Machine Learning, Human Experts, and the Valuation of Real Assets

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

June 23, 2019

Abstract

We study the accuracy and usefulness of automated (i.e., machine-generated) valuations for illiquid and heterogeneous real assets. We assemble a database of 1.1 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 valuations predict auction prices dramatically better than valuations based on a standard hedonic pricing model. Moreover, they help explaining price levels and sale probabilities even after conditioning on auctioneers' pre-sale estimates. Machine learning is particularly helpful for assets that are associated with high price uncertainty. It can also correct human experts' systematic biases in expectations formation—and identify ex ante situations in which such biases are likely to arise.

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

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 Luxembourg School of Finance, 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 Will Goetzmann, Katy Graddy, Stefano Lovo, Ramana Nanda, Christophe Pérignon, Luc Renneboog, Kelly Shue, David Sraer, and conference participants at the HEC Paris Data Day, the Northeastern University Finance Conference, and the ULB Art Market Workshop 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).

1 Introduction

Many asset markets see a rising importance of automated (i.e., machine-generated) valuations. This is particularly true for durable assets, such as property or collectibles, where trading is infrequent and assets are heterogeneous. In such markets, asset values are tightly (but in a complex manner) linked to a wide range of utility-generating (but not necessarily easy-to-quantify) characteristics. Human eyes and experience have therefore historically been deemed indispensable in valuing houses or paintings. However, advances in computer vision and machine learning may change this, which could have a major impact on how these markets function. Newly-developed automated valuation mechanisms can make markets more liquid, by leading to increased information availability and novel disintermediated trading opportunities for asset owners. In the real estate market, so-called "i-buyers" are already making offers to potential property sellers in an instantaneous and completely online fashion (The Economist (2018)). But having more accurate valuation mechanisms could also be productivity-enhancing and risk-reducing for long-existing businesses such as real estate agents and auction houses, banks lending against durable assets, and insurance companies.

Despite the fast-increasing importance of automated valuations, not much is known about their accuracy and about the conditions under which they may be helpful. In this paper, we therefore study machine-generated valuations in a setting where human expertise has up until now been considered an essential factor, namely the art auction market. Art prices are particularly difficult to predict because every object is unique and potential bidders' tastes—and thus willingness-to-pay—may exhibit substantial variation. The art auction market is also an appealing setting to examine the potential of automated valuations because we have a "human" benchmark, namely the pre-sale estimates that are published by auction houses for each lot and reflect their evaluation of the object.

We use data on 1.1 million painting auctions at more than 350 different auction houses worldwide between 2008 and 2015 from a proprietary database of art sales. For each lot, the database contains information related to the artist, the artwork, and the auction. It also provides an image of each artwork. We use the data for the period 2008–2014 (our "training set") to generate different types of statistical hammer price predictions for auctions in the first half of the year 2015 (our "test set"). These automated valuations can be compared to—and used together with—the auctioneers' pre-sale estimates for the same works.

Our first—and main—set of valuations are generated through popular machine-learning techniques based on neural networks. Neural networks can be considered as a method to define very large parametric models, in which the parameters are learned from the observations in an iterative and stochastic manner. Our most basic valuation, which we label $\hat{p}_{ML}^{txt}$, relies on independent variables derived from the textual and numerical data in the database (e.g., artist, year of creation, materials, size, etc.). Next, we use so-called "convolutional neural networks", which are often used in image-recognition tasks, to develop an algorithm that also considers the image of each work. This gives us a second market value estimate, which we denote by $\hat{p}_{ML}^{img}$. Crucially, neither pricing algorithm is exploiting any auction-related data (e.g., auction house, location), which are endogenous to auctioneers' evaluation of the object.

The second type of valuation that we use is also "automated", but relies on a more traditional and less sophisticated method, namely hedonic regressions. 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 $\hat{p}_{HR}$ for all artworks in the test set.

We then use this unique data set to study the distributional characteristics of our automated valuations. We also examine their relative performance in predicting prices, and whether they have explanatory power even after conditioning on auctioneers' pre-sale estimates. Finally, we analyze in depth why and when machine learning is particularly useful for valuing assets. We can summarize our empirical findings as follows.

First, we show that the long right tail in the distribution of art auction prices—with certain works by certain artists fetching very high prices—is mirrored in the distribution of machine-learning valuations but not in that of hedonic estimates. The lack of non-linearities and interaction effects in a standard hedonic model imply that valuations will be much closer to each other than actual prices.

Second, when regressing hammer prices against the different valuations, we find that our machine-learning valuations perform dramatically better as predictors (R^2 of 0.720 and 0.741 for $\hat{p}^{txt}_{ML}$ and $\hat{p}^{img}_{ML}$, respectively) than valuations based on the hedonic model (R^2 of 0.047). Interestingly, the incremental predictive power of images appears to be relatively limited, suggesting that machine learning may still be ineffective in associating distinctive image characteristics to economic value. Crucially, even our best machine-learning algorithm does not do as well as auctioneers (R^2 of 0.912), but a comparison of the explanatory power shows that it explains only about 20% less of the variation in price outcomes. (Note that auction house experts have access to more information about the artwork's quality and history. Also, we cannot rule out the possibility that auction house estimates *affect* bids.)

Third, we study whether machine-learning valuations have any predictive power after controlling for pre-sale estimates. In other words, we are testing the informational efficiency of auctioneers' estimates. We find that machine-learning valuations help explaining hammer prices conditional on auctioneers' evaluations. Adding our most sophisticated machine-learning valuation to a model with only the auction house estimate enables us to explain 4.1% of the hammer price variation left unexplained by auctioneers' estimates. It is also 18.3% more likely to lead to a more accurate than to a less accurate price prediction.

Fourth, we document substantial predictability of "buy-ins" (i.e., auctions where the highest bid remains below the secret reserve—typically set at a level just below the low estimate—and thus goes unsold). When the auction house estimate is low relative to the machine-learning valuation, the buy-in probability is about 25%, while this probability approaches 50% when the pre-sale estimate is high relative to the automated valuation. Machine-learning valuations thus do not only help predict prices conditional on selling, but also—because of the relation between auction house estimates and reserve prices—the probability of selling in the first place.

Fifth, we hypothesize that machine learning should be more beneficial in settings where prediction is more difficult. Certain artists are associated with a wider heterogeneity in the characteristics of their output or in potential buyers' tastes. The prices of works by such artists will then consistently exhibit more dispersion. They will also be more difficult to predict accurately—especially by human experts—as disentangling the different drivers of cross-sectional and temporal variation in prices becomes more challenging. We show empirically that there is indeed persistence in the artist-level deviations of prices from pre-sale estimates. Moreover, we find that our machine-learning valuations have a substantially higher incremental explanatory power—over and above auctioneers' pre-sale estimates—for those artists with the highest levels of relative "price uncertainty". Machine learning also is more helpful for works by artists with lower average price levels and higher transaction volume, ceteris paribus.

Sixth, and finally, we show that machine-learning valuations can help to overcome human experts' biases. We start by presenting evidence that auctioneers' prediction errors exhibit systematic patterns. We hypothesize that auction houses have a tendency to ignore negative information and are reluctant to adjust their valuations downwards. Empirically, we find that pre-sale estimates are indeed too high on average for works by artists that are associated with relatively low prices or returns in recent years. We also document that, for these subsamples of assets, machine-learning valuations are especially likely to improve prediction accuracy. We then show that, more generally, machine learning can help us identify ex ante situations in which such biases are likely to arise. Indeed, from recent transaction data (and without information on the actual estimate), machine-learning algorithms can predict—to an economically very significant extent—whether an artwork is likely to be under- or overestimated by auction houses.

1.1 Related Literature

Our paper contributes to different strands of literature. First, there exists a growing body of work that applies machine-learning techniques to "predict" asset prices and expected returns. A number of recent papers use machine learning to study the cross-section of equity returns (see Gu et al. (2018) and the references therein). Closer to our empirical setting, Lee and Sasaki (2018) find strong predictive power of online home value estimates from Zillow.com in explaining house transaction prices, even when controlling for house and neighborhood characteristics. However, it is unknown what information enters the estimates. Another caveat of their study is that the authors do not have access to human experts' valuations for the same properties. Also related to our work, because similar in data and methods, is a recent paper by Glaeser et al. (2018) that uses computer vision techniques to link (neighboring) houses' appearance to home values.

