	1,187,666	0.65	61,225	1,037	3,271	13,000	1.14	0.71	0.88	1.26
Pablo Picasso	2,380	0.72	1,198,106	4,750	11,166	170,000	1.23	0.79	1.00	1.40
Andy Warhol	2,351	0.76	1,299,384	14,000	105,877	478,632	1.09	0.76	0.91	1.26
Victor Vasarely	1,283	0.67	38,607	3,237	23,401	55,000	1.14	0.80	0.92	1.33
Christie's	110,764	0.77	217,080	4,467	15,480	60,700	1.24	0.75	0.94	1.40
Sotheby's	84,116	0.70	304,368	13,000	35,993	120,000	1.28	0.80	1.00	1.43
Bonhams	66,908	0.64	16,712	1,200	3,213	9,833	1.09	0.72	0.85	1.21
Paris	148,572	0.58	24,807	1,004	2,772	9,068	1.33	0.81	0.98	1.43
New York	105,357	0.70	270,879	3,500	14,000	60,000	1.12	0.67	0.87	1.29
London	74,041	0.66	279,438	11,076	31,775	109,528	1.21	0.79	0.93	1.36

Table 1 also includes some statistics on price-to-estimate ratios, which we compute by dividing the hammer price by the mean—or "mid"—pre-sale estimate (and subsequently winsorizing at 0.10 and 10).[7] We see that the median (mean) price-to-estimate ratio associated with successful sales is slightly below (above) one.

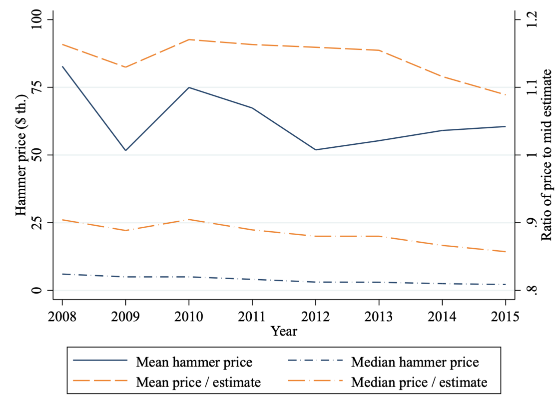
In Figure 2, we show time trends for some key statistics. Panel (a) shows the number of observations and sale rate (i.e., one minus the percentage of buy-ins) for each year of our sample period. Panel (b) shows the yearly mean and median hammer price and price-to-estimate ratio. Later on, we will use the data for the final year of our sample period to test the predictive power of different valuations. In this sense, it is reassuring that Figure 2 shows no dramatic changes in terms of sample composition for 2015.

[7] The size of the spread between the low and the high estimate does not show much variation once controlling for auction house and low estimate, and is thus unlikely to contain any relevant information about the auctioneer's confidence in her own estimate. For example, in our training data, 175 out of the 177 lots with a low estimate of $100,000 offered at Christie's in the U.S. have a high estimate of $150,000.

Figure 2: **Time-series variation in sample size and prices**

Panel (a) of this figure shows the yearly number of observations (against the left axis), and percentage of lots that sold successfully (against the right axis) in our database. Panel (b) shows the yearly mean and median hammer price (against the left axis), and mean and median ratio of hammer price to mean pre-sale estimate (against the right axis) based on lots that sold successfully.

(a) Number of observations and sale rates



(b) Hammer prices and price-to-estimate ratios

3 Valuing Art Using Machine Learning

3.1 Overview of Methodological Approach

In this section, we explain how we can generate alternative valuations that auctioneers *could have set* given the information that they had access to at the time of deciding on their pre-sale estimates. We use works auctioned over the years 2008–2014 to develop an algorithm that predicts the price of any artwork based on its characteristics. In the next section, we will then test the performance of our algorithm using auction data from the year 2015. (We choose our out-of-sample test set to follow our training sample period to avoid that we use information from *after* a sale to predict its outcome.)

The machine-learning technique that we employ is neural networks, which can be seen as very large parametric models. Neural networks consist of different interconnected "layers" of nodes (or "neurons"), where the first one is called the "input layer" and represents the variables from which predictions are made, and the final one is called the "output layer" and contains predictions compiled from the inputs. In between input and output layers, there are other layers that apply linear and non-linear "activation functions" to the nodes in the previous layer so as to extract the information that is most helpful for making predictions. Neural networks' parameters—also called "weights"—are typically learned from observations in an iterative and stochastic manner.

Given that our data set contains not just textual information on artworks' characteristics but also their images, we use a specific type of neural networks that is popular in image-recognition tasks, namely "convolutional neural networks" (CNNs). Such networks have the capacity to learn very complex functions of images' pixel values, while taking advantage of the spatial structure of an image in which nearby pixels are correlated. Given sufficient training data, CNNs are able to "predict" very reliably an artwork's genre, creator, and semantic content, as well as human aesthetic judgments (Karayev *et al.*, 2014; Tan *et al.*, 2016; Strezoski and Worring, 2017). In theory our algorithm can thus pick up any relation between artwork subject and composition (shape, color, etc.) on the one hand and prices on the other hand.

The following subsections detail the input variables that enter the algorithm, the architecture of the neural network, and how we estimate the network.

3.2 Input Variables

Next to the image, we derive the following explanatory variables from the textual information in the database:

i. **Artist**. The data that we will use for training our neural network (i.e., auctions over the period 2008–2014) includes lots by 117,000 different artists.

ii. **Artist nationality.** Together, these artists represent almost 170 different nationalities.

iii. **Artist birth and death year.** For more than 90% of all lots, the database includes information on the birth year of the artist. In the training data, the median birth year is 1897. If the artist has already died at the time of the auction, we typically also have information on the year of death.

iv. **Artist style.** The database classifies about two thirds of all works in one of the following style categories: (1) Old Masters; (2) 19th Century European; (3) Impressionist and Modern; (4) Post-War and Contemporary; (5) American; (6) Latin American; (7) Asian.

v. **Artwork creation year**. We have precise information on the creation year for about half of all observations. A large majority of the works for which we have this information date from the twentieth century.

vi. **Artwork width and height**. Artwork size is included in the database for nearly all observations. We winsorize width and height at 10 and 200 centimeters. The median width in the training data is 55 centimeters, while the median height is 52 centimeters.

vii. **Artwork materials**. We create 18 indicator variables for the following terms that appear frequently in the description of the materials and support: (1) oil; (2) watercolor; (3) acrylic; (4) ink; (5) gouache; (6) bronze; (7) mixed media; (8) pastel; (9) lithograph; (10) poster; (11) etching; (12) pencil; (13) canvas; (14) board; (15) panel; (16) paper; (17) masonite; (18) wood. These categories are not mutually exclusive. In the training data, only 2.8% of all lots fall outside of any of these categories. For more than 75% of all lots, exactly two dummies equal one (as would be the case, for example, if the description reads "oil on canvas").

viii. **Artwork title**. We create eight indicator variables for the following groups of terms that are used frequently in artwork titles: (1) untitled, sans titre, senza titolo, ohne titel, sin titulo, o.t.; (2) composition, abstract, composizione, komposition; (3) landscape, paysage, paesaggio, seascape, marine, paisaje; (4) still life, flowers, nature morte, bouquet de fleurs, nature morta, vase de fleurs; (5) figure, figura, character; (6) nude; (7) portrait, mother and child; (8) self-portrait, self portrait. (To come up with this classification, we consider the 50 most frequent titles in our sample, and manually create groups of related words.) These categories are not mutually exclusive. In the training data, at least one of these indicator variables equals one for 21.6% of all works.

ix. **Artwork markings**. We create three dummy variables that equal one if the artwork is (1) signed; (2) dated; or (3) inscribed by the artist. These categories are not mutually exclusive. In the training data, at least one of these indicator variables equals one for 82.0% of all works.

x. **Auction house.** The biggest auction houses are clearly Christie's and Sotheby's, but the training data set covers lots at nearly 370 auction houses in total (some of which have different locations).

xi. **Auction location.** The database specifies the location (typically, a city) for each auction. The training data includes sales in 230 different locations.

xii. **Auction month.** We create a variable capturing the month of the sale.

xiii. **Auction year.** The algorithm will also have access to the year of sale. When generating out-of-sample estimates of market values for the test set (i.e., 2015), it will do so as if these observations are from the final year of the training set (i.e., 2014). In principle it can put more weight on more recent observations.

Appendix Table A.1 gives more statistics regarding the different input variables described above.

3.3 Network Architecture

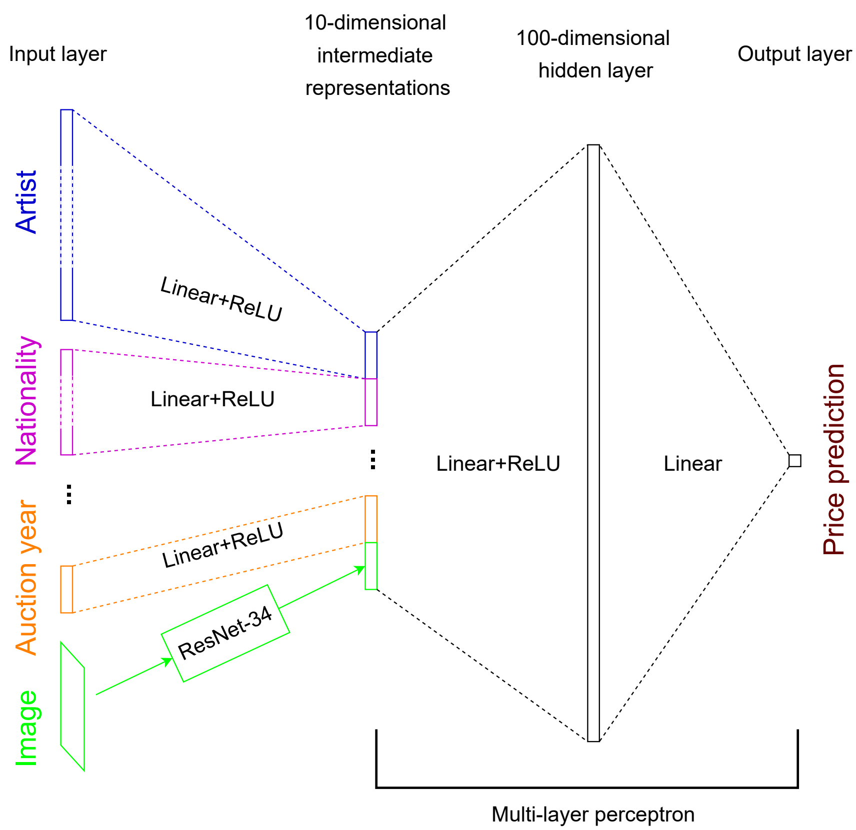
The architecture of our neural network is visualized in Figure 3. In the input layer, each non-visual input variable is represented by a vector of dummies, which is typically referred to as "one-hot encoding" in machine learning. So we have separate indicator variables for each artist, for each artist nationality, etc.[8] For each input variable category (cf. items i–xiii above), we then project this initial representation onto a 10-dimensional vector. This projection is done using a combination of a linear operation (a fully-connected layer whose parameters are learned during training) and a non-linear activation function called a "rectified linear unit" (ReLU).[9] There are different reasons for introducing this 10-dimensional bottleneck. First, it reduces the number of network parameters. Second, it avoids overfitting related to the high dimensionality of some of the initial dummy variable representations (e.g., artist, auction house). The 10-dimensional projection can be seen as

[8]The only exception is artwork size, which is simply represented as a two-dimensional vector capturing width and height. To avoid overfitting, for artist birth year, artist death year, and artwork creation year, we create a separate indicator variable for each year between 1800 and 2003, plus separate dummy variables for all years before 1800, for all years after 2003, and for missing values.

