

# In Art We Trust

 Yuexin Li,<sup>a,\*</sup> Xiaoyin Ma,<sup>b,c</sup> Luc Renneboog<sup>d</sup>

<sup>a</sup> Renmin University of China, Beijing 100872, China; <sup>b</sup> Ant Group China, Hangzhou, China; <sup>c</sup> Erasmus University, 3062 PA Rotterdam, Netherlands; <sup>d</sup> Tilburg University, 5037 AB Tilburg, Netherlands

\*Corresponding author

Contact: [yuexinli@ruc.edu.cn](mailto:yuexinli@ruc.edu.cn),  <https://orcid.org/0000-0001-8522-553X> (YL); [maxy@uvt.nl](mailto:maxy@uvt.nl) (XM); [Luc.Renneboog@uvt.nl](mailto:Luc.Renneboog@uvt.nl),

 <https://orcid.org/0000-0002-6068-3910> (LR)

Received: March 24, 2021

Revised: March 23, 2022

Accepted: September 17, 2022

Published Online in Articles in Advance:  
December 23, 2022

<https://doi.org/10.1287/mnsc.2022.4633>

Copyright: © 2022 The Author(s)

**Abstract.** Whereas trust is the cornerstone of any market’s functioning, it is of particular importance in markets that are unregulated, illiquid, and opaque, such as the art market. This study examines the role of authenticity, as captured by provenance information in auction catalogs, on the probability of auctioned oil paintings, watercolors, and prints being sold; their price formation; and returns. Auction catalogs include four authenticity dimensions: pedigree (ownership “blockchain,” descentance information; type of past owners, such as renowned collectors; and past sales records), exhibition history (e.g., in famous museums or galleries), literature coverage (e.g., in catalogues raisonnés or authoritative press), and certification (e.g., artist’s physical testimonial, expert opinions). We find that trust, proxied by provenance information, increases the probability of a work being sold by up to 4%, leads to hammer price premiums up to 54%, and increases annualized returns by 5%–16%. To address potential endogeneity problems between the provision of provenance and past prices/price expectations, we perform quasi-natural experiments in difference-in-differences settings on auction houses’ provenance policy changes following authenticity litigation and on a contamination effect of the discovery of fakes and forgeries on the oeuvre of forged artists. We also test transactions less affected by past prices, such as estate sales following the death of a collector. The findings on the relation between provenance and prices are robust to artist reputation, artistic style, auction house reputation, art market liquidity, and artist career timing.

**History:** Accepted by Tomasz Piskorski, finance.

 **Open Access Statement:** This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. You are free to download this work and share with others, but cannot change in any way or use commercially without permission, and you must attribute this work as “*Management Science*. Copyright © 2022 The Author(s). <https://doi.org/10.1287/mnsc.2022.4633>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.”

**Funding:** Y. Li gratefully acknowledges supports from the National Natural Science Foundation of China [Grant 72204257] and the Area Studies Fund of Renmin University of China [Grant AS2022002]. This study is partially funded by the Tilburg Alumni Fund.

**Supplemental Material:** The online appendix and data are available at <https://doi.org/10.1287/mnsc.2022.4633>.

**Keywords:** auction • hedonic pricing • art investment • art returns • auction house

## 1. Introduction

Acquiring an asset not only requires an assessment of the risk–return trade-off; it also involves an act of trust in the soundness of information reflecting fundamental value and in the fairness of the overall (trading) system (Guiso et al. 2008). Trustworthy and reliable information reduces the information asymmetry between trading parties. Equity and bond investors depend on annual reports, independent audit reports, credit ratings, analysts’ forecasts, and information from financial intermediaries, among others. Reliable information is even more important for markets for alternative investments, such as art markets, which differ from financial markets in several aspects.

First, except for the market for prints, art objects offered for sale are *unique* pieces, the value of which is determined by the characteristics of the art object (e.g., topic, medium, and artist reputation) and also by subjective nonmonetary ownership utility (Renneboog and Spaenjers 2013) and by the resale option value that is affected by demand factors (e.g., wealth concentration, equity market evolution, income inequality, or changes in art-collecting audiences) (Goetzmann et al. 2011, Lovo and Spaenjers 2018, Pénasse et al. 2021). This implies that pinpointing the “fundamental” value of an art object is much more difficult than for standardized financial assets. Second, in contrast to the considerable amount of information generated about corporations

with traded equity and debt, information in the art market is *opaque* because, for the majority of auctioned art, only sparse information is available, such as the artist's name, title, topic, medium, measurements, auction house, date of sale, and lot number, which are the variables traditionally included in hedonic pricing regressions. Prices may also be slow to reflect changing valuations as the lack of short selling induces delays in the incorporation of negative information in prices. Third, the art market is *illiquid*. A decision to sell a painting can take long to be executed as the right type of auction might not be available for a specific type of art. Fourth, the secondary art market comprises hundreds of intermediaries (auction houses or dealers) around the world and is *not regulated*. A problem in the art world is that fakes and forgeries are occasionally discovered, which may erode trust, reduce market participation, and lead to lower valuations. It is rumored that up to 40%–50% of the high-end modern art market consists of forgeries (Thompson 2010).<sup>1</sup>

The art auction market is an important alternative financial market. High-net-worth individuals (HNWIs) hold on average 9% of their investment portfolios in art and other types of collectibles (e.g., Bordeaux wines, classic cars, and collector watches). The total value of collectibles held by HNWIs is estimated at more than US\$4 trillion (Deloitte Luxembourg and ArtTactic 2016). Art sales through auction houses and internet auctions have grown rapidly over the past two decades (Deloitte Luxembourg and ArtTactic 2014, 2016), and global art sales exceeded US\$40 billion in 2015 and 2016 (Pownall 2017). The finance and economics literature focuses on the risk–return relationship of art (Mei and Moses 2002, Renneboog and Spaenjers 2013, Korteweg et al. 2016, Lovo and Spaenjers 2018, Li et al. 2022); its macroeconomic market drivers (Goetzmann et al. 2011); sentiment and hype (Pénasse et al. 2014); supply shocks (Pénasse et al. 2021); and behavioral anomalies, such as anchoring (Beggs and Graddy 2009, Graddy et al. 2023), gender bias (Cameron et al. 2019, Adams et al. 2021), or overextrapolation (Pénasse and Renneboog 2022). However, this broad literature has little to say on the fundamental question of art trade, that is, how information disclosure and trust affect the market.

In the art market, information on provenance is especially important as it may increase confidence in the authenticity of the art object offered for sale. This is particularly true if the provenance information offers details about an object's pedigree (e.g., can the ownership chain uninterruptedly be traced back to the artist?), the literature (e.g., is the art object referenced in art history books or catalogues raisonnés?), exhibitions (has the object been exhibited by museums or galleries?), and certification (is there any evidence in the form of old photographs of the art object, certificates by the artist, certified witness statements, etc.?). Pedigree refers to

past ownership, which ideally can be traced back to the primary market with sales by the artists themselves or their galleries. Pedigree also comprises inheritance information within the artists' families and about collectors' families. It can be regarded as the biography of a piece of art that lived through the decades, possibly even centuries, and records the owners down to the present day. Traceability may translate into trustworthiness. There is a general upward trend of tracking ownership and providing more detailed provenance. In the current digital age, pedigree can be recorded by means of blockchain technology, which could revolutionize the logging of ownership as records cannot be forged. Companies such as Verisart provide digital certification of art transactions, which can be applied to both physical transactions as well as digitalized art by means of nonfungible tokens (Chohan 2021, Fairfield 2022). The blockchain initiative related to physical art transactions is still in infancy, and detailed analyses of provenance (or “blockchain” by means of textual analysis on auction catalogs), as undertaken in the present study, will remain important for many decades. Our study of the art auction market comprises paintings created in the preblockchain area; we consider an époque of artistic painting spanning more than 500 years (from medieval old masters to contemporary art).