Second, some recent papers have studied the relative strengths and weaknesses of "men" (or traditional statistical methods) vs. "machines" in financial-economic decision-making, e.g., in investment management (Abis (2017)), new venture financing (Catalini et al. (2018)), household credit provision (Fuster et al. (2018)), or the selection of corporate directors (Erel et al. (2019)).

Third, our paper relates to the literature on price expectations formation in real asset markets. A number of studies have analyzed whether auction house pre-sale estimates are unbiased and informationally efficient (Bauwens and Ginsburgh (2000), Ashenfelter and Graddy (2003), Mei and Moses (2005), McAndrew et al. (2012)). These papers, which often use relatively small samples, come to conflicting conclusions. Beggs and Graddy (2009) look into anchoring effects in the art market. Lovo and Spaenjers (2018) build a model in which art owners form expectations of resale revenues based on the distribution of tastes and the state of the economy. A larger literature studies price expectations and selling decisions in the housing market (see, for example, Glaeser and Nathanson (2017), Bailey et al. (2018), Andersen et al. (2019), and the references therein).

2 Art Auction Data

2.1 Art Auctions

Art auctions are typically organized as "English" (i.e., ascending-bid, open-outcry) auctions. Prior to the auction, auction house experts publicly share a low and a high estimate for each lot.[1] Auction houses base these estimates on the quality, condition, rarity, etc. of the lot, considering recent auction prices for similar objects. Each consignor sets a reserve price (in agreement with the auction house), which is the lowest price she is willing to accept. Auction houses do not disclose this reserve, but it cannot exceed the low estimate. If the highest bid at the auction meets or exceeds the reserve price, the object will be sold at this price—the "hammer price".[2] If the highest bid remains below the reserve price, the item is said to be "bought in"; it does not sell and instead

[1] These estimates can be found online and in printed sales catalogues. Exceptionally, these estimates are only available "on request", typically for very valuable items.

[2] 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.

returns to the consignor.

2.2 Data

The analysis in this paper relies on proprietary data coming from the Blouin Art Sales Index (BASI), which tracks auction sales at hundreds of auction houses worldwide, including the two most important ones, namely Christie's and Sotheby's. The data have been used before by Korteweg et al. (2016).

We use data on paintings offered at auctions over the period between start-2008 and mid-2015. In total, our data set contains information on close to 1.1 million lots at 371 different auction houses—some of which have different locations—of works by about 125,000 individual artists. 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. (The year 2008 is the first one for which the underlying database has complete coverage of buy-ins.) For each lot, the database contains information related to the artist (artist name, birth and death year, nationality), the artwork (title, size, year of creation, markings (e.g., signed, dated), some details on materials), and the auction (auction house, auction date, pre-sale low and high estimate, a buy-in indicator, hammer price (if sold)). All estimate and price data are converted to U.S. dollars using the spot rate at the time of the sale. Crucially, we also have access to a high-quality image of each painting through BASI.

In the analysis below, we will use the 965,062 observed lots over the period 2008–2014 as our "training set", on which we will develop our (hedonic and machine-learning) price prediction algorithms. Our "test set" comprises the 99,203 painting auctions recorded for the first six months of 2015.[3] Table 1 shows some descriptive statistics for both subsamples of our data set. About two

[3]The results are not very different when using data for the complete calendar year. However, limiting the

thirds of all consigned lots sell successfully. In our training data, the average (median) hammer price is \$63,160 (\$3,733), with a long right tail of very expensive paintings. The difference between median and mean prices points to substantial skewness in the value distribution. In the following, we will work with natural logs of prices.

Table 1: **Descriptive statistics**

This table reports descriptive statistics separately for the data set that we use for training our automated valuation methods (painting auctions over the period January 2008 to December 2014) and for the data set that we use for testing the accuracy of the different price predictions (painting auctions over the period January to June 2015).

	2008–2014 Training set	Jan–June 2015 Test set
N observations	965,062	99,203
N distinct artists	115,792	36,222
N distinct auction houses	368	266
% with pre-sale estimates	97.9%	96.7%
Median low estimate (\$)	3,000	2,000
Mean low estimate (\$)	38,458	41,114
Median high estimate (\$)	4,300	3,000
Mean high estimate (\$)	54,673	58,185
% sold	64.7%	64.9%
Median hammer price (\$)	3,733	2,484
Mean hammer price (\$)	63,160	72,672
Median price-to-low-estimate ratio	1.040	1.038
Mean price-to-low-estimate ratio	1.475	1.442
Median price-to-high-estimate ratio	0.786	0.778
Mean price-to-high-estimate ratio	1.051	1.023

analysis to the first six months after the end of the training data allows for a fairer comparison between experts and machines. The pre-sale estimates for the lots that we consider have been determined around the end of 2014. If we also included lots for the second half of 2015, then we would compare automated valuations based on information until the end of 2014 to pre-sale estimates that incorporate information about transactions and market conditions until at least the summer of 2015.

2.3 Auctioneers' Pre-Sale Estimates

As explained before, the auction house typically shares a "low" and a "high" estimate prior to an auction. Table 1 includes some descriptive statistics for both estimates and for price-to-estimate ratios. The median sale is associated with a price that is virtually equal to the low pre-sale estimate and about 80% of the high estimate; the mean price-to-estimate ratios are higher. Below, we will use the logged average of the low and the high estimate, which we can label as $\hat{p}_{AH}$, as a predictor of the price outcome.[4]

Auction houses will typically argue that their estimates serve as "as an approximate guide to current market value and should not be interpreted as a representation or prediction of actual selling prices". Yet, at the same time, the estimates are said to be representing auction house experts' opinion "about the range in which the lot might sell at auction" (quotes taken from Sotheby's website).

One concern could be that auction houses may strategically choose to be relatively aggressive or conservative in their estimates, in an attempt to affect bidder participation or bids conditional on participation. Yet, auction theory states that "honesty is the best policy" (Milgrom and Weber (1982)). Moreover, even if estimates are shaded upwards or downwards, higher estimates should still imply higher expected prices, which is what will really matter for our empirical analysis since we will focus on the predictive power of estimates.

Auction houses' pre-sale estimates will be based on some of the easily-observable tangible characteristics of the artwork presented and used below in the development of our price prediction algorithms, such as the identity of the artist, materials, size, etc. However, auctioneers will also

[4]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.

take into account artwork condition, provenance, and other quality factors not observable to the econometrician working with art sales databases such as ours. Therefore, we might reasonably expect auction house estimates to be better predictors of prices than valuations—automated or not—based on a smaller information set.

There exists another reason for why auctioneers might be expected to "beat" our algorithms. Namely, any random noise element in auctioneers' estimates will spill over into bids if potential buyers anchor on those estimates (Mei and Moses (2005)). By contrast, our own automated valuations were of course not available to potential buyers at the time of the auctions that we study.

3 Automated Valuations

We will now use the training set—works auctioned over the years 2008–2014—to develop algorithms that predict the price of any artwork based on its characteristics. Both when using machine learning techniques and when running standard linear hedonic regressions, we train the algorithms on hammer prices for successful sales only.[5] All prices are log-transformed. Moreover, as we will drop artworks with mid estimates below $1,000 in our empirical tests in the next section, we winsorize all prices at this level prior to our training.[6]

[5]Crucially, results are very similar when using an imputed price equal to 75% of the low estimate for bought-in works. (Ashenfelter and Graddy (2011) estimate the average reserve to be approximately equal to 71% of the low estimate. McAndrew et al. (2012) find an average reserve of 73.5% of the low estimate.)

[6]We also winsorize a handful of prices at $50 million. Any price variation above this level is arguably largely idiosyncratic. In any case, we do not have any works with an estimate exceeding this level in our test data.