[9]ReLUs are the most commonly used non-linear functions in modern neural networks. ReLUs associate to a real number its positive part: $ReLU(x) = max(0, x)$. Each ReLU is preceded by a normalization to speed up and improve training.

encouraging the network to consider in a similar way many different artists or auction houses. Third, associating 10 dimensions to each type of input variable, including low-dimensional ones (e.g., the variable category measuring markings, which only has three different dummies), ensures that our algorithm is not biased to attaching higher importance to variables that are initially associated with more dimensions.

Figure 3: **Graphical representation of our neural network**



We also represent the images with a 10-dimensional vector, which is computed using a so-called "ResNet". ResNets are one of the most standard CNN architectures for images (He *et al.*, 2015). We here use a network type that is known to have a good performance for image classification, while being small enough to be trained in a reasonable time.

The intermediary representations of the non-visual input variables and of the image are appended into a 140-dimensional (14×10 dimensions) descriptor of each artwork. This vector is then used as input to a "multi-layer perceptron" (MLP) with one "hidden layer" (with 100 nodes). This MLP is the function that makes the actual price predictions by applying linear and non-linear operations to the 10-dimensional intermediate representations of the input variables.

The total number of parameters in our network architecture is of the order of one million. Most of them correspond to the conversion of the artist dummies into a 10-dimensional representation. Approximately 15,000 correspond to the MLP that generates price predictions starting from the intermediate representations. (Note that the different parts of the network are trained—and the weights are thus optimized—simultaneously.) Intuitively, our network is thus not prone to much overfitting, because—helped by the initial projections of all input variables into 10-dimensional representations—the number of parameters in the prediction function is much smaller than the number of observations in the training set.

3.4 Estimation

The data used for estimating the network correspond to all 965,062 included lots over the 2008–2014 period. A randomly sampled 1% of these observations is used as validation data. The performance of the network on this small subsample is used during development to decide on meta-parameters related to the network architecture (e.g., the dimensionality of the intermediate variable representations, the number of hidden layers and nodes in the MLP) and to the optimization.

We train the algorithms on hammer prices, but we also include buy-ins at an imputed price equal to 75% of the auction house's low estimate, motivated by our knowledge of average reserve-to-estimate ratios (e.g., McAndrew *et al.*, 2012). All prices are log-transformed. Moreover, as we want to focus on economically meaningful variation in art prices—and will perform our tests on data for auction houses and artists with a minimal level of recognition—we winsorize all prices at $1,000 prior to training.[10]

The weights of the neural network are optimized to minimize a loss in the training set. More precisely, given N observations (x_1, y_1), ..., (x_N, y_N) in the training set, where x_i denote the values of artwork i on the input variables and y_i denotes the logged sale price, we minimize the following

[10]Otherwise the algorithm would spend as much effort to try to differentiate between a $100 and a $200 transaction as between a $1 million and a $2 million sale. Moreover, in the lowest segment of the auction market (entirely outside of the main auction houses), price differences are arguably largely idiosyncratic and mainly driven by intermediary rather than artwork characteristics; purchased artworks will have little resale value. We also winsorize a handful of prices at $50 million.

loss function capturing squared prediction errors:

$$L(w) = \sum_{i=1}^{N} (f_w(x_i) - y_i)^2, \tag{1}$$

where w are the parameters of our network and f_w is the function associated to our network with parameters w. This loss is minimized using the popular gradient-based optimizer of Kingma and Ba (2017).

To regularize our training, we put each variable equal to zero with a probability of 0.2. This process of randomly zeroing out a subset of the input variables during training is similar in spirit to dropout procedures that are more typically applied to intermediary nodes in the network to avoid overfitting (Srivastava *et al.*, 2014). Our approach trains the network so that it can make predictions even if some information about the artist or artwork is missing. It also enables us to study the relative impact of the different variables on predictions, and to see how removing certain information changes the predictive power of the network.

3.5 Illustration

One of the caveats of using deep-learning networks such as ours is that it can be hard to understand what exactly the intermediate representations are learning. In our relatively parsimonious architecture, this is especially true for the 10-dimensional vector of image features that the network generates from the original pixel values. We therefore visualize in Figure 4 a set of artworks (shown in the first column) with their "nearest neighbors" (shown in the subsequent columns) based on the image features. More precisely, we identify for ten images (by ten different high-volume artists) the nine other images in the training data with the highest (cosine) similarities. The figure shows that the way in which the network considers different artworks as being "similar" is relatively complex: it includes elements of both semantic content and style.[11]

[11]Note that using a different output variable than price (e.g., human judgment) would lead to different "nearest neighbors", because the parameters for the projection of each image into a 10-dimensional vector would be different.

Figure 4: **“Nearest neighbors” (based on image features) of selection of artworks**

The first column of this matrix of artwork images shows ten randomly chosen artworks by ten different high-volume artists in our training data. The nine other images on each row then show the artworks that our network identifies as the first artwork’s “nearest neighbors” in terms of image features.

4 Assessing Machine-Learning Valuations

4.1 Data Filters

In this section, we verify how well the price predictions generated by our neural network line up with actual transaction prices out-of-sample. We also compare the predictive power of our baseline model to a number of alternative estimates and algorithms. We do so using data for (a subset of all) artworks auctioned in 2015, which the network did not "see" while learning. We impose two data filters. First, we drop a very small fraction (0.5%) of sales where the hammer price is below 10% of the low estimate or above ten times the high estimate. Some of these outliers may be cases where either the price or the estimate is incorrectly recorded in the database, or some (to us) unobservable event happened between estimation and auction of the artwork (e.g., a re-attribution). Second, in order to ensure that our analysis is focusing on economically meaningful objects and trading places, we focus on artworks by artists and auction houses that are associated with an average mid estimate over the training period of at least $5,000 and $10,000 respectively. This filter reduces the number of observations in our test set by about two thirds, but in terms of aggregated dollar sales the filtered-out lots represent less than 5% of the initial sample. The final data set that will be used for the tests in this section still contains nearly 60,000 auctions at 81 different auction houses in 78 locations on five different continents, and works by more than 12,000 different artists.

4.2 Machine-Learning Valuations and Prices

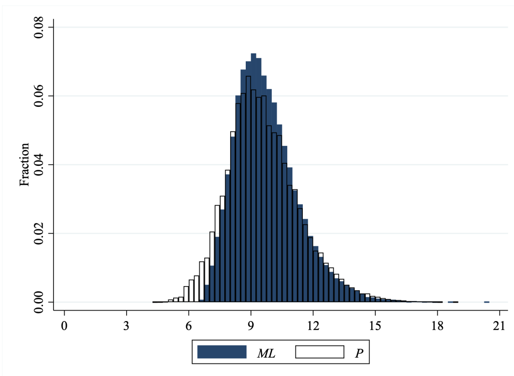
Let us denote by P the (log) realized hammer prices for the objects in our test set. If a work is bought in, we impute a value of 75% of the low estimate.[12] ML denotes the out-of-sample (log) price predictions generated by the neural network for each year-2015 auction. In panel (a) of Figure 5, we compare the distributions of these two variables to each other for all observations in our test set. The machine-learning predictions exhibit less dispersion in the left tail, which can be explained by the winsorization applied during the training of the neural network.

Panel (b) of Figure 5 shows the distribution of prediction errors $P - ML$ (winsorized at -4 and

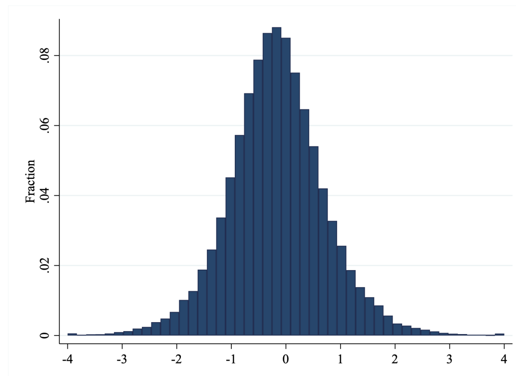
[12] The results described in this section are qualitatively and quantitatively similar when only using successful sales.

Figure 5: **Machine-learning valuations and prices in test data**

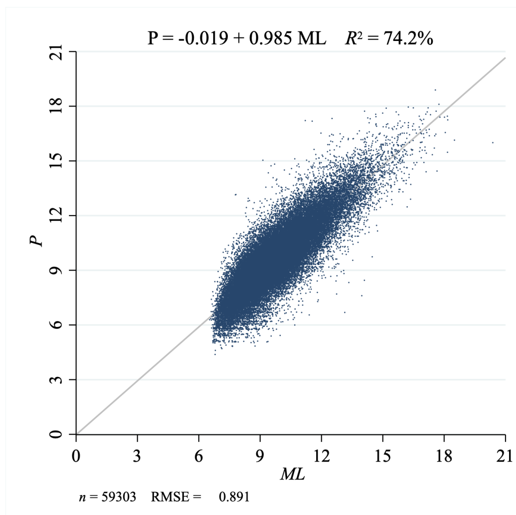
Panel (a) of this figure compares the distribution of machine-learning predictions ML to that of realized transaction prices P for our test set. Panel (b) shows the distribution of prediction errors $P - ML$, winsorized at -4 and +4. Panel (c) shows the results of a linear regression of P on ML. The line shows the predicted linear fit.



(a) Distribution of ML and P



(b) Distribution of prediction errors $P - ML$



(c) Relation between ML and P

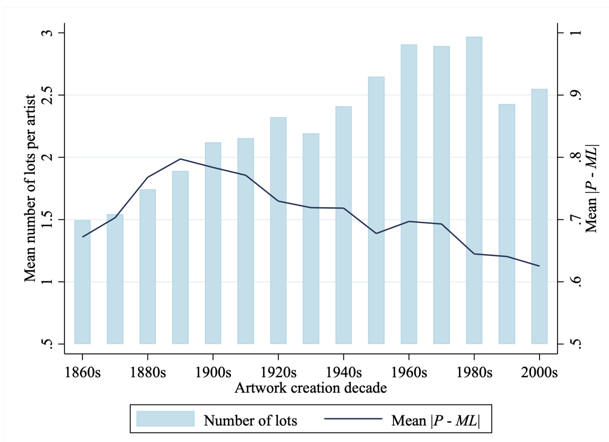
+4). We see a rather nicely-behaved bell curve with a mean and median just below zero. The interquartile range goes from -0.70 to 0.36, meaning that half of all prediction errors fall in this interval that corresponds to deviations of prices from our machine-learning predictions of -50% to $+43\%$. The mean (median) *absolute* prediction error $|P - ML|$ equals 0.69 (0.55).