The pedigree dimension also includes comprehensive lists of specific types of people appearing in the ownership chain, such as nobility and royalty, wealthy business professionals, influential politicians, celebrities, sportspeople, and other prominent individuals. The second provenance category, exhibition history, documents previous exhibitions in museums, galleries, and art fairs. Prominent exhibitions can serve as a filter for authenticity and a quality indicator because exhibitions are curated and often accompanied by an exhibition catalog, which may include new research on the exhibited art, its artist, or style. Third, literature coverage may be important; this includes books, catalogs, and scholarly articles covering the art piece. The most important reference work is the catalogue raisonné, which includes all the known artworks by an artist. The fourth category, certification, provides physical or non-physical proof of authenticity issued by the artist, the artist's close family, and art experts. Whereas all four provenance dimensions and their constituting elements undoubtedly enhance trust in art markets, some provenance variables may also embed the quality of the art object (the highest quality works may be exhibited more often or appear more in scientific art history books). In addition, prices of paintings may be inextricably related to the status or glamor of previous owners if a buyer is willing to pay a premium for a painting that was owned by, for example, royalty or celebrities (but the price premium may also reflect that such people may be wealthy enough to have bought in expert advice for their acquisitions and paid for technical analyses to verify authenticity).

We demonstrate that the probability of being sold, price levels, and returns are all significantly affected by the various dimensions of provenance. The probability of being sold increases by about 2% when the catalog documents information on pedigree, by 4% for information on the exhibition history, and by 3% when the painting is mentioned in the literature. The effect on price is even more pronounced: pedigree, exhibition background, references in the art literature, and the presence of certification drive prices up by 21%, 42%, 54%, and 14%, respectively. The annualized repeat sales returns increase by, respectively, 11, 16, and 5 percentage points when the catalog provides information on exhibitions, literature, and certification, respectively. In the models yielding these results, we control for artwork characteristics (e.g., topic, measurements, or medium) as well as artist, year, seasonality, and auction house branch fixed effects.

As the estimated relationship between provenance and probability of being sold, hammer prices, and returns may be affected by endogeneity (i.e., the decision to offer a painting for sale and the decision to provide detailed provenance information may depend on expected prices), we conduct four additional analyses to alleviate such concerns. First, we examine the provenance effects on prices of paintings by artists with some (attributed) works that have been discovered to be fakes and forgeries in a difference-in-differences (DiD) setting. We study whether the provision of provenance stems from a possible contagion effect from a forgery to all of an artist's artworks (in terms of buy-in probability, prices, and returns) (Section 3.3.1). We find that the impact of provenance increases substantially in a context of increased uncertainty about authenticity. Second, we exploit the Christie's provenance policy change in 2012 (following a litigation case) relative to Sotheby's, also in a DiD setting (Section 3.3.2). The Christie's increase in the quality of its provenance led to a price premium of 37% relative to Sotheby's transaction prices. Third, to address the effect of past prices on provenance provision, we examine the impact of incremental provenance information on prices, controlling for initial provenance at the first sale and for the previous hammer price, and conclude that incremental provenance is also priced. We report that provenance remains strongly related to prices (Section 3.3.3). Fourth, as a seller's decision to offer a painting to an auction house as well as the auction house's decision to provide provenance information may be affected by recent price evolutions (for similar paintings, e.g., by the same artist or school or for the art market as a whole), we examine the impact of provenance on the probability of being sold and the hammer prices for the subsample of sales for which we expect these decisions to have been taken exogenously. For instance, such auction sales may result from estate sales after an owner's death. Also, in this case in which

market timing would play little role, we find price premiums for provenance (Section 3.3.4). All endogeneity tests corroborate our baseline results.

As our models are estimated with high-dimensional fixed effects models, which could result in a decrease in predictive power, we perform least absolute shrinkage and selection operator estimations (LASSO), and the findings confirm earlier results. A possible issue is that provenance provision can be related to liquidity of an artist's oeuvre, and liquidity can, in turn, proxy for quality and require a premium. We study whether the provenance price premium coincides with a liquidity premium and demonstrate that this is not the case. Whereas we control for auction house branch fixed effects in all regressions, we still perform a separate analysis on auction houses and provenance in order to study whether a substitution effect exists between auction house types and provenance. We document that, for art of different price ranges by auction house type (large international, medium-sized, and small auction houses), provenance generates a price premium. Thus, there is no substitution effect, nor are provenance premiums limited to art in the highest price quantiles or offered by specific types of auction houses. Moreover, an artist's reputation is not a substitute for provenance provision; on the contrary, established artists' work may be more subject to being forged.

The rest of the paper proceeds as follows. Section 2 describes the methodology and data. Section 3 documents the empirical results and endogeneity tests. Section 4 reports extensions and robustness tests. Section 5 concludes.

## 2. Data and Methodology

### 2.1. Data and Variables

Whereas approximately half the global art market is private and comprises a primary market (galleries selling the oeuvre of living artists) as well as part of secondary markets (private sales among collectors or organized by dealers), we can study only the public segment of the secondary market, namely, the auction market organized by hundreds of auction houses around the world. This public market comprises more than half the global market and also leads in terms of price setting by providing publicly available price benchmarks for privately sold art. We focus on the market for oil paintings, watercolors, and drawings, which comprise the largest proportion of the fine art auction market. From the online database Blouin Art Sales Index, we collect all sales of paintings from 2007 to 2016. Our sample starts in 2007 as information on provenance and buy-ins (i.e., items that do not reach the undisclosed reserve price and remain unsold) is of poor quality in earlier years. Our data set includes 1,812,807 transactions of which 1,195,640 objects (66%) were sold at auction by 608 auction houses

(branches) all over the world. The paintings and drawings were created by about 150,000 artists. The average (median) hammer price is US\$53,142 (US\$3,400) with a standard deviation of US\$638,181—all real terms (2007 deflated). For each transaction, we collect all the artist, artwork, and transaction characteristics, such as artist name, title of the art object, medium (oil/acryl, watercolor, print), measurements (height and width), attribution, creation year, whether signed and/or dated by the artist, sold or unsold at auction, hammer price, auction lot number, low and high price estimates, auction date, auction house (branch), and detailed provenance information. We apply textual analysis to the provenance text in the auction catalogs to obtain 40 characteristics (see Section 2.1.2) that we categorize into four dimensions: pedigree, exhibition, literature, and certification.

**2.1.1. Traditional Hedonic Variables.** We follow Renneboog and Spaenjers (2013) by including the traditional hedonic pricing variables as control variables in our regressions.<sup>2</sup>

**2.1.1.1. Artist Characteristics.** We include artist fixed effects and a dead artist dummy (*Deceased*). The former captures each artist's uniqueness and reputation; the latter captures that prices may increase after the death of an artist because of a supply shock (Pénasse et al. 2021). In our data set, 76.7% of the auctioned paintings were from deceased artists.

**2.1.1.2. Artwork Characteristics.** We consider the following wide range of price-determining variables that capture the attribution, signature, medium, measurements, and topic of the work of art.

- Attribution: We consider six levels of attribution that capture various degrees of uncertainty/closeness to a specific artist: *Attributed* (to), *Studio* (of), *Circle* (of), *School* (of), *After*, and (in the) *Style* (of) an artist. About 3.4% of the observations in our sample carry such an attribution.

- Signature: We include *Signed*, *Dated*, and *Inscribed* variables; 80.4% of artworks are signed, about 36.3% are dated, and 11.4% are inscribed.

- Medium: The indicator variables *Oil*, *Watercolor*, and *Drawing* represent the media used. About 68.1% of the transactions are oil paintings, 20.5% are watercolors, and 11.4% are drawings.

- Measurements: *Height* and *Width* are included in centimeters (in addition to the squared values *Height\_2* and *Width\_2*).