3.1 Machine Learning

3.1.1 Methodology

The machine-learning technique that we employ in this paper is neural networks. Neural networks can be seen as a way to define very large parametric models. Their parameters—also called "weights"—are typically learned from observations in an iterative and stochastic manner. The backbone of the architecture we use is a "multi-layer perceptron", which alternates between linear operations and non-linearities.

To generate representations from the artwork images, we use convolutional neural networks (CNNs). CNNs are often used in image-recognition tasks. They are neural networks designed to 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. CNNs are able to predict very reliably semantic and texture information from an image. Given sufficient training data, they can predict the genre, creation date, or creator of an artwork, as well as human aesthetic judgments (Karayev et al. (2013), Tan et al. (2016), Strezoski and Worring (2017)).

Because neural networks can have hundreds of thousands—or even millions—of parameters, they can often perfectly explain any data they are trained with. Therefore, it is crucial to apply them first to validation data during training (to optimize meta-parameters such as their architecture and training procedure), and then to a test set that they have not seen during training and that has not been considered for validation (to test their out-of-sample performance).

3.1.2 Variables and Valuations

Using the methods described above, we come up with two different machine-learning valuations. Our most basic valuation, which we label $\hat{p}_{ML}^{txt}$, relies only on the following independent variables derived from the textual and numerical information in the database:

- **Artist and artist nationality**. Table 1 showed that we have more than 100,000 distinct artists in the training data. Together, these artists represent 168 different nationalities.

- **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.

- **Artwork size**. Width and height are included in the database for nearly all observations. We winsorize these size variables at the 0.5% lowest and highest values in each year. After winsorizing, the median (mean) width in the training data is 55 (66) centimeters, while the median (mean) height is 52 (63) centimeters.

- **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 81.9% of all works.

- **Artwork title**. We create eight indicator variables for the following groups of terms that are used frequently in artwork titles:[7] (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. These categories are not mutually exclusive. In the training data, at least one of these indicator variables equals one for 22.5% of all works.

[7] To come up with this classification, we consider the 50 most frequent titles in our sample, and manually create groups of related words.

- **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 2.7% 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").

The algorithm also has access to the year of sale. When generating out-of-sample estimates of market values for the test set (i.e., early 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.

Next, we develop an algorithm that considers the image of each painting in the way described before. This gives us the machine-learning valuation denoted by $\hat{p}_{ML}^{img}$. This valuation thus relies on non-visual characteristics, visual characteristics, and any interactions between the two.

Crucially, neither $\hat{p}_{ML}^{txt}$ nor $\hat{p}_{ML}^{img}$ uses any auction-related variables, such as the identity and location of the auction house, or the month in which the auction takes place. Where and when a lot is being offered is endogenous to auctioneers' evaluation of the market value of the object, so a valuation or prediction algorithm that uses such information cannot be considered fully "automated". Moreover, a method that necessitates such information would not be useful to most market participants, as it would not allow to value an item as long as it is not coming up for auction.

3.2 Hedonic Regressions

A natural benchmark, in particular for $\hat{p}_{ML}^{txt}$, is the price estimate generated by a standard linear hedonic regression model applied to the artwork's characteristics. More specifically, we can estimate the following model using ordinary least squares on the observations in the training set:

$$p_{i,t} = \alpha + X_i^{'}\beta + \gamma_t + \varepsilon_i, \tag{1}$$

where $p_{i,t}$ is the log-transformed hammer price of painting i, X_i is a vector of hedonic variables, and γ_t are year fixed effects. We here use the following hedonic variables: artist fixed effects, artwork height and width (and their squares), and the artwork marking, title, and material dummies introduced before. Unlike hedonic models estimated to measure quality-controlled changes in average price levels over time (e.g., Renneboog and Spaenjers (2013)), we do not include controls for auction-related variables. Instead, we limit ourselves to truly exogenous artwork characteristics, for reasons outlined before.

Table 2 shows the hedonic regression coefficients. (We do not show standard errors or significance levels, but nearly all coefficients are highly statistically significant.) The results are in line with findings in the previous literature. For example, substantially higher prices are paid for works that are bigger, signed or dated, self-portraits, and created with oil. The (in-sample) R^2 is 0.795, mainly thanks to the many artist fixed effects; estimating the same model without artist dummies yields an R^2 of 0.175.

We can then use the estimated coefficients reported in Table 2 to generate out-of-sample price predictions $\hat{p}_{HR}$ 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.

Table 2: **Hedonic regression coefficients**

This table reports estimated ordinary least squares coefficients for the hedonic regression model shown in Eq. (1). The dependent variable is the logged hammer price. Height and width are measured in meters. The model is estimated over all transactions in our training data set, which covers the period 2008–2014.

Year fixed effects	Yes	Materials: oil	0.597
Artist fixed effects	Yes	Materials: watercolor	-0.002
Height	1.283	Materials: acrylic	0.329
Height squared	-0.317	Materials: ink	-0.208
Width	1.430	Materials: gouache	0.265
Width squared	-0.301	Materials: bronze	0.710
Markings: signed	0.226	Materials: mixed media	0.139
Markings: dated	0.152	Materials: pastel	0.080
Markings: inscribed	0.062	Materials: lithograph	-2.434
Title: untitled	-0.185	Materials: poster	-0.743
Title: composition	-0.165	Materials: etching	-1.799
Title: landscape	-0.155	Materials: pencil	-0.311
Title: still life	-0.068	Materials: canvas	0.260
Title: figure	-0.078	Materials: board	0.099
Title: nude	-0.109	Materials: panel	0.266
Title: portrait	-0.246	Materials: paper	-0.162
Title: self-portrait	0.618	Materials: masonite	0.137
		Materials: wood	0.192
N			607,963
R^2			0.795

4 Results

We will in this section relate auction outcomes in the test data (i.e., in the first half of 2015) to our different predictions. We impose a number of data filters. We exclude a low number of works by artists who did not have any auctions over the training period; our hedonic model would not even generate a price prediction for such artists. We also drop objects with mid estimates below \$1,000, for which variation in (logged) prices is substantial but not very meaningful economically.[8] Finally, we drop sales where the sale 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 (e.g., a re-attribution).

4.1 Distributions of Valuations

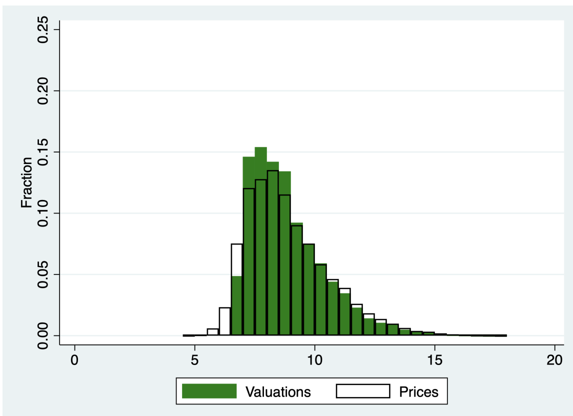
We start our analysis by plotting the distributions for all of the previously-introduced value estimates, and by comparing them to the distribution of realized hammer prices. The results are shown in Figure 1. Strikingly, the long right tail of prices is captured adequately by our machine-learning valuations, but not at all by the hedonic ones. (The left tail of prices constitutes of items that had auction house estimates in the lower \$1,000s, but sold for less than \$1,000.) Hedonic regressions are not well-suited to capture the whole distribution of prices because of the low dimensionality of the parameter space. Take, for example, the works of Pablo Picasso. All Picasso artworks will have very similar hedonic valuations, largely driven by the estimated coefficient on the artist fixed effect. Differences between hedonic Picasso predictions will be due to the average price differences—aggregated across all artists—between works with different sizes,

[8] Such objects are unlikely to have any resale value. Virtually none of these lots were offered by the main auction houses Christie's or Sotheby's.