The scatter plot in panel (c) shows how auction prices P line up with our machine-learning valuations ML. We also show the results of a linear regression model, and the predicted linear fit in the plot. We see that the regression line is almost exactly 45 degrees: the slope coefficient is almost perfectly one, while the intercept is very close to zero. The R-squared shows that our automated valuations explain about three quarters of the (out-of-sample) variation in auction prices.

In Figure 6, we plot the mean absolute prediction error in the test set by artwork creation decade (conditional on the artwork creation year being available). At least for modern and contemporary art, the prediction errors made by the algorithm decrease over time; the most recent artworks appear to be the easiest to price in an automated fashion. In the same figure, we also plot the average number of sales per artist for each creation decade, as a higher number of sales for a particular type of artworks may improve the performance of our network. The relation between availability of data and model performance does not seem to entirely explain the lower prediction errors for more recent art. (We will formally examine the importance of artist-level volume and artwork age in the relative contribution of machine learning in the next section.)

Figure 6: **Performance by artwork creation period**

This figure shows the average number of lots per artist in our test set (against the left axis) and the mean absolute prediction error $|P - ML|$ (against the right axis) for each artwork creation decade since the 1860s.



4.3 Assessing Individual Variable Importance

To open up the black box of price formation—or at least price *predictability*—in the art market, we show in Table 2 the predictive power of a number of variations on our benchmark model. These alternative models generate new predictions for all objects in the test data set after removing their values on one or more of the input variable categories that were introduced in Section 3. We then recompute the R-squared based on the (updated) machine-learning valuations and the (unchanged) price outcomes. Our approach is a valid one to study the predictive power of models with less inputs, because we explicitly trained the algorithm to be able to make predictions even in the presence of missing information.

Table 2: **Assessing individual variable importance**

This table reports the R-squareds of linear regressions of logged transaction prices (P) in the test data set against different predictions generated by our neural network. The first column shows results for models that include the artwork image (ML), while the second model shows results for models that do not include the artwork image (ML_{txt}). The first row shows results for models that include all predictive variables, while the next rows show results for models that exclude different sets of variable types. The roman numbers between square brackets refer to the variable list in Section 3.

	$R^2_{ML,P}$	$R^2_{ML_{txt},P}$
Benchmark model	74.2%	71.8%
Without auction-related info [x-xii]	67.5%	64.0%
Without auction-related info [x-xii] + artist identifiers [i]	43.8%	38.7%
Without auction-related info [x-xii] + artist/style info [ii-iv]	64.3%	60.0%
Without auction-related info [x-xii] + artwork year [v]	66.7%	63.0%
Without auction-related info [x-xii] + artist identifiers [i] + artist info [ii-iv] + artwork year [v]	18.6%	13.8%
Without auction-related info [x-xii] + artwork size [vi]	60.1%	54.5%
Without auction-related info [x-xii] + artwork materials [vii]	61.2%	54.6%
Without auction-related info [x-xii] + artwork title [viii]	67.2%	63.5%
Without auction-related info [x-xii] + artwork markings [ix]	67.2%	63.8%
Without auction-related info [x-xii] + artwork characteristics [vi-ix]	48.8%	37.0%

As a starting point, the entry in the first row and the first column of Table 2 repeats the R-squared for our benchmark model, namely 74.2%. In the rows below, we show how this R-squared changes if we drop certain sets of non-visual inputs. We first drop the auction-related info (auction house, auction location, auction month), which pick up the price predictability created by the endogenous matching of artworks to auctions. By ignoring the data coming out of this "selection" stage, we show how much predictive power our model would have in a setting where it is unknown where each object will (or can) be sold. In order to focus on how such ex-ante predictability is affected by asset-specific features, also the next rows ignore the information on where a work is auctioned, and further eliminate different sets of artist-related or artwork-related characteristics.

We see that dropping the artist dummies has a substantial effect on the predictive power of our model, although the R-squared is still only reduced by about one third as long as the network has information on artist nationality and period. If *all* information related to the artist and the creation period is removed, the R-squared drops to 18.6% only. We further see that artwork size and materials matter much more than (our proxy for) the title and the presence of a signature or other markings. Removing information on all "physical" artwork characteristics (size, materials, title, markings) has less of a negative effect on predictive power than simply removing the artist identifiers, at least in this set of variations where the network can still rely on the artwork images.

We then repeat all models without images, leading to predictions that we denote by ML_{txt}. The resulting R-squareds are shown in the second column of Table 2. One can conclude from the first row that the incremental explanatory power of images is relatively limited. The R-squared is only 2.4 percentage points lower without images. So either visual characteristics are not very important in driving prices (once controlling for artist identity, size, materials, and so on), or economic value *is* associated with certain distinctive image characteristics but machine learning is ineffective in identifying such relations. At least two arguments can be made in favor of the first interpretation. First, as we illustrated in Figure 4, the algorithm appears capable of identifying visual similarities between different artworks. Second, we see in the last row of Table 2 that the relative difference in predictive power between models with vs. models without images becomes much larger once we remove non-visual artwork descriptors. This suggests that the network *is* able to pick up meaningful relations between artwork characteristics and prices based on the images.

4.4 Comparison with Linear ("Hedonic") Regression Model

We now generate an alternative type of valuation that can be considered "automated" as well—and therefore serve as a useful benchmark—but relies on a more traditional and less sophisticated method. Following Rosen (1974) and real estate scholars, academics studying the art market have linked prices to artwork characteristics, typically employing linear regression models (e.g., Anderson, 1974; Renneboog and Spaenjers, 2013). We estimate a standard hedonic model on the training set, and use the regression coefficients to generate out-of-sample hedonic valuations for all artworks in the test set.[13] More specifically, we estimate the following model using ordinary least

[13] Given the many categorical input variables, shrinkage methods like lasso and ridge are not very practical in our empirical context.

squares on the observations in the training set:

$$y_i = \alpha + X_i'\beta + T + \varepsilon_i, \tag{2}$$

where y_i is the log-transformed price associated with auction i, X_i is a vector of hedonic variables, and T are auction year fixed effects. We can use the following earlier-introduced variables in our hedonic model: artist fixed effects, artwork height and width (and their squares), and the artwork material, title, and marking dummies, and auction house, location and month dummies.[14] Appendix Table B.1 shows the hedonic regression coefficients. The results are generally in line with findings in the existing literature. For example, substantially higher prices are paid for works that are bigger, signed or dated, self-portraits, and created with oil. We can then use the estimated coefficients to generate out-of-sample price predictions HR for all lots without missing values on any of the variables included in the hedonic regression model. In line with what we did before, we make predictions as if the out-of-sample observations are from the year 2014 by using the coefficient on the fixed effect for that year.

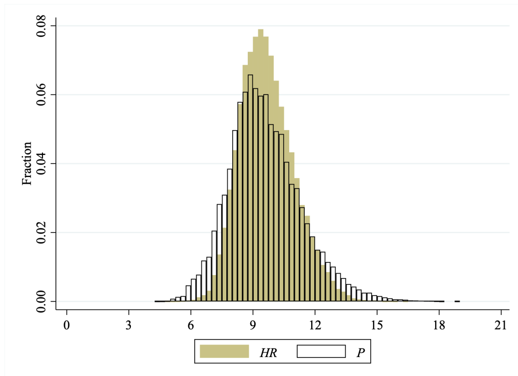
In Figure 7, we show the distribution of hedonic valuations, the distribution of prediction errors, and the relation between hedonic valuations and transaction prices, mirroring the different panels of Figure 5, which used our machine-learning valuations. In panel (a), we can observe that the distribution of hedonic valuations is more concentrated than that of prices—and also than that of machine-learning valuations. In panel (c), we see that the R-squared is substantially lower than before: 67.6% compared to 74.2%. The predicted linear fit is also further away from a 45-degree benchmark.

A major conceptual difference between a machine-learning approach and a hedonic model is that the latter does not exploit interaction effects between different variables. The upshot is that, once conditioning on some important determinants of variation in price levels, the hedonic valuations will show little dispersion. We illustrate this in Figure 8, which shows a number of

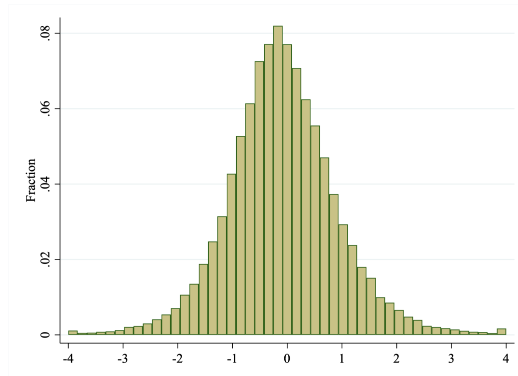
[14]Other artist-level variables would be dropped during estimation because of the artist dummies. We do not include artwork creation year as a separate variable, because, first, artwork creation year can safely be assumed to be relevant only in the context of a specific artist's career, and, second, the model would then not be able to make a prediction if the creation year is missing.

Figure 7: **Hedonic valuations and prices in test data**

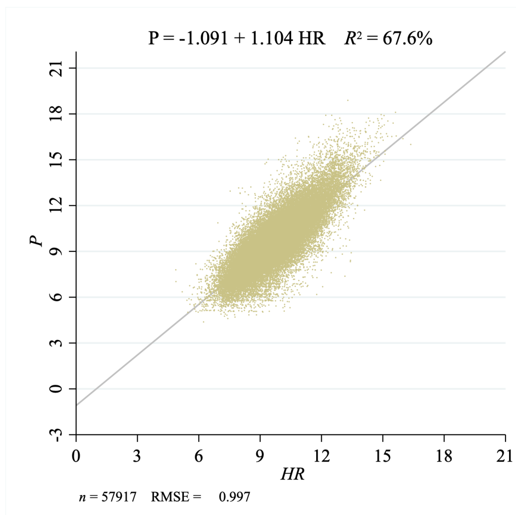
Panel (a) of this figure compares the distribution of hedonic valuations HR to that of realized transaction prices P for our test set. Panel (b) shows the distribution of prediction errors $P - HR$, winsorized at -4 and +4. Panel (c) shows the results of a linear regression of P on HR. The line shows the predicted linear fit.



(a) Distribution of HR and P



(b) Distribution of prediction errors $P - HR$



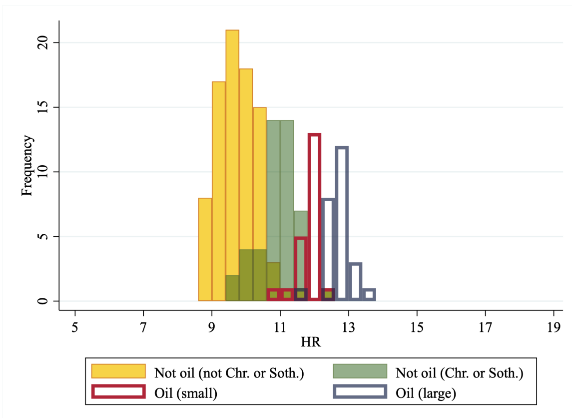
(c) Relation between HR and P

histograms for the works by Pablo Picasso in our test set. Panel (a) shows histograms for hedonic valuations HR, panel (b) for our earlier-generated neural network price predictions ML, and panel (c) for price outcomes P. Each panel includes two distributions of predictions or prices for oil paintings (based on whether the artwork is relatively large or small), and two distribution for non-oil artworks (based on whether the lot was auctioned at one of the two major auction houses or not).