- Topic: As the aesthetic and financial appreciation can depend on a painting's topic, we categorize the paintings based on the keyword analysis of the titles. We search for keywords in the seven languages most used in the art auction world (and its catalogs): Dutch, English, French, German, Italian, Portuguese, and Spanish. We

partition the paintings in the following categories: *Abstract*, *Animals*, *Landscape*, *Seascape*, *Cityscape*, *Nude*, *People*, *Self-Portrait*, *Portrait*, *Religion*, *Still Life*, *Study*, and *Other Topics*. *Untitled* is used as the omitted benchmark in our regressions.

**2.1.1.3. Transaction Characteristics.** We include indicator variables capturing the timing of the sale and the reputation and location of the auction house:

- Year and month: We control for year effects as well as seasonality as the most important auction seasons are in spring (May and June) and fall (November and December).

- Auction houses: We distinguish between different fine art auction houses based on reputation/size. In the case of Sotheby's and Christie's, we introduce dummy variables for their London, New York, and other branches (e.g., *Sotheby's London*, *Sotheby's New York*, and *Sotheby's Other Branches*). For two other important British auction houses, Bonhams and Phillips, we distinguish between their London sales rooms and other branches (e.g., *Bonhams London* and *Bonhams Other Branches*). We also create two dummies to account for the sales by important (large or middle sized) European and American auction houses (*Auction European* and *Auction American*).

**2.1.2. Provenance Variables.** About 14.2% of the observations in our database provide pedigree information; the proportions for exhibitions, literature, and certification amount to 3.7%, 4.0%, and 3.9%, respectively, and are presented in Table 1.

*Pedigree* considers the history of past owners and, thus, refers to ownership chains. For example, a painting might have been in the collection of the artist's family, prominent collectors, royal and noble families, wealthy families, CEOs, and celebrities. If an uninterrupted ownership chain between the artist and the current owner can be traced, the artwork has a higher probability of being authentic, and thus, there is a possibility of a premium at auction. It is also possible that a "glamour" premium is paid for a painting once owned by a famous individual.<sup>3</sup> As such, "ennobling" provenance can turn an ordinary object into an extraordinary one.

Furthermore, the ownership chain may reveal whether the painting was acquired directly from the artist, from the artist's family, or from the sitter (the person depicted in the painting) or has uninterrupted information on the descent of the painting through the generations. All of the aforementioned factors indicate the painting's authenticity. Obviously, falsified provenance can never be excluded although one expects that auction houses carefully verify provenance information in order to avoid lawsuits and loss of reputation from the auction of forged paintings. We also check whether paintings were sold throughout each work's history by one or

**Table 1.** Descriptive Statistics on Provenance

	N	Mean	Standard deviation	Minimum	Maximum
<b>Pedigree</b>					
Pedigree (Text Length)	256,560	103.99	152.79	0	9,034
<b>Owned by</b>					
Prominent Collectors	256,560	2.07%	14.20%	0	1
Royalty / Nobility	256,560	2.39%	15.30%	0	1
Wealthy Families	256,560	0.68%	8.22%	0	1
CEOs	256,560	0.06%	2.45%	0	1
Influential People ( <i>Time</i> 100)	256,560	0.08%	2.89%	0	1
Celebrities	256,560	0.20%	4.46%	0	1
Famous Sportspeople	256,560	0.12%	3.52%	0	1
Corporate Collection	256,560	0.16%	4.01%	0	1
Private Collection (Anonymous)	256,560	26.00%	43.80%	0	1
<b>Descent</b>					
Directly from Artist	256,560	10.51%	30.67%	0	1
From Artist's Family	256,560	5.29%	22.39%	0	1
From Sitter	256,560	0.33%	5.76%	0	1
Other Descent Information	256,560	9.81%	29.75%	0	1
<b>Past Sales Channel</b>					
Sold by Sotheby's or Christie's	256,560	14.96%	35.67%	0	1
Sold by Bonhams or Phillips	256,560	0.90%	9.42%	0	1
Sold by Historic Auction Houses	256,560	1.33%	11.46%	0	1
Sold by Other Important Auction Houses	256,560	1.30%	11.30%	0	1
Sold by Prominent Dealers	256,560	5.75%	23.29%	0	1
<b>Other Collections</b>					
Other Pedigree Information	256,560	42.48%	49.43%	0	1
<b>Number Count by Painting</b>					
Prominent Collectors (Number Count)	256,560	0.0213	0.1690	0	8
Descent (Number Count)	256,560	0.1090	0.3610	0	11
Sold by Sotheby's or Christie's (Number Count)	256,560	0.1780	0.4660	0	12
Sold by Bonhams or Phillips (Number Count)	256,560	0.0092	0.0983	0	5
Sold by Other Important Auction Houses (Number Count)	256,560	0.0137	0.1230	0	3
Sold by Historic Auction Houses (Number Count)	256,560	0.0150	0.1370	0	5
Sold by Prominent Dealers (Number Count)	256,560	0.0624	0.2640	0	6
<b>Exhibition</b>					
Exhibition (Text Length)	67,713	208.87	283.31	1	6,828
Exhibition (Number Count)	67,713	1.9760	2.6480	0	46
Prominent Exhibition	67,713	6.20%	24.10%	0	1
Prominent Art Fair	67,713	0.39%	6.20%	0	1
Prominent Museum	67,713	17.20%	37.80%	0	1
Other Museum	67,713	29.90%	45.80%	0	1
Cultural City	67,713	74.10%	43.80%	0	1
Gallery Exhibition	67,713	14.90%	35.60%	0	1
<b>Number Count by Painting</b>					
Prominent Exhibition (Number Count)	67,713	0.0708	0.3050	0	18
Prominent Art Fair (Number Count)	67,713	0.0041	0.0674	0	4
Prominent Museum (Number Count)	67,713	0.2730	0.7860	0	20
Other Museum (Number Count)	67,713	0.2992	0.4579	0	1
Cultural City (Number Count)	67,713	1.5740	2.1210	0	41
<b>Literature</b>					
Literature (Text Length)	72,906	241.79	388.37	0	22,413
Literature (Number Count)	72,906	1.5300	2.3970	0	150
Catalogue Raisonné	72,906	15.70%	36.40%	0	1
Cover Page	72,906	1.66%	12.80%	0	1
Illustration	72,906	45.90%	49.80%	0	1
Authoritative Press	72,906	1.15%	10.60%	0	1
Other Literature	72,906	48.10%	50.00%	0	1
<b>Number Count by Painting</b>					
Catalogue Raisonné (Number Count)	72,906	0.1690	0.4100	0	6
Cover Page (Number Count)	72,906	0.0181	0.1490	0	6
Illustration (Number Count)	72,906	0.8450	1.6170	0	89
Authoritative Press (Number Count)	72,906	0.0122	0.1180	0	4
<b>Certification</b>					
Certification (Text Length)	70,556	66.90	63.35	6	4,101

**Table 1.** (Continued)

	N	Mean	Standard deviation	Minimum	Maximum
Certification by					
Artist (Physical)	70,556	31.70%	46.50%	0	1
Artist Family (Physical)	70,556	6.16%	24.00%	0	1
Association (Physical)	70,556	15.10%	35.80%	0	1
Expert (Physical)	70,556	2.72%	16.30%	0	1
Other People (Physical)	70,556	27.60%	44.70%	0	1
Artist (Nonphysical)	70,556	5.31%	22.40%	0	1
Artist's Family (Nonphysical)	70,556	2.20%	14.70%	0	1
Association (Nonphysical)	70,556	4.45%	20.60%	0	1
Expert (Nonphysical)	70,556	2.88%	16.70%	0	1
Other People (Nonphysical)	70,556	6.15%	24.00%	0	1