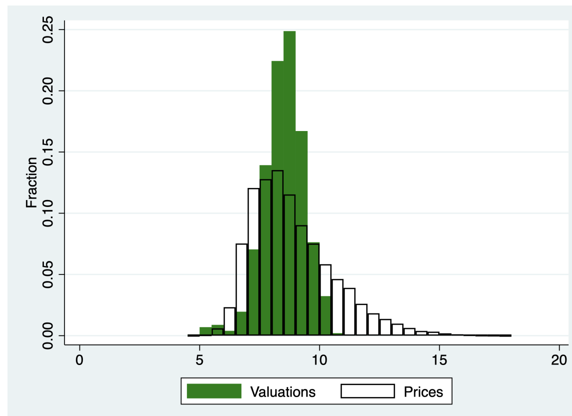
markings, materials, etc. By contrast, our machine-learning predictions can take into account non-linearities and interaction effects, for example that large oil paintings by Picasso carry an above-average premium, especially if they are, say, blue or from 1902.

Figure 1: **Distributions of valuations and prices**

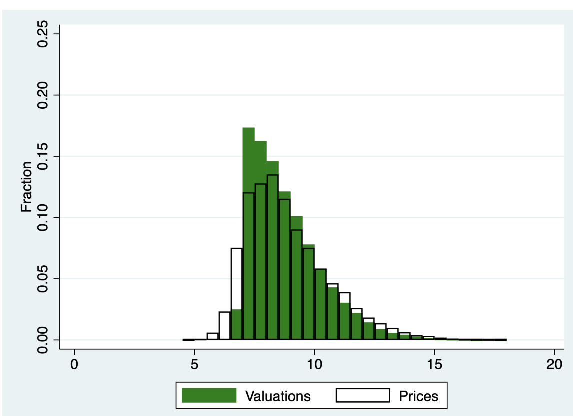
This figure shows the distributions of logged hammer prices and different valuations over all transactions in our test data set, which covers the period January–June 2015. The valuations are auction house estimates ($\hat{p}_{AH}$) in subfigure (a), hedonic valuations ($\hat{p}_{HR}$) in subfigure (b), and machine-learning valuations without and with relying on image information ($\hat{p}^{txt}_{ML}$ and $\hat{p}^{img}_{ML}$) in subfigures (c) and (d).



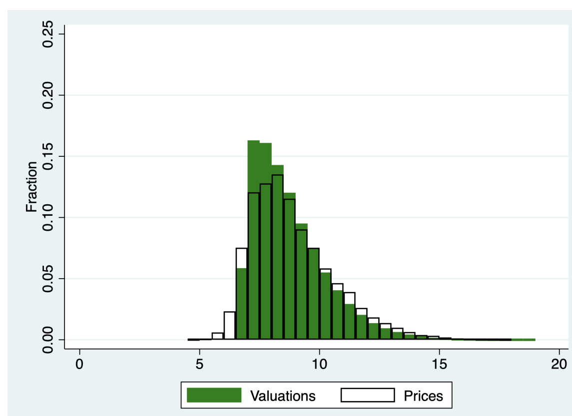
(a) Valuation used: $\hat{p}_{AH}$



(b) Valuation used: $\hat{p}_{HR}$



(c) Valuation used: $\hat{p}^{txt}_{ML}$



(d) Valuation used: $\hat{p}^{img}_{ML}$

4.2 Comparison of Predictive Power

To analyze how hammer prices line up with the different valuations, we estimate the following regression model using ordinary least squares in the test data:

$$p_i = \alpha + \beta \hat{p}_i + \varepsilon_i, \tag{2}$$

where p_i is the log hammer price of artwork i and $\hat{p}_i$ is the valuation for the same artwork. Columns 3–4 of Table 3 show the results for the different machine-learning valuations, which can be compared to the auction house pre-sale estimates in column 1 and the hedonic valuations in column 2.

Table 3: **Valuations and prices**

This table reports estimated ordinary least squares coefficients for the regression model shown in Eq. (2). The dependent variable is the logged hammer price. The models are estimated using the transactions in our test data set, which covers the period January–June 2015. Standard errors, which are two-way clustered at the artist and auction month level, are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)
Valuation:	$\hat{p}_{AH}$	$\hat{p}_{HR}$	$\hat{p}^{txt}_{ML}$	$\hat{p}^{img}_{ML}$
Valuation	1.026 ***	0.433 ***	0.965 ***	0.966 ***
	(0.005)	(0.091)	(0.012)	(0.015)
Constant	-0.242 ***	5.206 ***	0.298 **	0.338 *
	(0.054)	(0.811)	(0.112)	(0.138)
N	41,498	40,502	41,498	41,498
R^2	0.912	0.047	0.720	0.741

What do we learn from these results? First, auction house estimates explain about 91% of the variation in hammer prices. Second, a standard hedonic model performs extremely poorly in this out-of-sample setting, despite the large in-sample R^2 documented before. Third, even when only using non-visual variables, our machine-learning predictions do dramatically better than the hedonic predictions. They do not explain as much of the variation in price results as pre-sale estimates, but the R^2s in columns 3 and 4 of Table 3 are still 79% and 81%, respectively, of that in column 1. Fourth, the incremental explanatory power of images is relatively limited. The R^2 in column 4 is only 2.9% higher than that in column 3. This result suggests that machine learning may still be ineffective in associating distinctive image characteristics to economic value.[9] Fifth, average realized prices line up with both auction house estimates and machine-learning valuations near a 45-degree line: the constants are relatively close to zero, and the slope coefficients are close to one.

These can also be visualized through scatter plots. The different panels in Figure 2 show the valuations $\hat{p}_{AH}$, $\hat{p}_{HR}$, $\hat{p}^{txt}_{ML}$, and $\hat{p}^{img}_{ML}$ on the horizontal axis and hammer prices on the vertical axis. The plots based on the machine-learning valuations have a shape similar to that based on the auction house estimates, but exhibit more noise. The plot showing the relation between hedonic estimates and hammer prices looks very different. As indicated before, hedonic valuations tend to be clustered together much more.

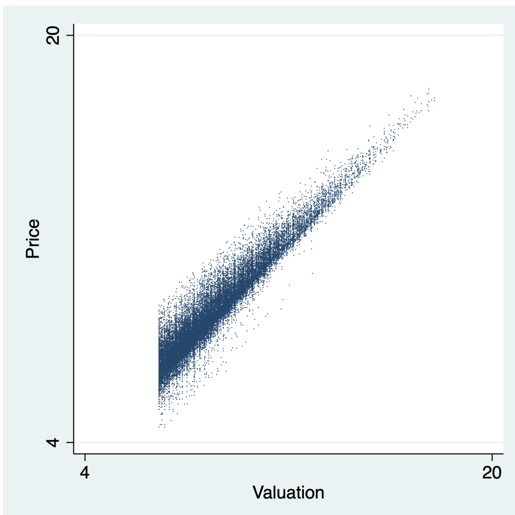
4.3 Test of Efficiency of Auctioneers' Pre-Sale Estimates

We now turn to studying whether hedonic and machine-learning valuations have some additional explanatory power in predicting hammer prices after controlling for pre-sale estimates. Another way of seeing this is as a test of the informational efficiency of auctioneers' estimates. Column 1 of Table 4 repeats the results reported in the first column of Table 3. The next columns then add the other valuations, which were first orthogonalized with respect to $\hat{p}_{AH}$, to the regression model. The bottom rows in columns 2–4 show, first, the increase in R^2 relative to the model in column 1 scaled by the hammer price variation left unexplained (i.e., $1-R^2$) by that benchmark

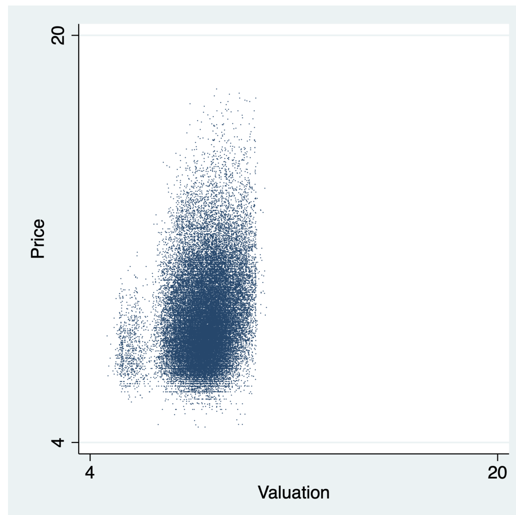
[9] Additional unreported analysis does not point to large differences across art movements (e.g., Old Masters, Impressionism & Modern, etc.).