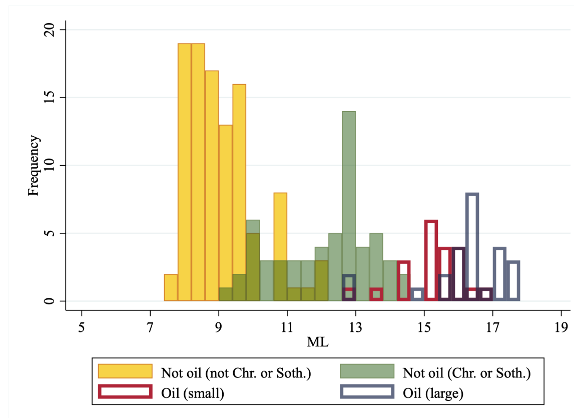
Figure 8 shows that prices are affected by the materials and size of the artwork, and correlate with the identity of the auction house—and that these patterns are reflected in both HR and

Figure 8: **Predictions for Pablo Picasso**

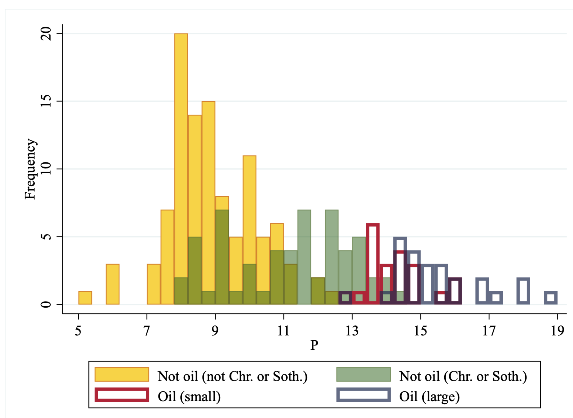
Panel (a) of this figure shows four distributions of HR for the works by Pablo Picasso in our test set: oil paintings with a width of at least 50cm ("large"); oil paintings with a width of less than 50cm ("small"); non-oil artworks sold at Christie's or Sotheby's; and non-oil artworks sold elsewhere. Panels (b) and (c) show the same histograms for ML and P, respectively.



(a) Histograms for HR



(b) Histograms for ML



(c) Histograms for P

ML. We also see, however, that the dispersion that exists in prices for Picasso works is not all reflected in the hedonic valuations. Each of the four categories of works considered are associated with a narrow distribution of HR. This is not surprising given the additive and linear structure of a standard hedonic model; for example, all small Picasso oil paintings will have relatively similar hedonic valuations, mainly driven by the (aggregation of the) coefficients on the artist, materials, and size variables in the model. Our machine-learning valuations ML show much more variation, even within each subsample.[15] While it is difficult to speculate on what exactly

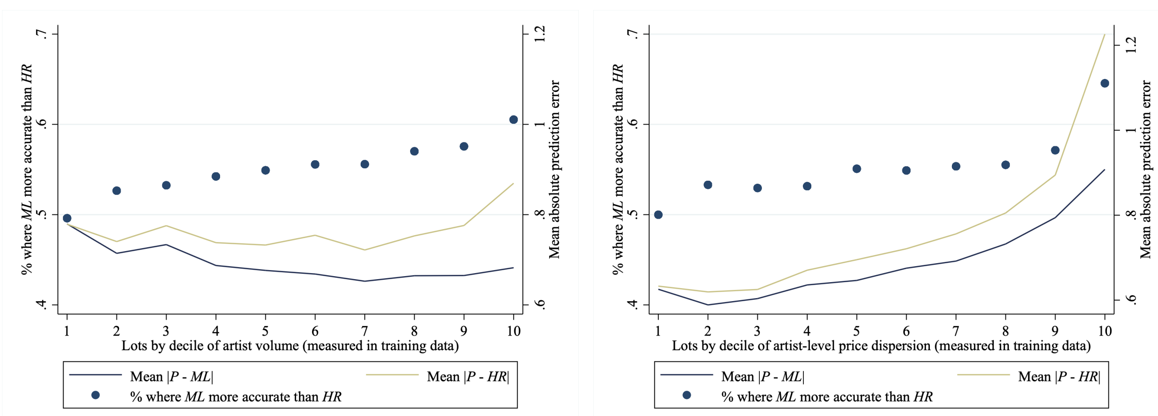
[15]The Picasso artwork associated with both the second-highest HR and ML (and the highest ML_{txt}) in this analysis is "Les Femmes d'Alger (Version 'O')", a large oil painting that went on to become the most expensive artwork ever to sell at auction (at the time) by selling for a hammer price of $160 million in May 2015.

drives this variation, it is necessarily related to (potentially artist-specific) interactions between different artwork characteristics—including visual ones—that the neural network has discovered to be predictive of prices.

Given the above, we might expect that machine-learning is particularly useful—relative to standard statistical tools—for artists that are associated with a large and varied *oeuvre*. We address this hypothesis in Figure 9. We first classify all lots in our test data set in ten deciles according to the associated artist's number of auction lots in the training data. We then compute the average absolute prediction errors based on both ML and HR for each decile, and also compute what is the fraction of observations for which ML is closer to P than HR. The results are shown in panel (a). Panel (b) then repeats the exercise but using deciles based on the artist-level standard deviation of (logged) transaction prices in the training data, as a proxy for the variation in the output of an artist.

Figure 9: **Performance of machine-learning algorithm vs. hedonic model**

Panel (a) of this figure shows the fraction of lots for which ML is a more accurate prediction of P than HR (against the left axis), and the mean absolute prediction errors $|P - ML|$ and $|P - HR|$ (against the right axis) for deciles of lots in the test set sorted by the number of auctions of works by the artist in the training data. Panel (b) repeats the exercise for deciles of lots in the test set sorted by the standard deviation of prices for the artist in the training data.



(a) ML vs. HR by artist volume decile

(b) ML vs. HR by artist price dispersion decile

We see that our machine-learning valuations are indeed more likely to be more accurate than hedonic valuations for works by high-volume and high-dispersion artists. Such artists also tend to be more expensive on average, meaning that machine learning will contribute more for the economically more important lots.[16] Panel (a) of Figure 9 additionally illustrates that prediction errors for *ML* are lower—even in absolute terms—for artists with more transactions historically. Panel (b) shows that our neural network struggles to accurately price artists with a high standard deviation of prices, even if it does better than a hedonic model.

5 Testing the Informational Efficiency of Estimates

5.1 Pre-Sale Estimates and Prices

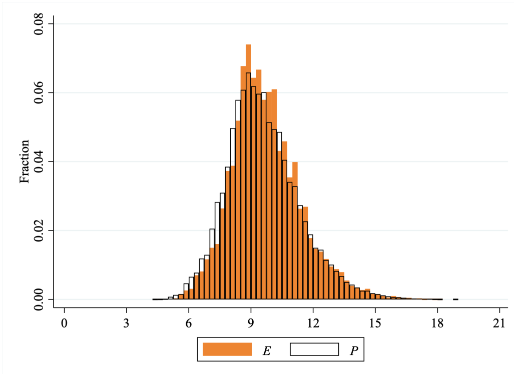
In this section, we will analyze how helpful our machine-learning valuations are in the presence of pre-sale valuations issued by the auction house organizing the sale. We denote by E the (logged) mean pre-sale estimate. Similar to Figures 5 and 7, Figure 10 shows the distribution of E and of prediction errors $P - E$, and also the results of a simple linear regression of P on E. What is striking in panel (a) is the discontinuities in the distribution of E, which is due to the fact that auction houses use standard estimate intervals (e.g., 1-2 million, 1.5-2.5 million, etc.). The distribution of prediction errors in panel (b) is of course reflecting the presence of buy-ins, for which we impute prices at 75% of the low estimate.

In panel (c) of Figure 10, we can observe that auction house estimates explain substantially more of the variation in hammer prices than our machine-learning algorithm. We want to stress, however, that an ex-post comparison of predictive power between auction house estimates and our automated valuations should not be construed as a horse race between "man" and "machine". On the one hand, as explained above, an estimate cannot at face value be considered as the auction house's truthful expectation of what the item will sell for. On the other hand, two factors will artificially drive up the relative "performance" of human-generated estimates. First, auctioneers take into account artwork-level (e.g., condition, provenance) and artist-level information (e.g., art-historical reputation) that is not observed by our algorithms, even when it could in principle be quantified

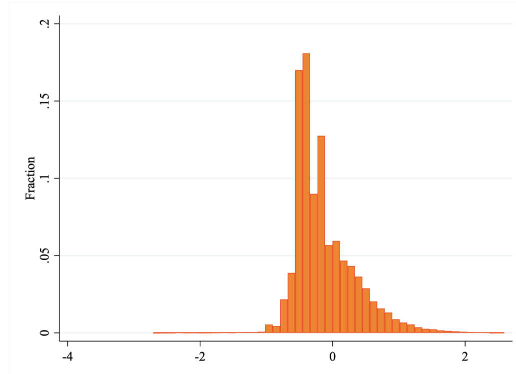
[16] If we sort lots by artist-level average price (measured in the training data), we get a pattern that is qualitatively similar to that shown in panel (b) of Figure 9. However, in a multivariate regression with a dummy variable that equals one if *ML* is more accurate than *HR* as the dependent variable, and variables measuring the different artist-level variables as independent variables, the artist-level average price is not statistically significant at any traditional level, while artist volume and price dispersion are highly significantly positive.

Figure 10: **Pre-sale estimates and prices in test data**

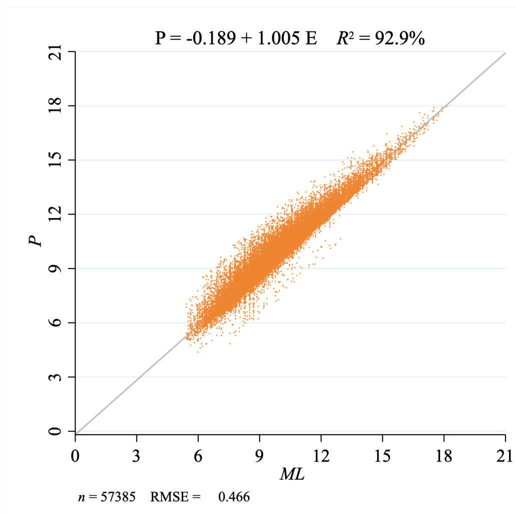
Panel (a) of this figure compares the distribution of auction house pre-sale estimates E to that of realized transaction prices P for our test set. Panel (b) shows the distribution of prediction errors $P - E$. Panel (c) shows the results of a linear regression of P on E. The line shows the predicted linear fit.



(a) Distribution of E and P



(b) Distribution of prediction errors $P - E$



(c) Relation between E and P

and "fed" to the machine. Second, auction houses' estimates will be endogenously correlated with prices if bidders anchor on those estimates to form beliefs about future resale revenues are affected by auction house estimates. (We will explore the implications of this possibility in our discussion section.)