*Notes.* This table presents the descriptive statistics of the provenance variables. *Pedigree (Text Length)* stands for the number of characters. *Prominent Collectors, Royalty/Nobility, Wealthy Families, CEOs, Influential People (Time 100), Celebrities, and Famous Sportspeople* are indicator variables equal one if the artwork has been in the collections of those types of collectors, respectively. *Corporate Collection* and *Private Collection (Anonymous)* equal one if the artwork was at one point part of corporate and private collections, respectively. *Directly from Artist, From Artist Family, and From Sitter* equal one if the paintings were acquired directly from the preceding respective categories. *Descent* equals one if the artwork's pedigree information contains any descendance information (not included in the preceding categories). *Sold by Sotheby's or Christie's, Sold by Bonhams or Phillips, Sold by Historic Auction Houses, Sold by Other Important Auction Houses, and Sold by Prominent Dealers* equal one if the artwork was once sold via the respective channels. *Other Pedigree Information* equals one if the artworks have other unclassified pedigree information. *Exhibition (Text Length)* is the number of characters of exhibition information, and *Exhibition (Number Count)* is the number of past exhibitions by the painting. *Prominent Exhibition, Prominent Art Fair, Prominent Museum, Other Museum, Cultural City, and Gallery Exhibition* equal one if the painting was at one point exhibited in the preceding types of exhibitions/museums/fairs/cities, respectively. *Literature (Text Length)* is the number of characters on literature information, and *Literature (number count)* is the number of times that the painting is referred to in the art literature. *Catalogue Raisonné, Cover Page* (of an art history book), *Illustration* (in an art history book), and *Authoritative Press* equal one if the artwork was illustrated in the preceding ways, respectively. *Other Literature* equals one if the catalog information refers to other literature information not included in the preceding categories. *Certification (Text Length)* is the number of characters related to certification. *Artist (Physical), Artist Family (Physical), artists' Association (Physical), Expert (Physical), and Other People (Physical)* equal one if the painting has a physical certification (e.g., "photo certificate of authenticity by artist") issued by the preceding sources, respectively. *Artist (Nonphysical), Artist Family (Nonphysical), Association (Nonphysical), Expert (Nonphysical), and Other People (Nonphysical)* equal one if the painting has nonphysical certification (e.g., "the authenticity was orally confirmed by Paul Vogt, Essen") issued by the preceding sources, respectively. Variables with "Number Count" are count variables. For each variable, we report the number of observations (N), the conditional mean, the standard deviation, the minimum, and the maximum.

more prominent auction houses or established dealers as it is likely that they more carefully collect and verify the provenance. In this respect, we consider both historic auction houses and dealers that no longer exist following mergers or termination as well as contemporary ones.

To apply textual analysis in the pedigree dimension, we develop a name list based on more than 150 databases.<sup>4</sup> The variables incorporated in the *Pedigree* are as follows:

- Past ownership: *Prominent Collectors, Royalty/Nobility, Wealthy Families, CEOs, Influential People (Time 100), Celebrities, Famous Sportspeople, Corporate Collection, and Private Collection (Anonymous)*, all from around the world. For instance, for *Royalty and Nobility*, we search for nobility titles in seven languages (English, Latin, Dutch, French, German, Italian, and Spanish). For the sportspeople, we collect the names of the best-paid ones as well as world champions and superstars (in boxing, golf, basketball, tennis, soccer, football, baseball, racing, motorcycle, cricket, track, auto racing, mixed martial arts, motorsport, and hockey).

- Descent: Purchased *Directly from Artist, From Artist's Family, or From Sitter*.<sup>7</sup> We also report *Other Descent Information*.<sup>8</sup>

- Past sale channel: *Sold by Sotheby's/Christie's, Sold by Bonhams/Phillips, Sold by Historic Auction Houses* (e.g.,

*Dorotheum, Dowell's, Hôtel Drouot, Sold by Other Important Auction Houses,*<sup>11</sup> and *Sold by Prominent Dealers*.<sup>12</sup>

- Other collections: *Other Pedigree Information* (indicates that other pedigree information is available not falling in any of the aforementioned categories).

Our first provenance dimension is *Pedigree*. For 256,560 paintings, pedigree information is available in auction catalogs. The average pedigree text length amounts to 104 characters. Of all the paintings with pedigree information, 2.1% were once part of a prominent collection, 2.4% were held by royal or noble collectors, 0.7% by (other) wealthy families, 0.06% by collectors who are also CEOs, 0.08% by influential people, and 0.3% by celebrities and sportspeople.<sup>13</sup> The credibility of the authenticity increases when the ownership can be traced back to the artist or people close to the origin. We document that there is evidence for 10.5% of the paintings that an earlier owner purchased the painting directly from the artist, 5.3% of the paintings were acquired from the artist's family, and 0.33% from the sitter. For 9.8% of artworks, the pedigree text gives additional information about ownership of descendants (not included in the aforementioned categories). When studying previous sales records, about 15.0% of the paintings with pedigree information were sold by Sotheby's and Christie's at one point in the painting's history, 0.9% by Bonhams and Phillips, 1.3%

by historically important auction houses, and 5.8% by prominent dealers.

Our second provenance dimension is labeled *Exhibition*, which embeds information about the number and importance of exhibitions (by museums, at art fairs, in museums, by galleries, and in culturally important cities). Past exhibitions may vet the painting because an exhibited painting is then often examined by experts and curators, who reflect in an exhibition catalog on the position of the painting within the total oeuvre of an artist or within an artistic school or era. Therefore, an often-exhibited painting may yield a premium at auctions. Our sample comprises 67,713 paintings with exhibition information. On average, a painting was exhibited twice. Among all observations with exhibition information, about 6.2% were exhibited at least once at prominent exhibitions, 0.4% at prominent art fairs, 17.2% in prominent museums, 29.9% at lesser known museums, 74.1% in cultural cities, and 14.9% in galleries.<sup>14</sup>

Our third provenance dimension is *Literature*. We consider whether artworks are included in catalogues raisonnés, which offer a comprehensive listing of all known artworks by the artists; are illustrated on the cover page of art books; or are included in art books published by an authoritative press (e.g., a university press)<sup>15</sup> or in any other publication. Our sample comprises 72,906 paintings with literature information with an average text length of 242 characters and, on average, 1.5 literature-related references of the painting. Among all paintings with literature information, about 15.7% are mentioned in the catalogue raisonné; 1.7% are even on the cover page of art books, 45.9% are illustrated in art books, and 1.2% are mentioned in art books published by an authoritative press.

In the fourth dimension, *Certification*, we search for two aspects: (i) the person or agency who has issued the certification (this can be artists themselves, their family members, associations,<sup>16</sup> experts, or other parties) and (ii) the form of the certification (physical certificate<sup>17</sup> versus nonphysical confirmation, for example, an oral statement by the artist about the painting's authenticity).<sup>18</sup> For 70,556 paintings, certification information is mentioned in the provenance text. Among the paintings with certification information, 31.7% have physical certification issued by the artist, 6.2% by the artist's family, 15.1% by the artist's association, 2.7% by experts, and 27.6% by other parties. In addition, about 5.3% of the observations are presented with nonphysical certification by the artist, 2.2% by the artist's family, 4.5% by their association, 2.9% by experts, and 6.2% by other parties. The general *Certification* dummy has a correlation close to zero with the other main provenance dimensions (*Pedigree*, *Exhibition*, and *Literature*), which, in turn, exhibit moderate positive correlations between 0.35 and 0.45. Delving deeper into the more detailed variables within each of the four main dimensions, we

find very low correlations, which suggests that all the detailed variables together need to be considered to obtain a reliable picture of a painting's authenticity.