Figure 2: **Valuations and prices**

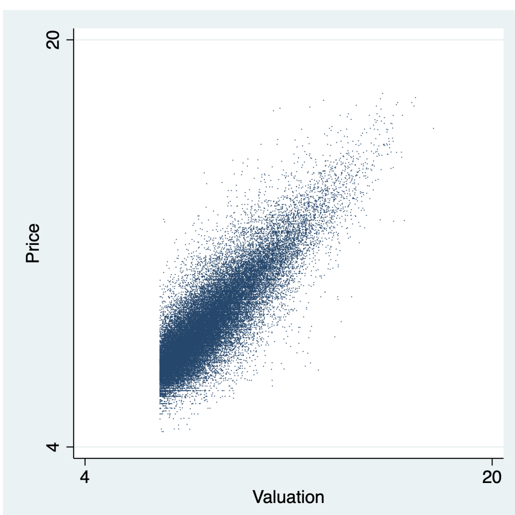
This figure plots logged hammer prices against different valuations over all transactions in our test data set, which covers the period January–June 2015. The valuations are auction house estimates ($\hat{p}_{AH}$) in subfigure (a), hedonic valuations ($\hat{p}_{HR}$) in subfigure (b), and machine-learning valuations without and with relying on image information ($\hat{p}_{ML}^{txt}$ and $\hat{p}_{ML}^{img}$) in subfigures (c) and (d).



(a) Valuation used: $\hat{p}_{AH}$



(b) Valuation used: $\hat{p}_{HR}$



(c) Valuation used: $\hat{p}_{ML}^{txt}$

(d) Valuation used: $\hat{p}_{ML}^{img}$

Table 4: **Efficiency of pre-sale estimates**

This table reports estimated ordinary least squares coefficients for regression models where the dependent variable is the logged hammer price. Column 1 only has $\hat{p}_{AH}$ as an explanatory variable, and thus repeats the first column of Table 3. Columns 2–4 then add the other valuations, which were first orthogonalized with respect to $\hat{p}_{AH}$. The models are estimated using the transactions in our test data set, which covers the period January–June 2015. Standard errors, which are two-way clustered at the artist and auction month level, are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)
Automated valuation (orthog.):		$\hat{p}_{HR}$	$\hat{p}^{txt}_{ML}$	$\hat{p}^{img}_{ML}$
$\hat{p}_{AH}$	1.026 ***	1.027 ***	1.025 ***	1.025 ***
	(0.005)	(0.005)	(0.005)	(0.005)
Automated valuation		-0.015 **	0.113 ***	0.137 ***
		(0.006)	(0.010)	(0.008)
Constant	-0.242 ***	-0.250 ***	-0.240 ***	-0.242 ***
	(0.054)	(0.057)	(0.049)	(0.052)
N	41,498	40,502	41,498	41,498
R^2	0.912	0.912	0.915	0.916
Increase in R^2 as % of variation unexplained by (1)		0.0%	3.0%	4.1%
% of predictions more accurate than (1)		51.3%	53.8%	54.2%

model, and, second, the likelihood that the predicted valuation coming out of the regression model is closer to the observed price than the prediction following from the benchmark model in column 1.

The results in Table 4 show that machine-learning valuations can help in predicting price outcomes conditional on pre-sale estimates. The increase in R^2 may not sound dramatic, but we need to consider the full extent of variation in artwork price levels that exists.[10] Moreover, from the results in the last column we can actually compute that adding a machine-learning valuation is $54.2\%/45.8\% - 1 = 18.3\%$ more likely to lead to a more accurate than to a less accurate price prediction. We thus argue the improvement in predictive power to be economically significant.

[10] Recent work on both art (Lovo and Spaenjers (2018)) and real estate (Sagi (2017)) has also highlighted the importance of a transaction-specific random noise component in durable asset prices. Some of the variation in prices is thus by construction non-predictable.

4.4 Predictability of Buy-Ins

We have so far considered the relation between our predictions and hammer prices, which are only observable if an item sells successfully. However, 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—decided jointly upon by auctioneer and consignor, but never above the 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 Table 5, which shows the results for 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. In line with our expectations, we find that machine-learning artwork valuations help predicting buy-ins.

Table 5: **Predictability of buy-ins**

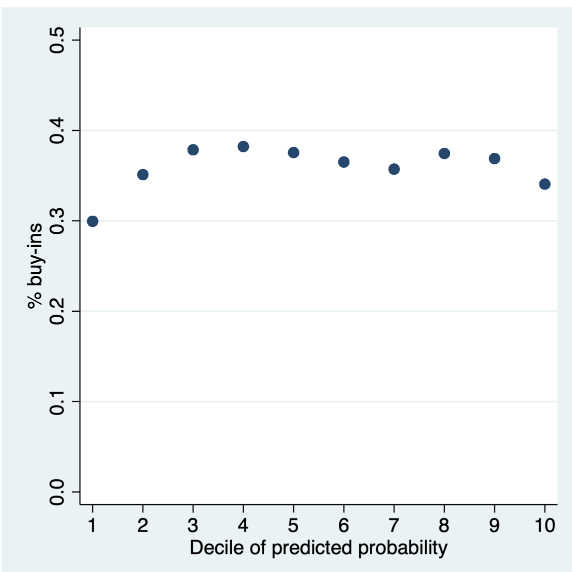
This table reports estimated probit coefficients for regression models where the dependent variable is a dummy variable that equals one if a lot is "bought in" (i.e., if the highest bid remains below the reserve price). Column 1 only has $\hat{p}_{AH}$ as an explanatory variable. Columns 2–4 then add the other valuations, which were first orthogonalized with respect to $\hat{p}_{AH}$. Each model also includes a constant, which is not shown here. The models are estimated using the transactions in our test data set, which covers the period January–June 2015. 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)
Automated valuation (orthog.):		$\hat{p}_{HR}$	$\hat{p}^{txt}_{ML}$	$\hat{p}^{img}_{ML}$
$\hat{p}_{AH}$	-0.009	-0.010 *	-0.009 **	-0.009 ***
	(0.004)	(0.004)	(0.003)	(0.004)
Automated valuation		0.003	-0.057 ***	-0.079 ***
		(0.006)	(0.003)	(0.004)
N	64,764	63,327	64,764	64,764
Pseudo R^2	0.001	0.001	0.010	0.017

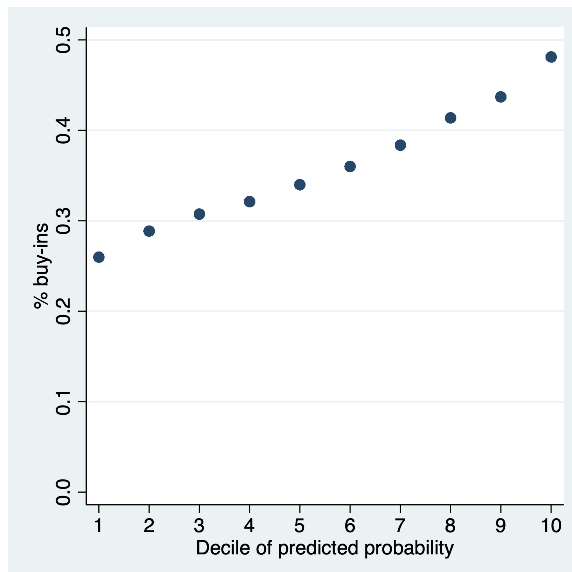
To evaluate the economic significance of our results, we plot in Figure 3 the realized out-of-sample buy-in frequency as a function of deciles of the predicted buy-in likelihoods that follow from the probit models in columns 1 and 4 of Table 5. We can see that we find substantial predictability of buy-ins when adding our machine-learning valuations to the model. Subfigure (b) shows that the buy-in probability is about 25% when $\hat{p}_{ML}^{img}$ is relatively high (keeping constant the auction house estimate), while this frequency approaches 50% when $\hat{p}_{ML}^{img}$ is relatively low.