5.2 Pre-Sale Estimates, Automated Valuations, and Auction Outcomes

We have established that, as we expected, pre-sale estimates turn out to be more highly correlated with prices than the predictions generated by our neural network. However, our primary goal is to analyze whether, *conditional on pre-sale estimates*, machine learning can help predicting auction outcomes. If auctioneers set informationally efficient estimates, this will not be the case. By contrast, if auction house estimates are affected by behavioral or strategic biases, or simply do not reflect all available information, then we should see that the relative level of ML predicts deviations of transaction prices from pre-sale estimates. In such case, we can also expect buy-in probabilities to be affected, as reserve prices are typically tightly linked to pre-sale estimates.

We start our analysis by running a regression of the prediction error $P - E$ (i.e., the logged price-to-estimate ratio) against E. We do so to check whether deviations of prices from pre-sale estimates are on average higher or lower for more valuable items. The results are shown in column 1 of Table 3. We see that realized price deviations from estimates are on average slightly higher for more expensive paintings, but the relation is economically insignificant. In column 2, we then add ML_{orth}, which is our machine-learning valuation ML orthogonalized with respect to E. (This orthogonalization neither affects the coefficient on the ML variable nor the R-squared, but allows to focus on the *additional* role played by our machine-learning valuations.) We see that our automated valuation substantially increases the R-squared of the regression model. Higher machine-learning valuations are associated with economically and statistically significantly higher price-to-estimate ratios.

To illustrate the economic significance of the results in column 2 of Table 3, we can show them visually. We sort all observations in the test data set on ML_{orth} and group them together in half-deciles. We compute the mean ML_{orth} for each of these twenty groups. The line in panel (a) of Figure 11 then shows the (mean) prediction error $P - E$ (i.e., the logged price-to-estimate ratio) as a function of the (mean) ML_{orth} for each half-decile. We see price deviations from pre-sale estimates are on average much higher for higher relative machine-learning valuations. The same panel also shows for each group what is the fraction of observations for which ML is closer to P than E. Not surprisingly, we see a hump-shaped scatter plot, where more extreme ML_{orth} values are less likely to be good predictors.[17]

[17] We focus on cross-sectional variation in this measure, rather than the level, as the latter is influenced by our definition of E. Namely, we work here with the mean of the low and the high pre-sale estimate, while other choices are of course imaginable.

Table 3: **Informational efficiency of pre-sale estimates**

Columns 1 and 2 of this table report estimated ordinary least squares coefficients for a linear regression model that has the auctioneer's prediction error (i.e., $P - E$) as the dependent variable. Columns 3–6 report estimated ordinary least squares and probit coefficients for regression models where the dependent variable is a dummy variable that equals one if a lot is "bought in" (i.e., if the highest bid remains below the reserve price). The models are estimated using the transactions in our test data set. Standard errors, which are two-way clustered at the artist and auction month level (except in columns 5 and 6 where clustering is only at the artist level), are reported in parentheses. The asterisks *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

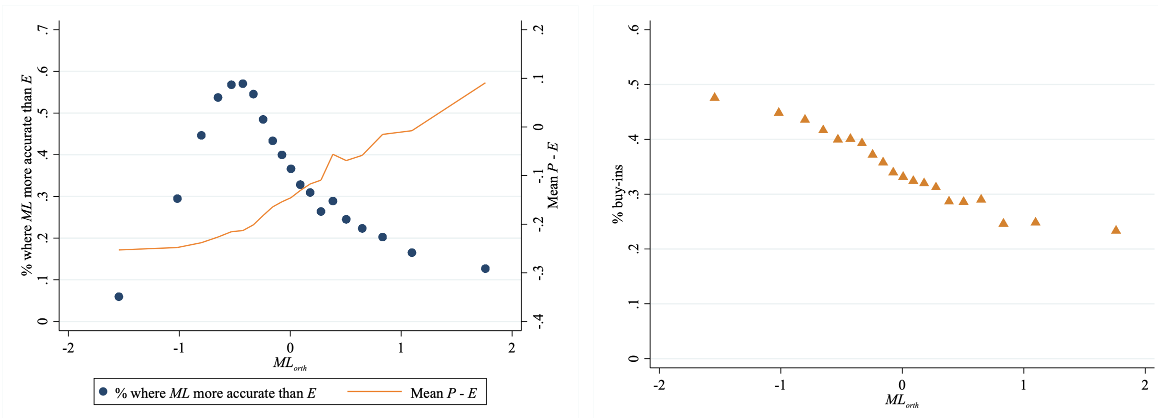
	(1)	(2)	(3)	(4)	(5)	(6)
	P - E		*Dummy = 1 if buy-in*			
	OLS		OLS		Probit	
ML_{orth}		0.116^{***}		-0.086^{***}		-0.242^{***}
		(0.009)		(0.007)		(0.010)
E	0.005^{*}	0.005^{*}	-0.016^{***}	-0.016^{***}	-0.044^{***}	-0.046^{***}
	(0.003)	(0.003)	(0.004)	(0.003)	(0.005)	(0.005)
Constant	-0.189^{***}	-0.189^{***}	0.501^{***}	0.501^{***}	0.031	0.040
	(0.031)	(0.024)	(0.033)	(0.026)	(0.048)	(0.049)
N	57,385	57,385	57,385	57,385	57,385	57,385
R^2	0.000	0.036	0.003	0.022	0.003	0.018

So far, we have considered the relation between our predictions and prices. Yet, our finding that we can improve on the pre-sale estimate to predict price outcomes suggests that there might also be some predictability of whether a lot will be bought in. More specifically, if the estimate is set relatively high for a certain work, then the reserve price—decided jointly upon by auctioneer and consignor, but never above the auctioneer's low estimate—is also likely to be relatively high. We can thus expect to see more buy-ins if our automated valuations are low compared to the pre-sale estimates. We test this hypothesis in columns 3–6 of Table 3, which shows the results for OLS and probit regressions, estimated over all lots in the test data set, where the dependent variable is a dummy that equals one if the item was bought in. The negative coefficient on E in each model tells us that in general buy-ins are somewhat less frequent for more expensive art. More importantly, we see in columns 4 and 6 that machine-learning artwork valuations help predicting buy-ins, in line with our expectation.

To evaluate the economic significance of our results, we plot in panel (b) of Figure 11 the realized out-of-sample buy-in frequency as a function of our ML_{orth} valuations grouped by half-decile as before. The graph shows that the buy-in probability is more than 45% when ML is low relative to the

Figure 11: **Informational efficiency of pre-sale estimates**

Panel (a) of this figure shows the average auctioneer prediction error (i.e., $P - E$) over all lots in our test set as a function of the orthogonalized machine-learning valuations ML_{orth}, which are averaged by half-decile. It also shows for each group of lots the fraction for which ML is closer to P than E. Panel (b) shows average buy-in rates as a function of the orthogonalized machine-learning valuations.



(a) $P - E$ as function of ML_{orth}

(b) Buy-in rate as function of ML_{orth}

auction house estimate, while this frequency decreases to around 25% when ML is relatively high. So the discrepancy between our machine-learning valuations and auctioneers' value assessments has substantial predictive power for the probability of selling.

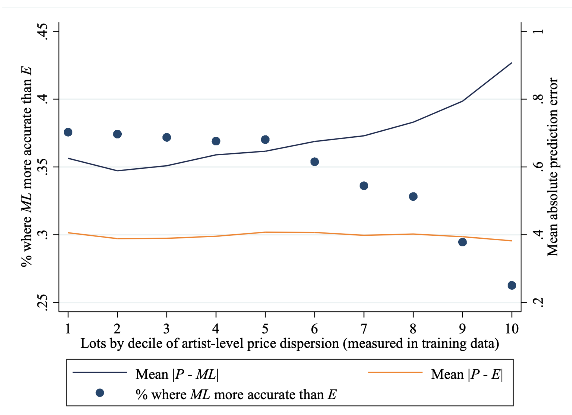
5.3 Variation in Added Value of Machine-Learning Valuations

The above analysis shows that a comparison to machine-learning valuations can help in assessing whether auctioneers' estimates are likely to under- or overshoot bidders' willingness-to-pay. However, it is likely that there exists heterogeneity in the added value of machine learning. In this subsection, we examine under which conditions we can expect our automated valuations to be more helpful.

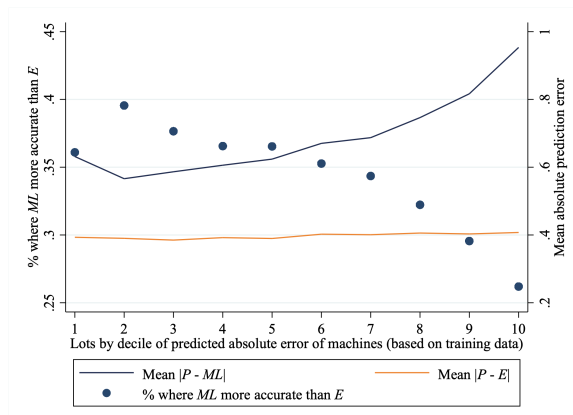
The *relative* contribution of automated valuations depends both on the accuracy of our neural network and on the prediction errors of auctioneers. So a first way to approach the issue at hand is to focus on heterogeneity in how well our automated valuations can be expected to predict prices. Our results in panel (b) of Figure 9 suggested that ex-ante artist-level price dispersion may be a good proxy for the complexity of the task faced by (both standard and more sophisticated) machine-based valuation methods. So in panel (a) of Figure 12, we show for each decile of lots in the test set—sorted by artist-level price dispersion in the training set—the mean absolute prediction error of our neural network (i.e., $|P - ML|$). Furthermore, we also show the average absolute prediction error of auctioneers (i.e., $|P - E|$) for each group of lots, and the proportion of lots for which ML is more accurate than E as a predictor of the price. As before, we see that the absolute prediction error of our machine-learning algorithm generally rises with artist-level price dispersion. Interestingly, however, $|P - E|$ does *not* rise accordingly, meaning that also the *relative* accuracy of ML decreases with our proxy for the range of possible prices associated with an artist.

Figure 12: **Drivers of performance of machine-learning algorithm vs. estimates**

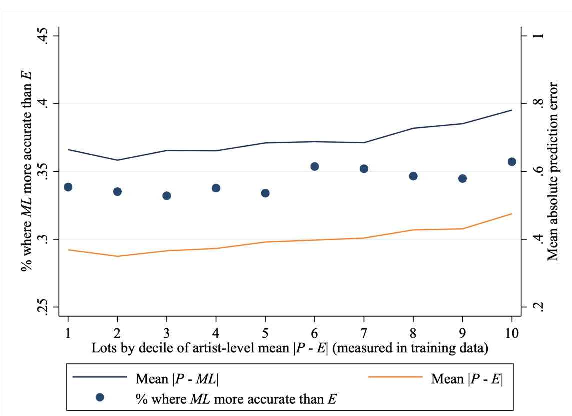
Panel (a) of this figure shows the fraction of lots for which ML is a more accurate prediction of P than E (against the left axis), and the mean absolute prediction errors $|P - ML|$ and $|P - E|$ (against the right axis) for deciles of lots in the test set sorted by the standard deviation of prices for the artist in the training data. Panel (b), (c), and (d) repeat the exercise for deciles of lots in the test set sorted by the predicted absolute error of machines, by the average $|P - E|$ for the artist in the training data, and by the average P for the artist in the training data, respectively.