## 2.2. Methodology

**2.2.1. Hedonic Linear Probability Regression.** To investigate the provenance effects, we estimate linear probability of being sold regressions with the dependent variable  $Sold_{i,t}$  indicating whether an art object was sold (versus bought in), controlling for a wide range of hedonic characteristics:

$$\begin{aligned} Sold_{i,t} = & \alpha_{i,t} + \sum_{p=1}^P \beta_p Pedigree_{p,i,t} + \sum_{l=1}^L \beta_l Literature_{l,i,t} \\ & + \sum_{e=1}^E \beta_e Exhibition_{e,i,t} + \sum_{c=1}^C \beta_c Certification_{c,i,t} \\ & + \sum_{m=1}^M \beta_m X_{m,i,t} + \sum_{t=1}^T \gamma_t D_{i,t} + \varepsilon_{i,t}, \end{aligned} \quad (1)$$

where  $Sold_{i,t}$  equals one if art object  $i$  at time  $t$  is sold and zero when bought in. The four provenance dimensions are represented by  $Pedigree_{p,i,t}$  (with characteristic  $p$  of item  $i$  at time  $t$ ),  $Literature_{l,i,t}$  (with literature characteristic  $l$ ),  $Exhibition_{e,i,t}$  (with exhibition characteristic  $e$ ), and  $Certification_{c,i,t}$  (with certification characteristic  $c$ ). The provenance dimensions are presented as dummy variables capturing whether the artwork's catalog comprises any information on these dimensions or the natural log of the number of characters used in the provenance text for each of these dimensions.  $X_{m,i,t}$  is the value of characteristic  $m$  of item  $i$  at time  $t$ .  $D_{i,t}$  is a time indicator variable for the year in which the art object is offered for sale in an auction. The coefficients,  $\beta_p$ ,  $\beta_l$ ,  $\beta_e$ , and  $\beta_c$  capture the relationships of the provenance dimensions for the probability of being sold. The coefficients  $\beta_m$  reflect the shadow price of each of the  $m$  characteristics, and the coefficients  $\gamma_t$  reflect the time trend, which can be used to construct an art price index.

**2.2.2. Hedonic Pricing Regression.** To measure the impact of provenance information on hammer prices, we resort to a hedonic pricing model, which has the advantage (relative to a repeat sales approach) of including all observed transactions.<sup>19</sup> We regress the natural logarithms of hammer prices (in 2007 real US\$) on provenance, controlling for a broad range of hedonic characteristics:

$$\begin{aligned} Ln(P_{i,t}) = & \alpha_{i,t} + \sum_{p=1}^P \beta_p Pedigree_{p,i,t} + \sum_{l=1}^L \beta_l Literature_{l,i,t} \\ & + \sum_{e=1}^E \beta_e Exhibition_{e,i,t} + \sum_{c=1}^C \beta_c Certification_{c,i,t} \\ & + \sum_{m=1}^M \beta_m X_{m,i,t} + \sum_{t=1}^T \gamma_t D_{i,t} + \varepsilon_{i,t}, \end{aligned} \quad (2)$$

where  $P_{i,t}$  represents the hammer price of art object  $i$  at time  $t$ , and the rest of the control variables are defined in Section 2.1.

**2.2.3. Repeat Sales Regression.** As we also intend to study art returns, we turn to a repeat sales analysis. Here, we analyze whether incremental changes of provenance occurring between two sales impact returns. The dependent variable  $Return_{i,(1,2)}$  in Equation (3) is the geometrically annualized return of the painting in the repeat sales transaction. We regress this return on (i) the changes in the provenance variables between the first and second transactions,  $\Delta Provenance_{a,i,(1-\delta,2-\delta)}$  (with provenance dimension  $a$  of item  $i$  between two transactions;  $a$  represents the various characteristics of each of the four provenance dimensions, *Pedigree*, *Exhibition*, *Literature*, and *Certification*), also controlling for the initial provenance predating the first sale  $Provenance_{a,i,1-\delta}$  (with provenance characteristic  $a$  of item  $i$  at time  $1-\delta$ ); (ii) possible changes in sales channel (changes in auction house (branch)) (i.e., *Auction House Upgrade* is a dummy variable equal to one if the second sale auction house has a better reputation than the first one);<sup>20</sup> and (iii) the other hedonic variables and fixed effects as defined in Section 2.1. In Equation (3), we include the same hedonic control variables  $X_{m,i,2}$  and time control variables  $D_{i,2}$  as in Equations (1) and (2). The subscripts 1, 2, (1,2), and  $\delta$  refer to the time of the first sale, second sale, holding period, and time lag, respectively. The coefficients of interest are  $\beta_a$ , which reflect the incremental provenance effects on the returns:

$$\begin{aligned}
 Return_{i,(1,2)} = & \alpha_{i,(1,2)} + \sum_{a=1}^A \beta_a \Delta Provenance_{a,i,(1-\delta,2-\delta)} \\
 & + \sum_{a=1}^A \beta_b Provenance_{a,i,1-\delta} \\
 & + \beta_u Auction\ House\ Upgrade_{i,(1,2)} \\
 & + \sum_{m=1}^M \beta_m X_{m,i,2} + \sum_{t=1}^T \gamma_t D_{i,2} + \varepsilon_{i,(1,2)}. \quad (3)
 \end{aligned}$$

The provenance information (auction catalog) is available 4–6 weeks prior to the auction, which is highlighted by means of  $\delta$  in Equation (3).<sup>21</sup>

### 3. Empirical Results

This section discusses the empirical results of the correlation between provenance information on the one hand and the probability of being sold, the hammer price (for our full sample), and the returns (for the repeat sample) on the other hand. We also provide robustness tests on subsamples and perform analyses by means of LASSO estimations to alleviate estimation concerns induced by high-dimensional fixed effects.

#### 3.1. Provenance Effects and Probability of Being Sold

Is the probability that a painting offered for sale is actually sold (versus bought in) correlated with the provision of provenance? As the auction house provides a price estimate in the catalog, which can affect the sales outcome, we include the lower estimate (which is assumed to be at or close to the reserve price) in the linear probability (Models 1 and 2 of panel A of Table 2). A low-price estimate that is set high increases the probability that the painting is not sold (which occurs when the highest bid does not exceed the reserve price). We also control for a large set of hedonic variables detailed in Section 2.1.1 as well as for the following fixed effects: artist, year, month, and auction house branch level.<sup>22</sup> Our sample comprises 1,707,136 observations with full hedonic information available as well as low price estimates.

The aggregated provenance dimensions *Pedigree*, *Exhibition*, *Literature*, and *Certification* are represented by either (i) an indicator variable capturing the provision of this type of information or (ii) the natural logarithm of the catalog’s text length of each of these dimensions. The provision of pedigree information goes hand in hand with a higher probability of being sold, which increases by 1.7% (column (1) in panel A of Table 2). Likewise, exhibition history and the appearance of a painting in the art (history) literature increase the probability of being sold by 3.8% and 2.5%, respectively. By contrast, the presence of a certificate does not seem to affect the probability of being sold (but affects prices; see Section 3.2). As expected, the negative coefficient of the low price estimate indicates that artworks with higher reserve prices are less likely to be sold. The specification that uses the provenance dimensions’ text length (column (2) in panel A of Table 2) shows consistent and similar patterns with the exception of a positive correlation for certification. It is important to note that these regressions should not be interpreted as causal relations. In fact, endogeneity (in the form of reverse causality) could very well be substantial because auction houses offer provenance information, especially for paintings with ex ante high probability of being sold or with the highest price potential (which can partially but not entirely be addressed by the inclusion of the price estimates). We undertake a set of endogeneity tests in Section 3.3 but first undertake a correlation analysis based on a more granular approach to the provenance information with Models 1 and 2 in panel B of Table 2.

In the pedigree dimension, past ownership by prominent collectors, royalty/nobility, wealthy families, celebrities, and famous sportspeople are all correlated to increase the probability of being sold by 5.2%, 6.6%, 8.6%, 3.7%, and 9.5%, respectively (column (1) in panel B of Table 2). If the painting was in the past part of a corporate collection, the current probability of being sold is 9.8% higher.