Figure 3: **Predictability of buy-ins**

This figure shows the frequency of buy-ins over all auction lots in our test data set, which covers the period January–June 2015, for each decile of predicted buy-in probability based on two different models from Table 5. In subfigure (a), the buy-in probabilities are predicted using the model in column 1 of Table 5, which only has $\hat{p}_{AH}$ as an explanatory variable. In subfigure (b), the predicted probabilities come from column 4, which adds $\hat{p}_{ML}^{img}$ to the model.



(a) Based on $\hat{p}_{AH}$ only



(b) Adding $\hat{p}_{ML}^{img}$

In sum, machine-learning valuations thus do not only help predict prices conditional on selling, but also—because of the relation between auction house estimates and reserve prices—the probability of selling in the first place.

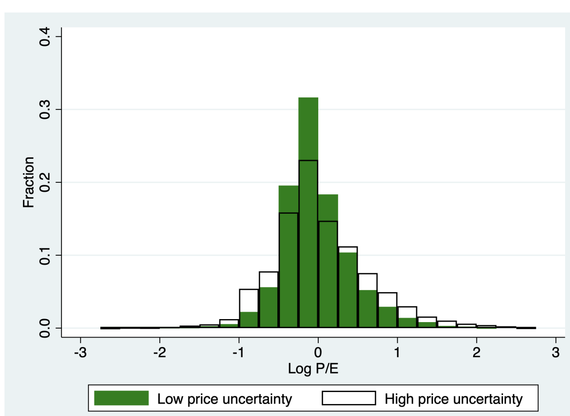
4.5 Variation in Prediction Difficulty

It is likely that some artists have a more heterogeneous output than others, and therefore exhibit more price dispersion. Also, some artists may be associated with more heterogeneity in (potential) buyers' tastes and preferences, and therefore even their quality-controlled prices fluctuate (Lovo and Spaenjers (2018)). Prices of both types of artists will be more difficult to predict accurately—both by humans and machines, but especially by the former—as disentangling the different drivers of cross-sectional and temporal variation in prices becomes more challenging.

We here first check whether there indeed exists artist-level persistence in the extent to which auction prices deviate from pre-sale estimates. We compute for each artist the average of the absolute deviations of hammer prices from pre-sale estimates (i.e., the average $|p - \hat{p}_{AH}|$) in the training data. We then rank all lots in the test data on this measure, and create four quartiles, where the first (fourth) quartile has the lots by the artists with the lowest (highest) "price uncertainty". Figure 4 compares the distributions of logged price-to-estimate ("P/E") ratios (i.e., $p - \hat{p}_{AH}$) for the extreme quartiles. We clearly see a wider distribution for high-uncertainty artists. So artists that have historically (i.e., in the training data) exhibited larger deviations from estimates indeed continue to show a higher relative price variation (i.e., in the test data).

Figure 4: **Persistence in prediction difficulty**

This figure shows different distributions of logged price-to-estimate ratios (i.e., $p - \hat{p}_{AH}$) in the test data set, which covers the period January–June 2015. We classify all lots in different quartiles based on the artist-level average of the absolute deviations of hammer prices from pre-sale estimates (i.e., the average $|p - \hat{p}_{AH}|$) in the training data set, which covers the period 2008–2014. We then show the distribution for the first quartile ("Low price uncertainty") and the fourth quartile ("High price uncertainty").



We now study whether machine-learning predictions are more helpful for the artworks that are harder to value (or, to be more precise, that are by artists who have historically been associated with less accurate auction house estimates). To do so, Table 6 repeats the analysis from Table 4 on split samples based for the first and the fourth quartile of price uncertainty. The regression coefficient on the (orthogonalized) $\hat{p}_{ML}^{img}$ is indeed much higher for the latter set of artists. Our predictions also have a much higher incremental explanatory power—over and above auctioneers' pre-sale estimates—in column 4 than in column 2.

Table 6: **Variation in efficiency of pre-sale estimates**

This table reports estimated ordinary least squares coefficients for regression models identical to those shown in columns 1 and 4 of Table 4. We classify all lots in different quartiles based on the artist-level average of the absolute deviations of hammer prices from pre-sale estimates (i.e., the average $|p - \hat{p}_{AH}|$) in the training data set, which covers the period 2008–2014. We then show estimation results for the first quartile ("Low price uncertainty") and the fourth quartile ("High price uncertainty"). Standard errors, which are two-way clustered at the artist and auction month level, are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)
	Low price uncertainty		High price uncertainty	
$\hat{p}_{AH}$	1.014 ***	1.014 ***	1.050 ***	1.052 ***
	(0.007)	(0.007)	(0.012)	(0.012)
$\hat{p}_{ML}^{img}$ (orthog.)		0.105 ***		0.186 ***
		(0.007)		(0.016)
Constant	-0.160 *	-0.160 *	-0.414 **	-0.443 ***
	(0.067)	(0.070)	(0.114)	(0.106)
N	9,857	9,857	10,110	10,110
R^2	0.922	0.924	0.879	0.888
Increase in R^2 as % of unexplained variation		2.8%		7.0%
% of predictions more accurate than benchmark		54.0%		55.8%

We can more generally examine what are the drivers of the likelihood that machine learning helps generating a more accurate valuation by running a probit regression in which the dependent variable is a dummy that equals one if $\hat{p}_{ML}^{img}$ is closer to p than $\hat{p}_{AH}$. We show the results for two specifications in Table 7. In column 1, the only explanatory variable is the price uncertainty measure created and used before. As expected, we see a strong positive correlation. In subsequent columns, we include two other covariates, namely the average estimate and number of lots for the artist (as measured in the training data). Adding these controls provides some additional insights. First, auction houses do relatively better (or, in other words, machines do relatively worse) for more expensive artists. This is intuitive, and different factors could play a role: auctioneers may have more market knowledge for such artists; auctioneers may do more effort to produce accurate estimates for such artists; hard-to-quantify factors like condition and provenance may be more important; etc. Second, auctioneers do well for relatively *illiquid* artists, at least when controlling for their price level. This may at first sound counter-intuitive, but note that the machine is not fed data on which artists can be considered "substitutes". Such art-historical knowledge may be relevant when valuing a work by an artist with few recent sales, as it will allow to consider prices in recent auctions by similar artists. Also, when not a lot of recent price points are available, auctioneers can still rely on soft information that they have about the current demand for a certain artist, while the machine is basically left in the dark.

Table 7: **Drivers of relative accuracy of machine-learning predictions**

This table reports estimated probit coefficients for regression models where the dependent variable is a dummy variable that equals one if $\hat{p}_{ML}^{img}$ is closer to p than $\hat{p}_{AH}$. The three independent variables are measured at the artist level using data from the training data set, which covers the period 2008–2014. Each model also includes a constant, which is not shown here. The models are estimated using the transactions in our test data set, which covers the period January–June 2015. 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)
Price uncertainty	0.174 ***			0.134 ***
	(0.038)			(0.038)
Log average estimate		-0.096 ***		-0.111 ***
		(0.006)		(0.006)
Log number of lots			-0.006	0.031 ***
			(0.005)	(0.005)
N	40,732	41,454	41,498	40,732
Pseudo R^2	0.001	0.007	0.000	0.008

4.6 Machine Learning vs. Human Biases

We have established that auctioneers' estimates are not always efficient, in the sense that they can be improved upon as a prediction of the hammer price. One reason may be that auctioneers' estimates are potentially biased in systematic ways.