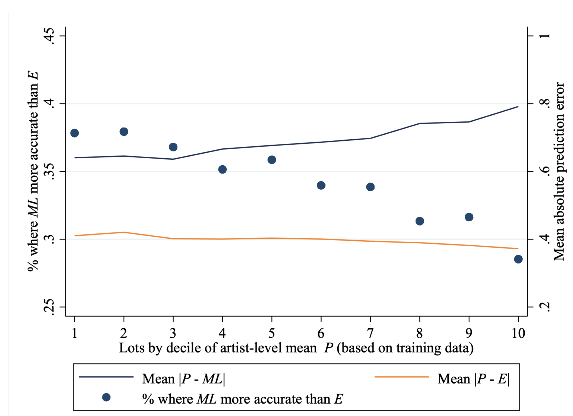
(a) By decile of artist-level price dispersion



(b) By decile of difficulty to value



(c) By decile of artist-level mean $|P - E|$



(d) By decile of artist-level mean P

We construct an alternative proxy for the difficulty of pricing lots in an automated fashion as follows. We consider our training data, and regress the (in-sample) absolute prediction errors of our previously-presented hedonic model back on all artwork- and artist-specific variables. This allows us to identify which artwork characteristics are associated with less accurate hedonic predictions on average. We then apply the regression coefficients to all observations in the test data set, which gives us an ex-ante proxy for each object of how difficult it will be for an automated model to come up with an accurate prediction. We then sort all lots in the test data into deciles based on this measure of the "difficulty for machines". This approach is in the spirit of Buchak *et al.* (2020), who aim to identify which kind of properties are hard vs. easy to value. The results are shown in panel (b) of Figure 12, and mirror those in panel (a).[18] The take-away is that there exists predictable heterogeneity in the size of the errors made by our neural network, but that this variation does not correlate with auctioneers' errors.

So far, we have attempted to differentiate lots based on whether we can expect them to be easy or hard to price by an automated algorithm. However, as we explained before, the *relative* contribution that we can expect from our machine-learning valuations also depends on how accurate we can expect auctioneers' estimates to be—assuming that systematic heterogeneity along this dimension exists. To examine this, we sort lots in the test set based on the associated artist's average $|P - E|$ in the training data.[19] So we are ranking artists by the accuracy of auctioneers during the years preceding 2015. The idea is that auction houses may find disentangling the different drivers of cross-sectional and temporal variation in prices persistently more challenging for certain artists than for others. The results are shown in panel (c) of Figure 12. We see that the mean absolute prediction error of the pre-sale estimates $|P - E|$ increases in our newly-built proxy for the expected noise in human valuations. However, also the machine-learning algorithm is associated with somewhat larger prediction errors when auctioneers' errors go up, and therefore the probability that ML is more accurate than E as a predictor of P only increases weakly with artists' mean $|P - E|$ in the

[18] The kink between deciles 1 and 2 can be explained by the fact that the in-sample error of the hedonic model—and thus the "predicted error"—will be very low for artists with only one or two lots in the training set, but lots by these artists will be associated with high prediction errors out-of-sample.

[19] We can also repeat the process outlined in the previous paragraph to predict auctioneers' errors, but this gives very similar results.

training data. One interpretation is that artists associated with higher absolute prediction errors of auctioneers are genuinely more difficult to value on average.

Finally, in panel (d), we sort lots in the test set by artist-level average prices, again measured in the training data. More information—outside of the auction data considered here—will be available to auctioneers for more expensive artists. Human prediction errors are indeed slightly smaller for more expensive artists. By contrast, *ML* becomes less accurate as the artist's average price goes up, which can be related to more valuable artists' wider dispersion of prices in the auction market (and potentially also the fact that hard-to-quantify factors like provenance and exhibition history may be more important for more established artists).

As a more formal analysis, we also run a probit regression in which the dependent variable is a dummy that equals one if *ML* is closer to *P* than *E*. In column 1 of Table 4, we include as independent variables the four different variables used to sort lots in Figure 12. The results are consistent with our earlier results and with expectations: *ML* is more likely to be more accurate than *E* for artists with less dispersed and lower average prices, artworks that can be predicted ex ante to be easier to value using a statistical function of artwork characteristics, and artists associated with higher absolute prediction errors by auctioneers historically.

In column 2 of Table 4, we add two additional variables that we can hypothesize based on our earlier results to be related to the relative performance of machine learning. First, we add a variable measuring for each artist the number of lots in the training data. Ceteris paribus, we expect that automated valuations are more accurate for more liquid artists, and that is indeed what we find. Second, we add a variable measuring the number of years since creation of the artwork. We saw before that machine-learning predictions tend be more accurate for more recent works. This result is confirmed here in a multivariate setting for the *relative accuracy* of machine learning.[20]

Finally, in columns 3 and 4 of Table 4, we repeat the models shown in columns 1 and 2, but adding auction house fixed effects. The magnitudes and statistical significance of the coefficients do not change much. So the drivers of relative machine-learning accuracy that we have identified seem

[20]When we do an analysis like the one shown in Figure 6 but with the mean $|P - E|$, we do not see much of a relation between creation period and auctioneers' prediction errors.

Table 4: **Determinants of added value of machine learning**

This table reports probit coefficients for regression models where the dependent variable is a dummy variable that equals one if ML is a more accurate predictor of P than E. The models are estimated using the transactions in our test data set, but the artist-level and artwork-level independent variables are constructed using the training data. Standard errors, which are clustered at the artist level, are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)
Artist-level price dispersion	-0.056^{**}	-0.083^{***}	-0.059^{***}	-0.091^{***}
	(0.023)	(0.029)	(0.023)	(0.030)
Artwork-level difficulty to value	-0.256^{***}	-0.317^{***}	-0.254^{***}	-0.310^{***}
	(0.037)	(0.050)	(0.037)	(0.051)
Artist-level mean $\|P-E\|$	0.249^{***}	0.289^{***}	0.230^{***}	0.274^{***}
	(0.063)	(0.077)	(0.068)	(0.079)
Artist-level mean P	-0.040^{***}	-0.043^{***}	-0.041^{***}	-0.045^{***}
	(0.007)	(0.008)	(0.007)	(0.009)
Artist-level # lots (logged)		0.037^{***}		0.038^{***}
		(0.007)		(0.007)
Years since artwork creation (logged)		-0.019^{**}		-0.021^{**}
		(0.008)		(0.009)
Auction house F.E.?	No	No	Yes	Yes
N	54,819	34,754	54,809	34,752
R^2	0.006	0.009	0.013	0.017

largely orthogonal to auction house identities. Yet, the R-squareds go up, which suggests a role for auctioneer effects in the magnitude of prediction errors, an issue that we will turn to next.

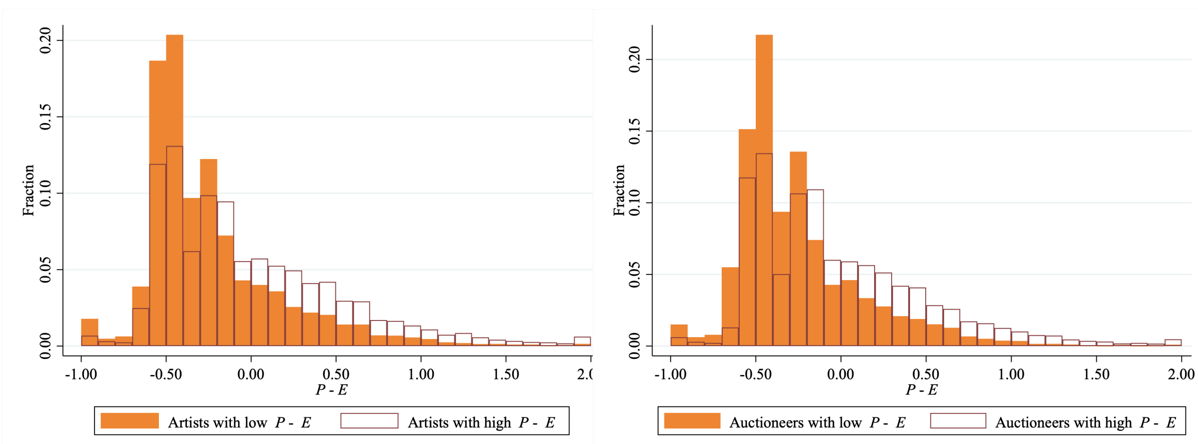
6 Predicting Auctioneers' Prediction Errors Directly

In the previous sections, we first generated automated art price predictions (i.e., ML), and then showed that the relative magnitudes of those machine-learning valuations help predicting the discrepancies between auction house pre-sale estimates and transaction prices (i.e., $P-E$). However, is it possible to *directly* predict auction houses' under- and overvaluations? If auctioneers are affected by biases that are systematic and persistent, then past prediction errors will be informative about future prediction errors. Such autocorrelation of prediction errors can then exist both at the level of auction house experts, and at the level of artworks that the auctioneers are likely to consider as substitutes.

To verify the plausibility of this hypothesis, we plot two histograms similar to that in Figure 1 with which we opened our paper. We first sort lots (in the test set) based on the mean $P - E$ (in the training data) of both the artist and the "auctioneer", with which we here mean a specific auction house–location combination (e.g., Christie's London). We then show the histograms for $P - E$ for the quartiles of lots associated with the lowest and highest average past $P - E$. Panel (a) shows the results for the sort at the artist level. Clearly, lots by artists that have been valued relatively low by auctioneers in recent years continue to get relatively low pre-sale estimates.[21] Panel (b) compares "low $P - E$" to "high $P - E$" auctioneers. Also here the persistence of prediction errors is very striking visually; auctioneers that issue relatively low estimates on average continue to do so.

Figure 13: **Persistence of prediction errors**

Panel (a) of this figure shows two different distributions of $P - E$, winsorized at -1 and $+2$, in the test set. We classify all lots in quartiles based on the artist-level average $P - E$ in the training set. We then compare the distribution for the first quartile ("Artists with low mean $P - E$") to the fourth quartile ("Artists with high mean $P - E$"). Panel (b) repeats the exercise after classifying all lots based on the average $P - E$ for each auction house–location combination.



(a) By past artist-level prediction errors (b) By past auctioneer-level prediction errors

Encouraged by these findings, we let our neural network predict auctioneers' prediction errors rather than prices. We use a network architecture that is identical to the one presented before, except that we add a variable measuring the auction house pre-sale estimate. We denote by ML_{P-E} our benchmark prediction—relying on both the image and all non-visual characteristics—of $P - E$. In line with our earlier analysis, we can then study how auctioneers' prediction errors and buy-in rates correlate with our ex-ante "prediction error prediction" ML_{P-E}. The results are shown in

[21] This stickiness in estimates may also explain the pattern in Figure 1, where artists that have had high prices recently—probably outperforming ex-ante expectations—are associated with pre-sale estimates that are relatively low.

columns 1, 3, and 5 of Table 5. ML_{P-E} explains almost 7% of the variation in $P - E$, and also has a strong correlation with the buy-in probability.