**Table 2.** Provenance Effects on Probability of Being Sold and Hammer Price

Panel A: Provenance effects on probability of being sold and hammer price					
Dependent variable	Sold[0,1]		Ln(Price)		
	(1) Indicator	(2) Text length	(3) Indicator	(4) Price impact, %	(5) Text length
Provenance as					
Pedigree	0.0173*** (0.0062)	0.0044*** (0.0014)	0.1885*** (0.0170)	20.74	0.0531*** (0.0045)
Exhibition	0.0379*** (0.0041)	0.0079*** (0.0008)	0.3499*** (0.0208)	41.89	0.0734*** (0.0041)
Literature	0.0250*** (0.0063)	0.0056*** (0.0011)	0.4288*** (0.0336)	53.54	0.0869*** (0.0063)
Certification	0.0102 (0.0133)	0.0115*** (0.0033)	0.1297*** (0.0226)	13.85	0.0614*** (0.0100)
Ln(Low Price Estimate)	-0.0661*** (0.0036)	-0.0668*** (0.0036)			
Artist fixed effects	Yes	Yes	Yes		Yes
Year fixed effects	Yes	Yes	Yes		Yes
Month fixed effects	Yes	Yes	Yes		Yes
Auction house branch fixed effects	Yes	Yes	Yes		Yes
Hedonic controls	Yes	Yes	Yes		Yes
Number of observations	1,707,136	1,707,136	1,111,220		1,111,220
R <sup>2</sup>	0.1750	0.1752	0.7805		0.7817
Panel B: Detailed provenance effects on probability of being sold and hammer price					
Dependent variable	Sold[0,1]		Ln(Price)		
	(1) Indicator	(2) Indicator	(3) Price Impact, %	(4) Price Impact, %	
Ln(Low Price Estimate)	-0.0665*** (0.0036)				
<b>Pedigree</b>					
Past Ownership					
Prominent Collectors	0.0524*** (0.0113)	0.2183*** (0.0405)	24.40		
Royalty / Nobility	0.0656*** (0.0148)	0.2724*** (0.0262)	31.31		
Wealthy Families	0.0855*** (0.0232)	0.3538*** (0.0479)	42.45		
CEOs	-0.0101 (0.0343)	0.1479 (0.1021)	15.94		
Influential People ( <i>Time</i> 100 list)	-0.0167 (0.0372)	0.0841 (0.0842)	8.77		
Celebrities	0.0372** (0.0170)	0.1288 (0.1308)	13.75		
Famous Sportspeople	0.0948*** (0.0289)	0.4062*** (0.1358)	50.11		
Corporate Collection	0.0976*** (0.0376)	0.1038** (0.0482)	10.94		
Private Collection (Anonymus)	-0.0051 (0.0084)	0.1793*** (0.0222)	19.64		
Descent					
Directly from Artist	0.0145*** (0.0056)	0.1296*** (0.0142)	13.84		
From Artist's Family	-0.0027 (0.0049)	0.0070 (0.0239)	0.70		
From Sitter	-0.0892*** (0.0185)	0.1068** (0.0515)	11.27		
Other Descent Information	0.0288*** (0.0045)	0.2121*** (0.0190)	23.63		
Past Sales Channel					
Sold by Sotheby's or Christie's	-0.0060 (0.0042)	0.2121*** (0.0243)	23.63		

**Table 2.** (Continued)

Panel B: Detailed provenance effects on probability of being sold and hammer price			
Dependent variable	Sold[0,1]	Ln(Price)	
	(1)	(2)	(3)
Provenance as	Indicator	Indicator	Price Impact, %
Sold by Bonhams or Phillips	-0.0277* (0.0147)	0.0687* (0.0387)	7.11
Sold by Historic Auction Houses	-0.0164** (0.0072)	0.0957*** (0.0363)	10.04
Sold by Other Important Auction Houses	-0.0133 (0.0083)	-0.0463 (0.0393)	-4.52
Sold by Prominent Dealers	0.0377*** (0.0064)	0.2842*** (0.0378)	32.87
<b>Other Collections</b>			
Other Pedigree Information	0.0274*** (0.0058)	0.1323*** (0.0133)	14.15
<b>Exhibition</b>			
Prominent Exhibition	0.0263*** (0.0081)	0.2327*** (0.0247)	26.20
Prominent Art Fair	0.0015 (0.0298)	-0.0055 (0.0648)	-0.55
Prominent Museum	0.0543*** (0.0083)	0.4631*** (0.0331)	58.90
Other Museum	0.0176*** (0.0042)	0.1878*** (0.0147)	20.66
Cultural City	0.0192*** (0.0034)	0.2165*** (0.0147)	24.17
Gallery Exhibition	0.0430*** (0.0077)	0.2535*** (0.0220)	28.85
<b>Literature</b>			
Catalogue Raisonné	0.0236** (0.0095)	0.3056*** (0.0448)	35.74
Cover Page	0.0509*** (0.0143)	0.4237*** (0.0513)	52.76
Illustration	0.0207*** (0.0079)	0.3660*** (0.0366)	44.20
Authoritative Press	0.0130 (0.0183)	0.3450*** (0.0876)	41.20
Other Literature	0.0149* (0.0081)	0.3186*** (0.0252)	37.52
<b>Certification</b>			
Artist (Physical)	0.0599*** (0.0172)	0.0828*** (0.0262)	8.63
Artist's Family (Physical)	-0.0166 (0.0152)	0.0674* (0.0385)	6.97
Association (Physical)	0.0348** (0.0148)	0.1199*** (0.0410)	12.74
Expert (Physical)	-0.0169 (0.0179)	0.3416*** (0.0517)	40.72
Other People (Physical)	0.0011 (0.0129)	0.1148*** (0.0302)	12.16
Artist (Nonphysical)	0.0348 (0.0224)	0.2810*** (0.0287)	32.45
Artist's Family (Nonphysical)	-0.0312* (0.0171)	0.0039 (0.0597)	0.39
Association (Nonphysical)	-0.0381 (0.0240)	0.1889*** (0.0352)	20.79
Expert (Nonphysical)	-0.0292 (0.0211)	0.1654*** (0.0482)	17.99
Other People (Nonphysical)	-0.0206 (0.0136)	0.1340*** (0.0393)	14.34
Artist fixed effects	Yes	Yes	

**Table 2.** (Continued)

Dependent variable	Panel B: Detailed provenance effects on probability of being sold and hammer price		
	Sold[0,1]	Ln(Price)	
	(1) Indicator	(2) Indicator	(3) Price Impact, %
Provenance as			
Year fixed effects	Yes	Yes	
Month fixed effects	Yes	Yes	
Auction house branch fixed effects	Yes	Yes	
Hedonic controls	Yes	Yes	
Number of observations	1,707,136	1,111,220	
R <sup>2</sup>	0.1755	0.7819	

*Notes.* This table presents the relation between Provenance and Probability of Being Sold and Hammer Price. The dependent variables are (1) *Sold*[0,1], which takes one if the painting is successfully sold (and zero when bought in) and (2) *Ln(Price)*, the natural logarithm of the real hammer price in real (2007) US\$. In panel a, columns (1) and (3), *Pedigree*, *Exhibition*, *Literature*, and *Certification* are dummy variables capturing if the artwork's catalog comprises any information on these respective dimensions. In panel A, columns (2) and (5), these four provenance variables stand for the natural log of the number of characters used in the provenance text for each of these respective dimensions. Column (4) in panel A reports the corresponding price impact (relative price change) of column (3) calculated as  $\exp(\text{estimated coefficient}) - 1$ . Panel B has a similar setup, but the independent variables are the detailed elements that constitute *Pedigree*, *Exhibition*, *Literature*, and *Certification* are included (for definitions, see Appendix A). In both panels, *Ln(Low Price Estimate)*, the log of the low price estimate (in real US\$) is a proxy for the reserve price. All regressions in both panels include hedonic controls (detailed in Section 2.1.1), and artist, year, month, and auction house branch level (AH) fixed effects. Standard errors (S.E.) are reported in parentheses and clustered at the auction branch level.