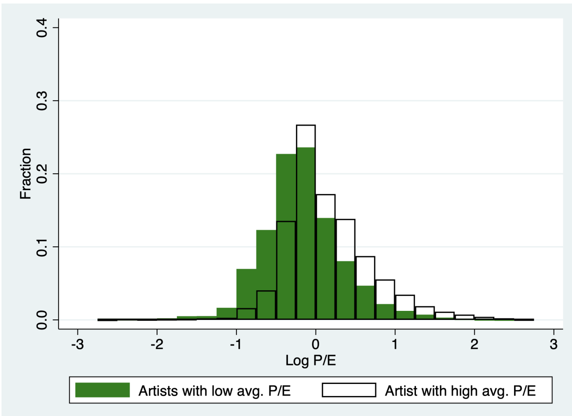
For example, we know from prior work that participants in real asset markets tend to be reluctant to adjust appraisals (or reserve prices), especially downwards.[11] We could therefore expect that recent changes in price levels—and in particular recent adverse price trends—are not always reflected in auction house estimates, which would make alternative (automated) valuations more helpful. Is this something that we see in the data? Like before, we classify all lots in the test data in quartiles, using two different metrics. First, we rank lots based on artist-level average logged price-to-estimate ratios (i.e., $p - \hat{p}_{AH}$) in the training data. Second, we can identify more than 10,000 resales of identical items in the training data, and compute artist-level average annualized returns. We then compare the distributions of price-to-estimate ratios for lots by artists with relatively low recent price surprises or returns to those by artists with relatively high recent price surprises and returns. The results are shown in Figure 5. As we were expecting, we see that low recent relative prices or returns (in the training data) are associated with low price-to-estimate ratios (in the test data).

We then compute a number of statistics for the different quartiles that we constructed. First, we measure the proportion of lots for which the machine-learning valuation exceeds the auction house estimate. Second, we report the proportion of transactions for which a prediction model like the one estimated before gives a more accurate prediction once including the machine-learning

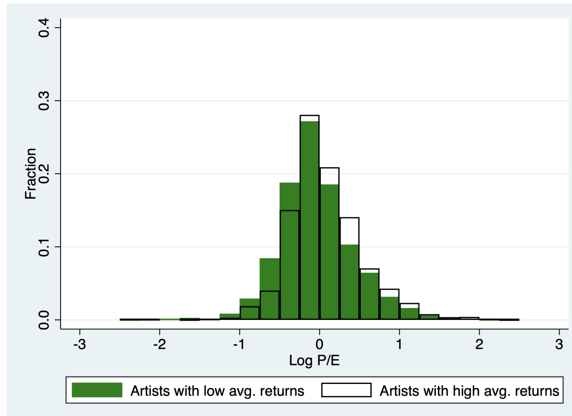
[11] This is well-known for sellers in the housing market (Genesove and Mayer (2001), Andersen et al. (2019)), but is also true for intermediaries in collectibles markets (Beggs and Graddy (2009), Dimson and Spaenjers (2011)). This can be traced back to cognitive biases such as loss aversion, anchoring, confirmation bias, belief perseverance, and preference for early information (Kahneman et al. (1982)). Intermediaries such as auction houses and dealers may also have strategic incentives to avoid price decreases.

Figure 5: **Systematic patterns in price deviations from pre-sale estimates**

This figure shows different distributions of logged price-to-estimate ratios (i.e., $p - \hat{p}_{AH}$) in the test data set, which covers the period January–June 2015. To create subfigure (a), we classify all lots in quartiles based on the artist-level average logged price-to-estimate ratio in the training data set, which covers the period 2008–2014. We then compare the distribution for the first quartile ("Artists with low avg. P/E") to that the fourth quartile ("Artists with high avg. P/E"). In subfigure (b), we do a similar exercise but creating quartiles based on the artist-level average log return on observed resales in the training data.



(a) Sample split on recent P/E ratios



(b) Sample split on recent returns

valuation (cf. columns 1 and 4 of Table 4). The results are reported in the first two rows of Panels A and B of Table 8. We find that the machine-learning valuations are much less likely to be above the pre-sale estimates and much more likely to improve accuracy for lots by artists with relatively low recent prices and returns.[12]

Of course, there may be other systematic patterns in how pre-sale estimates and sale prices differ, depending on recent price paths for certain artists, styles, etc—and on biases in expectations formation. Is it possible to predict ex ante which assets are likely to be under- or overvalued by human experts? To answer this question, we apply the machine learning methods used before to generate a prediction of the logged price-to-estimate ratio, i.e., $\widehat{p - \hat{p}_{AH}}^{img}_{ML}$. Crucially, the set of variables used by our machine-learning algorithm is exactly the same as when we were generating $\hat{p}^{img}_{ML}$, meaning that it does not include information related to the auction itself. The exercise thus

[12] We also repeat the analysis based on quartiles of artist-level price-to-estimate ratios *at the same auction house*, and the results are even stronger. This points to the importance of personal or institutional persistence in beliefs about market value.

Table 8: **Systematic patterns in relative accuracy of machine-learning predictions**

This table reports a number of statistics for different subsamples of the test data set, which covers the period January–June 2015. In Panel A, we classify all lots in quartiles based on the artist-level average logged price-to-estimate ratio (i.e., $p - \hat{p}_{AH}$) in the training data set, which covers the period 2008–2014. In Panel B, we do a similar exercise but creating quartiles based on the artist-level average logged return on observed resales in the training data.

Panel A: Lots ranked by average P/E of artist over 2008–2014

	Quartile 1	Quartile 2	Quartile 3	Quartile 4
% where $\hat{p}^{img}_{ML} > \hat{p}_{AH}$	34.2%	42.9%	47.1%	51.1%
% of predictions more accurate than benchmark	59.2%	53.8%	52.6%	53.0%
Average $\widehat{p - \hat{p}_{AH}}^{img}_{ML}$	-0.297	-0.208	-0.138	-0.009

Panel B: Lots ranked by average returns on resales of artist over 2008–2013

	Quartile 1	Quartile 2	Quartile 3	Quartile 4
% where $\hat{p}^{img}_{ML} > \hat{p}_{AH}$	36.3%	40.4%	44.4%	51.9%
% of predictions more accurate than benchmark	57.7%	53.4%	52.2%	53.6%
Average $\widehat{p - \hat{p}_{AH}}^{img}_{ML}$	-0.250	-0.189	-0.123	-0.027

consists of predicting price-to-estimate ratios *without knowing the estimate.*

As an initial check on the usefulness of this exercise, we show in the third row of Panels A and B of Table 8 the average values for $\widehat{p - \hat{p}_{AH}}^{img}_{ML}$ for each subsample. We see that it goes up with recent price surprises and returns.[13] In other words, the machine is more likely to predict that the auctioneer's estimate will be too optimistic (pessimistic) for artists with low (high) recent prices and returns.

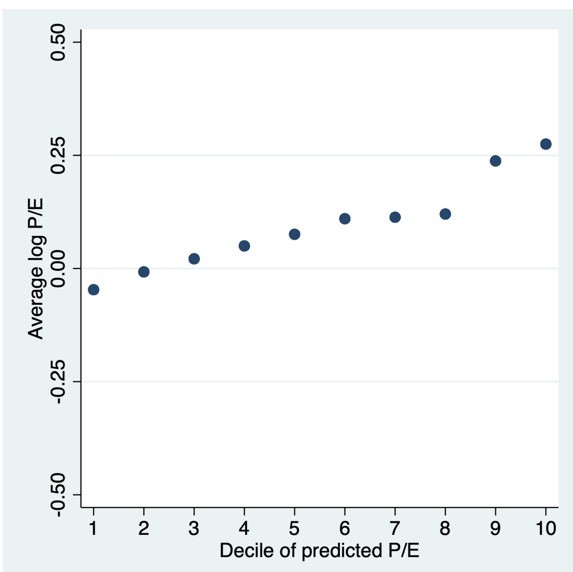
Next, we examine the relation between our machine-learning predictions of price-to-estimate ratios on the one hand and actual auction outcomes on the other hand. For this final exercise, we focus on auctions at Christie's and Sotheby's only, who are the main players in the auction market, have some of the best-known auctioneers, and sell the most expensive lots. We first create deciles (in the test data) of predicted price-to-estimate ratios. In subfigure (a) of Figure 6, we

[13] The prediction is negative on average because we included all buy-ins (with an imputed price of 75% of the low estimate) when training the algorithm.