Table 5: **Informational efficiency of pre-sale estimates (continued)**

Columns 1 and 2 of this table report estimated ordinary least squares coefficients for a linear regression model that has the auctioneer's prediction error (i.e., $P - E$) as the dependent variable. Columns 3–6 report estimated ordinary least squares and probit coefficients for regression models where the dependent variable is a dummy variable that equals one if a lot is "bought in" (i.e., if the highest bid remains below the reserve price). The models are estimated using the transactions in our test data set. Standard errors, which are two-way clustered at the artist and auction month level (except in columns 5 and 6 where clustering is only at the artist level), are reported in parentheses. The asterisks *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

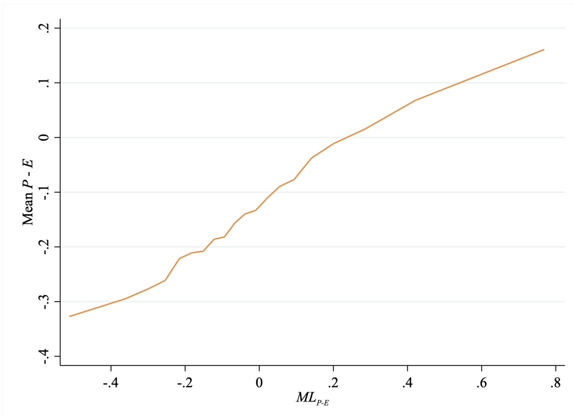
	(1)	(2)	(3)	(4)	(5)	(6)
	P - E		*Dummy = 1 if buy-in*			
	OLS		OLS		Probit	
ML_{P-E}	0.420***	0.360***	−0.290***	−0.237***	−0.854***	−0.688***
	(0.028)	(0.031)	(0.027)	(0.034)	(0.034)	(0.034)
ML_{orth}		0.073***		−0.059***		−0.166***
		(0.009)		(0.009)		(0.010)
E		0.003		−0.014***		−0.039***
		(0.002)		(0.003)		(0.005)
Constant	−0.127***	−0.154***	0.331***	0.478***	−0.452***	−0.044
	(0.007)	(0.023)	(0.008)	(0.031)	(0.009)	(0.047)
N	57,382	57,382	59,363	57,382	59,363	57,382
R^2	0.067	0.080	0.032	0.041	0.026	0.033

The results are visualized in Figure 14. Panel (a) illustrates the very strong sensitivity of realized to predicted logged price-to-estimate ratios. Panel (b) evidences that our ex-ante predictions of price-versus-estimate discrepancies line up with buy-in probabilities. It is striking that in both panels the relation between predictions and auction outcomes is even stronger than in Figure 11. Using data on past errors, machine learning can thus help to identify situations in which human experts are likely to be biased.

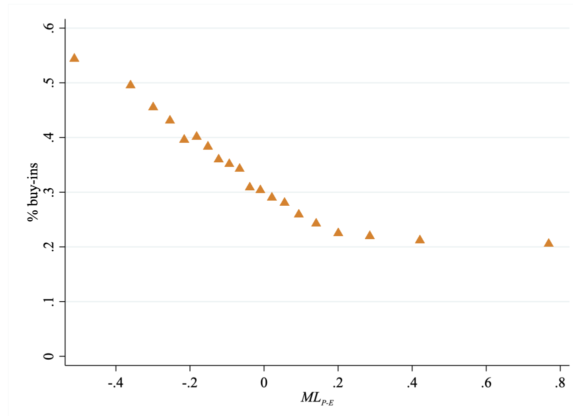
To shed more light on what drives the predictability of auctioneers' prediction errors, we document the R-squared of linear regressions of $P - E$ against our benchmark prediction and different variations thereon in Table 6. The first row shows that, as also reported in column 1 of Table 5, our predictions ML_{P-E} explain 6.7% of the variation in $P - E$. The second row of Table 6 shows results for predictions generated without information on the pre-sale estimate. The R-squared only drops to 5.3%—a relatively minor change given the importance that one might

Figure 14: **Predicting auctioneers' prediction errors directly**

Panel (a) of this figure shows the average auctioneer prediction error (i.e., $P - E$) over all lots in our test set as a function of the "prediction error predictions" ML_{P-E}, which are averaged by half-decile. Panel (b) shows average buy-in rates as a function of ML_{P-E}.



(a) $P - E$ as function of ML_{P-E}



(b) Buy-in rate as function of ML_{P-E}

have assumed the pre-sale estimate to have in predicting the ex-post realized price-to-estimate ratio. Auctioneers' prediction errors are thus to an economically significant extent predictable from a truly ex-ante perspective, i.e, without knowing the pre-sale estimate of the auctioneer. The third row of Table 6 also drops the information on the identity and location of the auction house. We see that the R-squared is lowered to 3.4%. Auction house effects are thus an important factor in the predictability of prediction errors. If we further drop information related to the artist or creation period, we see that the explanatory power deteriorates even more, as we could have expected.

Table 6: **Assessing individual variable importance in predicting auctioneers' prediction errors**

This table reports the R-squareds of linear regressions of auctioneers' prediction errors ($P - E$) in the test data set against different predictions generated by our neural network (ML_{P-E}). The first row shows results for the benchmark model that includes all predictive variables, while the next rows show results for models that exclude different sets of variable types. All models include the artwork image.

	R^2
Benchmark model	6.7%
Without pre-sale estimate	5.3%
Without pre-sale estimate + auction-related info	3.4%
Without pre-sale estimate + auction-related info + artist identifiers	1.7%
Without pre-sale estimate + auction-related info + artist/style info	3.4%
Without pre-sale estimate + auction-related info + artwork year	3.3%
Without pre-sale estimate + auction-related info + artist identifiers + artist info + artwork year	1.2%

Finally, we can of course combine the prediction of the prediction error (i.e., ML_{P-E}) with the relative magnitude of the artwork valuation (i.e., ML_{orth}) in one predictive model, which is what we do in columns 2, 4, and 6 of Table 5. We see that ML_{P-E} and ML_{orth} carry independent predictive power for auctioneers' prediction errors and buy-in rates, and can thus be considered as being complementary.

7 Discussion and Implications

There is a long lineage of research linking prices of infrequently-traded "real" assets—artworks and other collectibles, but also real estate—to their quality-determining characteristics. In most of such papers, predictability has not been a goal in itself. Instead, most researchers are interested in the value of a hedonic characteristic, or use hedonic models to control for time-series variation in average quality when estimating price trends.[22] The practical usefulness of hedonic models in terms of asset valuation has arguably remained relatively limited, as market values of artworks or houses are not always well-described by a linear function of their value-determining characteristics. Hedonic models simply cannot capture the complexity of art or real estate pricing. Hence, participants in markets for such assets have historically relied on the eyes and expertise of human valuers.

The advent of machine learning is challenging this role of human expertise. It has already led to a range of new business models built on automated valuation methods, such as "iBuyers" (Buchak *et al.*, 2020). While it has not been our aim to come up with the best possible art price prediction algorithm—we could have collected additional information on artist fame and networks, artwork provenance and exhibition history, etc.—our paper *does* shed more light on the potential and drivers of price predictability in markets for illiquid real assets. One of the implications of our findings is that, even if modern machine-learning techniques are unlikely to completely replace human judgment, they are likely to become important tools for investors and intermediaries, as they have the ability to explain much of the variation in market values in a time-efficient and relatively inexpensive manner.

Our work also shows how the asset valuations generated by machine learning can be used as a

[22] An exception is Ashenfelter (2008), which explains variation in Bordeaux wine prices using weather data by estimating a simple linear regression model, and shows that the market for young wines is inefficient.

benchmark to evaluate human experts' valuations. We demonstrate that art auctioneers' pre-sale estimates are informationally inefficient. The added value of machine learning is not uniform, but depends—in a systematic way—on a number of characteristics of the asset category, including the level and dispersion of prices, and the availability of past transaction information.

Moreover, tools similar to those used to predict *prices* can be employed to predict *pricing errors*. We find that art auction houses are systematically biased in predictable ways. Whether the "predictable prediction errors" of auctioneers mainly stem from behavioral or from strategic biases is not easy to tell. As we explained in the opening paragraphs of this paper, they may sometimes have similar effects. More fundamentally, however, we lack a good theoretical understanding of, first, what determines optimal estimates in a setting where auction houses compete for consignments and where both consignors' and bidders' behavior may be affected by the pre-sale estimate, and, second but related, how deviating from "honesty" could ever be an equilibrium longer-run strategy for auctioneers. This is definitely an avenue for further research.[23]

Overall, however, we would argue that the predictability shown in Section 6 of this paper is likely to be related to a combination of different factors. The fact that auction house effects are so important in explaining prediction errors and buy-in rates suggests that different auction houses implement different strategies in terms of pre-sale estimates; maybe there exists heterogeneity in the way in which auctioneers weigh the costs and benefits of higher vs. lower estimates. Other results in this paper point more to the importance of behavioral frictions, in particular the artist-level persistence—across all auction houses—in prediction errors.

Whatever are the precise sources of non-fundamental variation in auction house estimates, these inefficiencies and biases are not just a side show. In settings where buyers and sellers are affected by human experts' appraisals, the discrepancy between these appraisals and an unbiased proxy for market values will correlate with relevant economic outcomes. For example, when consignors set reserve prices in line with auction houses' expectations, buy-in rates will be higher if auctioneers are too optimistic, which is indeed what we have found. But also bidders may anchor on auction house

[23] In additional unreported analysis, we do not find a relation between prediction errors on the one hand and the auction house's share of the market for an artist or the overall concentration of the artist's market on the other hand.

estimates. While it is difficult to formally show the extent to which this is happening, one resultant prediction in the context of our analysis would be that higher relative automated valuations—or, equivalently, less aggressive pre-sale estimates—are associated with higher post-acquisition returns. We test this hypothesis using a small sample of artworks that we could identify that initially traded in 2015 (and are thus part of our test data set) and *retraded* over the time period 2016–2018. Column 1 of Table 7 regresses annualized log returns against the initial logged price-to-estimate ratio (i.e., $P - E$). We see a strong negative correlation between the relative level of the purchase price and the post-acquisition return Column 2 then adds our orthogonalized automated valuation (i.e., ML_{orth}). In line with our hypothesis, higher automated valuations relative to pre-sale estimates are associated with (weakly) higher returns.

Table 7: **Biased pre-sale estimates and post-acquisition returns**

This table reports estimated ordinary least square coefficients for regression models where the dependent variable is the annualized log return on the artwork's resale. The models are estimated using observed resales over the period 2016–2018 of items that transacted in 2015 (and are thus part of our test data set). Standard errors are reported in parentheses. The asterisks *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)
ML_{orth}		0.053 *
		(0.031)
$P - E$	-0.167 ***	-0.190 ***
	(0.056)	(0.057)
Constant	0.002	-0.007
	(0.024)	(0.024)
N	246	246
R^2	0.035	0.047

In sum, it is clear that automated valuation methods (and human error predictions) may be very useful for both buyers and sellers in markets for illiquid real assets. Therefore, they also have the potential to change equilibrium behavior and outcomes in such markets in the future.