\*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

With regard to information on descent, when a painting was purchased directly from the artist, the probability of being sold increases by 1.5%.<sup>23</sup> We also examine whether the reputation of past sales channels has an impact on current sales: if the painting was sold by a prominent dealer, the current probability of being sold is 3.8% higher, but when past sales were at one point executed by Sotheby's or Christie's, there is no positive effect on current probability of being sold (possibly because those paintings may at the time already have been sold at higher price levels). A past auction at Bonhams and Phillips (or other historically important auction houses) does not have any positive impact on prices.

When a painting was part of prominent exhibitions, it is sold more easily (the probability increases by 2.6%; column (1) of panel B in Table 2). If a painting was exhibited in museums in the past, it is now sold more easily; exhibitions in prominent museums increase this probability by 5.4% and 1.8% in the case of less prestigious museums. If the artwork was exhibited in cultural cities where possibly a more cultured audience of art lovers can be reached, we note an augmentation in the probability of being sold (by 1.9%). Past exhibitions by galleries facilitate a sale (by 4.3%), but displaying the painting at prominent art fairs does not seem to matter.

When a painting is mentioned in the art (history) literature, selling is expectedly easier: a mention in a catalogue raisonné, depiction on the cover page of art books, and inclusion of an illustration in an art book augments the probability of being sold by 2.4%, 5.1%, and 2.1%, respectively. Physical certification by artists or their associations or foundations (e.g., the Andy Warhol Foundation and

Keith Haring Foundation) is important and affects sales: the physical certificate issued by the artist increases the probability (by 6.0%) as does certification by an artist's association (by 3.5%). Interestingly, if the painting is accompanied by nonphysical certification (by the artist's family, experts, etc.) the painting does not sell more easily, which suggests that nonphysical certification does not remove possible doubts regarding authenticity.

### 3.2. Provenance Effects and Hammer Prices

We turn to the relationship between provenance and hammer prices in panels A and B of Table 2 in the same vein as for the probability of being sold. We first estimate Equation (2) for 1,111,220 auction transactions for which we have complete information on all the hedonic characteristics presented in Section 2.1.<sup>24</sup> The provenance variables in columns (3) and (5) of panel A of Table 2 are dummy and textual length variables (as defined in Section 3.1), respectively. After controlling for all the traditional hedonic variables and fixed effects, we find that the presence of provenance information goes hand in hand with higher price levels (column (3)).<sup>25</sup> If information is made available on the painting's pedigree, the price is 20.7% higher ( $\exp(0.1885) - 1$ ); with an exhibition history, the price augments by 41.9%; literature drives the price up by 53.5%; and with certification, the price increases by 13.9%. Thus, art (history) literature sources and an exhibition history affect prices most. Models including the provenance text length yield qualitatively similar results (column (5) in panel A). As mentioned in Section 3.1, the caveat on endogeneity applies, which we address in Section 3.3.

We dissect the four main dimensions of provenance into their detailed constituting elements and show the results in panel B of Table 2. For the pedigree components, we observe that past ownership by prominent collectors, royalty/nobility, wealthy families, and famous sportspeople has a strong impact on current price levels. Paintings reach prices that are, on average, 24.4% (31.3%) higher if they were held by prominent collectors (royalty/nobility). Past ownership by wealthy families or famous sportspeople correlates with current price premiums of 42.5% and 50.1%, respectively. When the auction catalog reports that the painting was at one point purchased directly from the artist, the current hammer price is 13.8% higher; a past purchase directly from a sitter now augments the price by 11.3%; and any descent information (not included in the preceding categories) increases prices by 23.6%. Thus, information related to the past ownership chain (purchases from the artist or his family), prominent collectors, or famous owners enhances the marketability of paintings and their hammer prices.

In addition, past sales channels also affect auction prices: the fact that a past auction sale was made by Sotheby's or Christie's, Bonhams or Phillips, or by historically important auction houses (e.g., Dorotheum or Hôtel Drouot) may be perceived as (past) recognition of quality and authenticity, which is mirrored in prices that are higher by 23.6%, 7.1%, and 10.0%, respectively. The current price premium may, thus, reflect the (past) notoriety of these auction houses. Similarly, past sales by prominent dealers are correlated with auction prices that are, on average, 32.9% higher.

Details about the second pillar—past exhibitions are also strongly related to price levels as they may reflect the intrinsic quality of paintings. For instance, if the painting had been included in prominent exhibitions, the auction price is 26.2% higher. Likewise, exhibitions organized by prominent museums or in important cultural cities enhance a painting's exposure to a larger art-loving audience and are related to higher prices (by 58.9% and 24.2%, respectively).

Likewise, it is important that a painting is named or depicted in the art literature: all aspects within the *Literature* dimension correlate with prices: the inclusion of the painting in the catalogue raisonné, being depicted on the cover page of or in art books, discussed in books published by an authoritative (university) press are related to higher prices (with significant price premiums of 35.7%, 52.8%, 44.2%, and 41.2%, respectively).

Certification also enhances trust: panel B of Table 2 shows that all physical and nonphysical (testimonial) certification is positively correlated with prices (column (2)). Certificates by experts have the biggest price impact (40.7%). In the case of a nonphysical certification by the artist (e.g., when the provenance text refers to a

testimonial of an oral statement by the artist), the price impact amounts to 32.5%.

### 3.3. Endogeneity

A first endogeneity concern may emerge when the provision of provenance information is driven by past prices of a painting (or of similar paintings). For a painting expected to attract a high price, more resources can be made available to research and document its provenance. To rule out (or at least attenuate the possibility of) the endogeneity concerns, we perform the following four analyses. (i) In a DiD setting, we examine the provenance effects on the artworks made by artists whose works were faked or forged (Section 3.3.1). (ii) We exploit the Christie's change in provenance policy in 2012 in a DiD setting and use the provenance policy by Sotheby's (which is regarded as having similar reputation and quality) as a control (Section 3.3.2). (iii) To address potential endogeneity induced by reverse causality between *Past Prices* (capturing higher quality) and *Provenance Changes*, we run a two-stage model on a repeat sales sample (Section 3.3.3). A second endogeneity concern relates to the possibility that sale decisions may be endogenous in that sellers may offer a painting for sale to an auction house after having observed recent prices for similar paintings (of the same artist, same school, or in general in the entire market). Therefore, we resort to (iv) a subsample analysis for auction sessions that comprise sales decisions that we expect to be made exogenously (Section 3.3.4). These are auction sales from estates of deceased previous owners. The heirs might be forced to sell to pay inheritance tax<sup>26</sup> or might not be able to afford or be disinterested in the upkeep of an estate, including art inheritance. Thus, the decision to offer art for auction in the period after the decease of the collector might not be (or be less) endogenous.