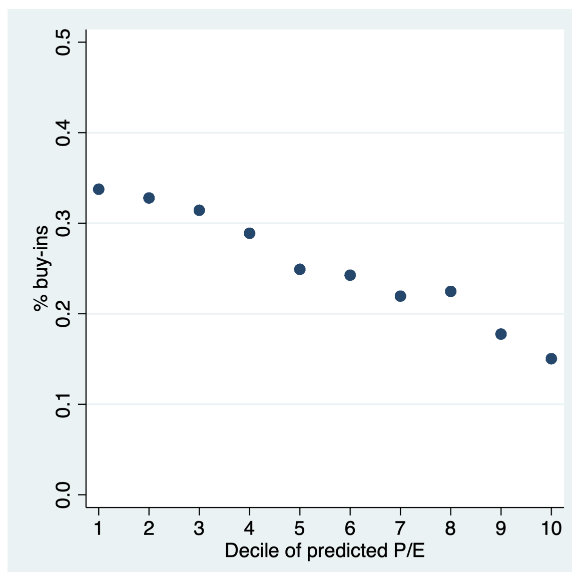
then show how the average logged price-to-estimate ratios line up with the predicted values for all successful sales. We see a near-linear relation; while the hammer price is on average below the pre-sale estimate for the lots with the lowest predictions, it tends to be substantially higher than the estimate for the lots with the highest predicted values. In subfigure (b), we show buy-in rates as a function of predicted price-to-estimate ratios. As expected, we see higher buy-in rates for those lots where the algorithm predicts a low price relative to the estimate (conditional on selling). In sum, these results show that machine learning can identify ex ante situations in which human experts are likely to be biased.

Figure 6: **Predictability of price deviations from pre-sale estimates**

Subfigure (a) shows the average logged price-to-estimate ratio (i.e., $p - \hat{p}_{AH}$) over all lots auctioned at Christie's and Sotheby's in our test data set, which covers the period January–June 2015, for each decile of predicted price-to-estimate ratio (i.e., $\widehat{p - \hat{p}_{AH}}^{img}_{ML}$). Subfigure (b) shows the frequency of buy-ins for each decile.



(a) Predictability of P/E ratios



(b) Predictability of buy-ins

5 Conclusion

We study the accuracy and usefulness of automated (i.e., machine-generated) valuations for illiquid and heterogeneous real assets. We assemble a database of 1.1 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 valuations predict auction prices dramatically better than valuations based on a standard hedonic pricing model. Moreover, they help explaining price levels and sale probabilities even after conditioning on auctioneers' pre-sale estimates. Machine learning is particularly helpful for assets that are associated with high price uncertainty. It can also correct human experts' systematic biases in expectations formation—and identify ex ante situations in which such biases are likely to arise.

Recent work has discussed the implications of machine learning for job occupations and the economy (e.g., Zeira (1998), Acemoglu and Restrepo (2018), Agrawal et al. (2018), Brynjolfsson et al. (2018)). Our results suggest that machine learning dramatically improves prediction of durable asset prices when compared to less sophisticated automated methods (i.e., hedonic regressions). It can also help to overcome certain systematic prediction errors exhibited by human experts. However, it does not seem ready to completely replace human judgement in a complex task such as valuing artworks; instead, human judgement assisted by machine learning might yield the best results.

References

Abis, Simona (2017), “Man vs. machine: Quantitative and discretionary equity management.” Working paper. URL https://www8.gsb.columbia.edu/researcharchive/getpub/25656/p.

Acemoglu, Daron and Pascual Restrepo (2018), “The race between man and machine: Implications of technology for growth, factor shares, and employment.” *American Economic Review*, 108, 1488–1542.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb (2018), *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Review Press.

Andersen, Steffen, Cristian Badarinza, Lu Liu, Julie Marx, and Tarun Ramadorai (2019), “Reference points in the housing market.” Working paper. URL https://ssrn.com/abstract=3396506.

Anderson, Robert C. (1974), “Paintings as an investment.” *Economic Inquiry*, 12, 13–26.

Ashenfelter, Orley and Kathryn Graddy (2003), “Auctions and the price of art.” *Journal of Economic Literature*, 41, 763–787.

Ashenfelter, Orley and Kathryn Graddy (2011), “Sale rates and price movements in art auctions.” *American Economic Review*, 101, 212–216.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel (2018), “The economic effects of social networks: Evidence from the housing market.” *Journal of Political Economy*, 126, 2224–2276.

Bauwens, Luc and Victor Ginsburgh (2000), “Art experts and auctions: Are pre-sale estimates unbiased and fully informative?” *Louvain Economic Review*, 66, 131–144.

Beggs, Alan and Kathryn Graddy (2009), “Anchoring effects: Evidence from art auctions.” *American Economic Review*, 99, 1027–1039.

Brynjolfsson, Erik, Tom Mitchell, and Daniel Rock (2018), “What can machines learn, and what does it mean for occupations and the economy?” *AEA Papers and Proceedings*, 108, 43–47.

Catalini, Christian, Chris Foster, and Ramana Nanda (2018), “Machine intelligence vs. human judgement in new venture finance.” Mimeo.

Dimson, Elroy and Christophe Spaenjers (2011), “Ex post: The investment performance of collectible stamps.” *Journal of Financial Economics*, 100, 443–458.

Erel, Isil, Lea H. Stern, Chenhao Tan, and Michael S. Weisbach (2019), “Selecting directors using machine learning.” Working paper. URL https://ssrn.com/abstract=3144080.

Fuster, Andreas, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther (2018), “Predictably unequal? The effects of machine learning on credit markets.” Working paper. URL https://ssrn.com/abstract=3072038.

Genesove, David and Christopher Mayer (2001), "Loss aversion and seller behavior: Evidence from the housing market." *Quarterly Journal of Economics*, 116, 1233–1260.

Glaeser, Edward L., Michael S. Kincaid, and Naik Nikhil (2018), "Computer vision and real estate: Do looks matter and do incentives determine looks." NBER Working Paper 25174. URL `http://www.nber.org/papers/w25174`.

Glaeser, Edward L. and Charles G. Nathanson (2017), "An extrapolative model of house price dynamics." *Journal of Financial Economics*, 126, 147–170.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu (2018), "Empirical asset pricing via machine learning." Working paper. URL `https://ssrn.com/abstract=3159577`.

Kahneman, Daniel, Paul Slovic, and Amos Tversky, eds. (1982), *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.

Karayev, Sergey, Matthew Trentacoste, Helen Han, Aseem Agarwala, Trevor Darrell, Aaron Hertzmann, and Holger Winnemoeller (2013), "Recognizing image style." *CoRR*, abs/1311.3715, URL `http://arxiv.org/abs/1311.3715`.

Korteweg, Arthur, Roman Kräussl, and Patrick Verwijmeren (2016), "Does it pay to invest in art? A selection-corrected returns perspective." *The Review of Financial Studies*, 29, 1007–1038.

Lee, Yong Suk and Yuya Sasaki (2018), "Information technology in the property market." *Information Economics and Policy*, 44, 1–7.

Lovo, Stefano and Christophe Spaenjers (2018), "A model of trading in the art market." *American Economic Review*, 108, 744–774.

McAndrew, Clare, James L. Smith, and Rex Thompson (2012), "The impact of reserve prices on the perceived bias of expert appraisals of fine art." *Journal of Applied Econometrics*, 27, 235–252.

Mei, Jianping and Michael Moses (2005), "Vested interest and biased price estimates: Evidence from an auction market." *The Journal of Finance*, 60, 2409–2435.

Milgrom, Paul R. and Robert J. Weber (1982), "A theory of auctions and competitive bidding." *Econometrica*, 50, 1089–1122.

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

Rosen, Sherwin (1974), "Hedonic prices and implicit markets: Product differentiation in pure competition." *Journal of Political Economy*, 82, 34–55.

Sagi, Jacob (2017), "Asset-level risk and return in real estate investments." Working paper. URL `https://ssrn.com/abstract=2596156`.

Strezoski, Gjorgji and Marcel Worring (2017), "Omniart: Multi-task deep learning for artistic data analysis." *CoRR*, abs/1708.00684, URL `http://arxiv.org/abs/1708.00684`.

Tan, Wei Ren, Chee Seng Chan, Hernán E. Aguirre, and Kiyoshi Tanaka (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)*, 3703–3707.

The Economist (2018), "Tech firms disrupt the property market." 13 September 2018.

Zeira, Joseph (1998), "Workers, machines, and economic growth." *Quarterly Journal of Economics*, 113, 1091–1117.