8 Conclusion

We study the biases in the valuations of intermediaries in the market for an important illiquid and heterogeneous real asset class. We assemble a database of over one million paintings auctioned between 2008 and 2015. We use a popular machine-learning technique—neural networks—to develop a pricing algorithm based on both non-visual and visual artwork characteristics. Our out-of-sample machine-learning valuations predict close to three quarters of the variation in auction prices. The incremental predictive power of visual artwork characteristics is limited on average. We show that the magnitude of machine learning valuations help explain price-to-estimate ratios and buy-in rates, evidencing that pre-sale estimates are not informationally efficient. The relative contribution of machine learning, conditional on the availability of human-generated valuations, is higher for artists with less dispersed and lower average prices. Finally, we document that machine-learning algorithms can predict directly whether an artwork will be valued too high or too low by auctioneers, illustrating the importance of systematic and persistent biases in human prediction errors.

References

Abis, S. (2020). *Man vs. Machine: Quantitative and Discretionary Equity Management*. Working paper.

Acemoglu, D. and Restrepo, P. (2018). The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment. *American Economic Review*, **108** (6), 1488–1542.

Agrawal, A., Gans, J. and Goldfarb, A. (2018). *Prediction Machines: The Simple Economics of Artificial Intelligence*.

Andersen, S., Badarinza, C., Liu, L., Marx, J. and Ramadorai, T. (2021). *Reference Dependence in the Housing Market*. Working paper.

Anderson, R. C. (1974). Paintings as an Investment. *Economic Inquiry*, **12** (1), 13–26.

Ashenfelter, O. (2008). Predicting the Quality and Prices of Bordeaux Wine*. *The Economic Journal*, **118** (529), F174–F184.

— and Graddy, K. (2006). Art Auctions. In V. A. Ginsburg and D. Throsby (eds.), *Handbook of the Economics of Art and Culture*, vol. 1, Elsevier, pp. 909–945.

Autor, D. H. (2015). Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives*, **29** (3), 3–30.

Beggs, A. and Graddy, K. (2009). Anchoring Effects: Evidence from Art Auctions. *American Economic Review*, **99** (3), 1027–1039.

Brown, G. R. and Matysiak, G. A. (2000). Sticky Valuations, Aggregation Effects, and Property Indices. *The Journal of Real Estate Finance and Economics*, **20** (1), 49–66.

Brynjolfsson, E., Mitchell, T. and Rock, D. (2018). What Can Machines Learn, and What Does It Mean for Occupations and the Economy? *AEA Papers and Proceedings*, **108**, 43–47.

Buchak, G., Matvos, G., Piskorski, T. and Seru, A. (2020). *iBuyers: Liquidity in Real Estate Markets?* Working paper.

Chambers, D., Dimson, E. and Spaenjers, C. (2020). Art as an Asset: Evidence from Keynes the Collector. *Review of Asset Pricing Studies*, **10** (3), 490–520.

Coleman, B., Merkley, K. J. and Pacelli, J. (2020). *Man versus Machine: A Comparison of Robo-Analyst and Traditional Research Analyst Investment Recommendations*. Working paper.

Dimson, E. and Spaenjers, C. (2011). Ex Post: The Investment Performance of Collectible Stamps. *Journal of Financial Economics*, **100** (2), 443–458.

Erel, I., Stern, L. H., Tan, C. and Weisbach, M. S. (2021). *Selecting Directors Using Machine Learning*. Working paper.

Eriksen, M. D., Fout, H. B., Palim, M. and Rosenblatt, E. (2020). Contract Price Confirmation Bias: Evidence from Repeat Appraisals. *The Journal of Real Estate Finance and Economics*, **60** (1), 77–98.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T. and Walther, A. (2020). Predictably Unequal? The Effects of Machine Learning on Credit Markets. *Journal of Finance*, forthcoming.

Gammon, M. (2019). What Auction House Estimates Really MeanâĂŤand What They Can Tell Us.

Genesove, D. and Mayer, C. (2001). Loss Aversion and Seller Behavior: Evidence from the Housing Market. *Quarterly Journal of Economics*, **116** (4), 1233–1260.

Glaeser, E. L., Kincaid, M. S. and Naik, N. (2018). *Computer Vision and Real Estate: Do Looks Matter and Do Incentives Determine Looks*. NBER Working Paper 25174.

— and Nathanson, C. G. (2017). An Extrapolative Model of House Price Dynamics. *Journal of Financial Economics*, **126** (1), 147–170.

Grennan, J. and Michaely, R. (2020). *Artificial Intelligence and High-Skilled Work: Evidence from Analysts*. Working paper.

He, K., Zhang, X., Ren, S. and Sun, J. (2015). Deep Residual Learning for Image Recognition. *arXiv:1512.03385 [cs]*.

Karayev, S., Trentacoste, M., Han, H., Agarwala, A., Darrell, T., Hertzmann, A. and Winnemoeller, H. (2014). Recognizing Image Style. *arXiv:1311.3715 [cs]*.

Kingma, D. P. and Ba, J. (2017). Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*.

Korteweg, A., KrÃďussl, R. and Verwijmeren, P. (2016). Does it Pay to Invest in Art? A Selection-Corrected Returns Perspective. *Review of Financial Studies*, **29** (4), 1007–1038.

Ma, M. X., Noussair, C. N. and Renneboog, L. (2019). *Colors, Emotions, and the Auction Value of Paintings*. Working paper.

McAndrew, C., Smith, J. L. and Thompson, R. (2012). The Impact of Reserve Prices on the Perceived Bias of Expert Appraisals of Fine Art. *Journal of Applied Econometrics*, **27** (2), 235–252.

Mei, J. and Moses, M. (2005). Vested Interest and Biased Price Estimates: Evidence from an Auction Market. *Journal of Finance*, **60** (5), 2409–2435.

Milgrom, P. R. and Weber, R. J. (1982). A Theory of Auctions and Competitive Bidding. *Econometrica*, **50** (5), 1089–1122.

Pownall, R. A. J. and Graddy, K. (2016). Pricing Color Intensity and Lightness in Contemporary Art Auctions. *Research in Economics*, **70** (3), 412–420.

Renneboog, L. and Spaenjers, C. (2013). Buying Beauty: On Prices and Returns in the Art Market. *Management Science*, **59** (1), 36–53.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, **82** (1), 34–55.

Salzman, D. and Zwinkels, R. C. J. (2017). Behavioral Real Estate. *Journal of Real Estate Literature*, **25** (1), 77–106.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, **15** (1), 1929–1958.

Strezoski, G. and Worring, M. (2017). OmniArt: Multi-task Deep Learning for Artistic Data Analysis. *arXiv:1708.00684 [cs]*.

Tan, W. R., Chan, C. S., Aguirre, H. E. and Tanaka, K. (2016). Ceci n'est pas une pipe: A deep convolutional network for fine-art paintings classification. In *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 3703–3707.

Velthuis, O. (2007). *Talking Prices: Symbolic Meanings of Prices on the Market for Contemporary Art*. Princeton studies in cultural sociology, Princeton, N.J.: Princeton Univ. Press, 4th edn.

Appendix A Descriptive Statistics of Input Variables

Table A.1: **Descriptive statistics for training set**

		% missing	Distinct	Mean	Median	Mean hammer ($) if 1
i.	Artist	0.6%	116,694			
ii.	Artist nationality	3.9%	168			
iii.	Artist birth year	7.0%		1883	1897	
	Artist death year	28.5%		1938	1958	
iv.	Style: Old Masters	31.9%		0.06		87,504
	Style: 19th Century European	31.9%		0.17		16,839
	Style: Impressionist and Modern	31.9%		0.14		126,806
	Style: Post-War and Contemporary	31.9%		0.18		117,635
	Style: American	31.9%		0.08		25,261
	Style: Latin American	31.9%		0.02		45,869
	Style: Asian	31.9%		0.04		129,423
v.	Artwork creation year	49.9%		1959	1963	
vi.	Artwork width (cm)	2.4%		64.8	55.3	
	Artwork height (cm)	0.4%		62.7	52.0	
vii.	Materials: oil	0%		0.65		67,661
	Materials: watercolor	0%		0.05		7,213
	Materials: acrylic	0%		0.06		86,983
	Materials: ink	0%		0.06		106,739
	Materials: gouache	0%		0.03		28,132
	Materials: bronze	0%		0.02		80,307
	Materials: mixed media	0%		0.04		14,345
	Materials: pastel	0%		0.02		52,330
	Materials: lithograph	0%		0.01		5,018
	Materials: poster	0%		0.01		5,984
	Materials: etching	0%		0.01		6,453
	Materials: pencil	0%		0.01		79,062
	Materials: canvas	0%		0.52		84,281
	Materials: board	0%		0.14		25,595
	Materials: panel	0%		0.07		55,429
	Materials: paper	0%		0.13		50,000
	Materials: masonite	0%		0.01		35,613
	Materials: wood	0%		0.03		49,479
viii.	Title: untitled	0%		0.07		67,999
	Title: composition	0%		0.02		37,955
	Title: landscape	0%		0.05		34,492
	Title: still life	0%		0.03		54,730
	Title: figure	0%		0.02		50,415
	Title: nude	0%		0.01		106,149
	Title: portrait	0%		0.03		98,464
	Title: self-portrait	0%		0.00		407,585
ix.	Markings: signed	0%		0.78		57,775
	Markings: dated	0%		0.37		86,938
	Markings: inscribed	0%		0.10		70,409
x.	Auction house	0%	369			
xi.	Auction location	0%	230			
xii.	Auction month	0%	12			

Appendix B Hedonic Regressions Results

Table B.1: **Hedonic regression results based on training set**

	P
Artist F.E.?	Yes
Artwork width (cm) / 100	1.118***
Artwork width (cm) / 100 – squared	−0.233***
Artwork height (cm) / 100	0.935***
Artwork height (cm) / 100 – squared	−0.197***
Materials: oil	0.446***
Materials: watercolor	0.084***
Materials: acrylic	0.253***
Materials: ink	−0.229***
Materials: gouache	0.188***
Materials: bronze	0.560***
Materials: mixed media	0.154***
Materials: pastel	0.063***
Materials: lithograph	−1.735***
Materials: poster	−0.709***
Materials: etching	−1.359***
Materials: pencil	−0.349***
Materials: canvas	0.178***
Materials: board	0.060***
Materials: panel	0.178***
Materials: paper	−0.196***
Materials: masonite	0.093***
Materials: wood	0.156***
Title: untitled	−0.182***
Title: composition	−0.079***
Title: landscape	−0.109***
Title: still life	−0.051***
Title: figure	−0.061***
Title: nude	−0.079***
Title: portrait	−0.184***
Title: self-portrait	0.395***
Markings: signed	0.194***
Markings: dated	0.114***
Markings: inscribed	0.030***
Auction house F.E.?	Yes
Auction location F.E.?	Yes
Auction month F.E.?	Yes
N	959,950
R^2	0.800