**3.3.1. Discoveries of Fakes and Forged Paintings.** Art markets offer a congenial environment for forgeries, which has a detrimental effect on the trustworthiness of the market. Large numbers of fake and forged paintings are rumored to circulate. According to a former director of The Museum of Modern Art, up to 40% of the high-end art market may consist of forged art (Thompson 2010).<sup>27</sup>

This may undermine the art world's confidence in the authenticity of paintings and, hence, distort price formation and depress value. In this respect, provenance is a prominent factor underlying the proper functioning of the art market although provenance may also be subject to forgery. The discovery of a fake painting may negatively affect the prices of the paintings by the forged artist and could shift transactions to the most trustworthy intermediaries who could offer an insurance (a guarantee to take forged paintings back). To investigate the impact of the discoveries of fakes/forgeries on the price

of artworks through the effect of the provenance dimensions, we collect the disclosure dates of fakes and forgery cases from three primary sources: the specialized art journals Artsjournal.com and Artnews.com and the general news database Factiva, which includes the worldwide print media in 28 languages. For each article that we retrieve on this topic, we collect the event date of the discovery, title of the fake or forged painting, name of the artist, name of the forger (if available), title of the original painting, and the auction house(s) involved in the case. We take the first date that the rumor, motivated suspicion, proof, or ruling by a judge was mentioned in the press as the date of discovery of fraud. We retain only the cases that were ultimately confirmed as fakes or forgeries. In this manner, we identified 54 cases of fraud related to paintings in our sample period.<sup>28</sup>

Equation (4) presents the DiD regression in which *Treated* equals one for the paintings of artists of whom one or more paintings were discovered as being faked or forged and whose paintings were offered for sale in an auction at a date after the date of the discovery. *Provenance* is a dummy variable capturing if the painting contains information on provenance (and on its dimensions). The control variables are the hedonic variables and fixed effects consistent with Equation (1):

$$\begin{aligned} \ln(\text{Price}_{i,t}) = & \alpha_{i,t} + \beta_r \text{Treated}_{i,t} + \beta_k \text{Treated}_{i,t} \\ & \times \sum_{a=1}^A \text{Provenance}_{a,i,t} + \sum_{a=1}^A \beta_a \text{Provenance}_{a,i,t} \\ & + \sum_{m=1}^M \beta_m X_{m,i,t} + \sum_{t=1}^T \gamma_t D_{i,t} + \varepsilon_{i,t}. \end{aligned} \quad (4)$$

The variable of interest is the interaction term *Treated* and *Provenance*. We expect a negative coefficient *Treated* because, after a fake is discovered for a particular artist, doubts about the authenticity of all of this artist's paintings may arise, which may translate into lower prices for their paintings without provenance information. The provision of provenance information may undo or reduce the negative effect of the discovery because, in case of mistrust, additional information as an authenticity signal is vital to restore confidence, which is why we expect a positive coefficient of the interaction term *Treated* × *Provenance*.

The results in column (1) of Table 3 reveal a sharp decline in the prices of paintings without provenance information by artists whose paintings have been forged (the parameter estimate of *Treated* is  $-0.2670$ , which represents a 23.4% price decline ( $\exp(-0.2670) - 1$ ). The provision of provenance has a significantly positive impact on the hammer price subsequent to the discoveries of fakes/forgeries (in column (1), *Treated* × *Provenance* = 0.7054), which results in a net positive price impact of 55.0% ( $\exp(-0.2670 + 0.7054) - 1$ ) for the

paintings with provenance of the affected artists. The provision of provenance, thus, more than compensates the negative impact of the discovery of a forged painting. The price increase may also reflect a supply shock as the number of authentic paintings (with strong and reliable provenance) may be scarcer.

Column (2) of Table 3 disentangles provenance into its four dimensions and depicts a similar and consistent result: there is a negative price impact on auctioned paintings of artists whose work is forged as soon as the media reports rumors or proof of forgeries and fakes. From that moment on, provenance information on pedigree, exhibitions, and literature becomes more important and undoes and even more than compensates the negative price effect of forgeries.

Because most of the artists affected by fraud in our databases were active in the 20th century (the vast majority were born between 1880 and 1940), we restrict our sample to these affected artists as well as the unaffected ones born in the same time span and, therefore, productive in the same era. We show that columns (3) and (4) of Table 3 yield very similar results.

In addition, we examine the impact of provenance information on the probability of being sold of (non-forged) paintings after the discovery of fakes and forgeries of paintings of a specific artist. We expect a positive coefficient for the interaction term *Treated* × *Provenance* as a painting by an affected artist that is offered for sale with provision of provenance information may sell more easily. We confirm that provenance has a significant positive impact on the probability of being sold (Appendix C). Higher bids arise in times of mistrust for items containing provenance information.

**3.3.2. Provenance Policy Change.** As disclosed in the opinion and order of *Waren v. Christie's Inc. 2018* court case, Christie's confirmed during the interrogatory responses that it had amended its provenance policy in 2012 and required *all* consignors to provide detailed provenance information.<sup>29</sup> This change to the Christie's disclosure policy on provenance was applied to all art types offered for auction (or private sale) at all of its branches. We treat this policy change as a quasi-natural experiment using a DiD setting by contrasting the shock's impact on Christie's sales (treated sample) with Sotheby's sales (which serves as control sample). The latter is often considered Christie's "twin" auction house as both auction houses are similar in terms of auction history, reputation, quality of auction lots offered, branch locations, international clientele, networks of sellers and buyers, and valuation expertise. Over the period 2007 to 2016, 198,076 paintings were sold through Christie's and Sotheby's, which represents 17.8% of our sample. In Table 4, the generic *Provenance* variable and its four main dimensions equal one if any information on provenance, pedigree, exhibition,

**Table 3.** Provenance Effects on Hammer Price After the Discovery of Fakes and Forgeries

Dependent variable	Full sample		Artists active in 20th century	
	(1) Ln(Price)	(2) Ln(Price)	(3) Ln(Price)	(4) Ln(Price)
<i>Treated</i>	-0.2670** (0.1168)	-0.1977* (0.1142)	-0.2747** (0.1267)	-0.2023 (0.1262)
<i>Treated</i> × <i>Provenance</i>	0.7054*** (0.0895)		0.7115*** (0.1017)	
<i>Treated</i> × <i>Pedigree</i>		0.3520*** (0.0926)		0.3459*** (0.0988)
<i>Treated</i> × <i>Exhibition</i>		0.2215** (0.0915)		0.2652*** (0.0729)
<i>Treated</i> × <i>Literature</i>		0.3529*** (0.1024)		0.3253*** (0.1245)
<i>Treated</i> × <i>Certification</i>		0.0656 (0.2314)		0.0163 (0.2626)
<i>Provenance</i>	0.2715*** (0.0201)		0.2526*** (0.0228)	
<i>Pedigree</i>		0.1863*** (0.0171)		0.1995*** (0.0210)
<i>Exhibition</i>		0.3450*** (0.0193)		0.3300*** (0.0253)
<i>Literature</i>		0.4178*** (0.0318)		0.4131*** (0.0351)
<i>Certification</i>		0.1302*** (0.0227)		0.1117*** (0.0247)
Artist fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes
Hedonic controls	Yes	Yes	Yes	Yes
Auction house branch fixed effects	Yes	Yes	Yes	Yes
Number of observations	1,111,220	1,111,220	548,684	548,684
R <sup>2</sup>	0.7773	0.7807	0.7963	0.7996

*Notes.* This table presents the DiD estimators related to the discoveries of faked or forged paintings. The dependent variable is the natural logarithm of deflated hammer price in US\$. *Treated* equals one if the auction date of a specific artist’s painting falls after the date of the discovery of fakes/forgeries for this artist. *Provenance*, *Pedigree*, *Exhibition*, *Literature*, and *Certification* are dummy variables equal to one if the painting contains any such information from the auction catalog. Full sample results are presented in columns (1) and (2), and results for the subsample of artists active in the 20th century are in columns (3) and (4). All regressions include hedonic controls (see Section 2.1.1), and artist, year, month, and auction house branch level (AH) fixed effects. Standard errors (S.E.) are reported in parentheses and clustered at the auction branch level.

\*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

literature, and certification is provided. The variables of interests are the interaction terms of *Treated* with the aforementioned provenance terms, which we expect to be positive not only because of the mere provision of provenance information (which we test), but also because of the provision of higher quality provenance (which we cannot test). We find that sales with provenance information by Christie’s since 2013 earn a DiD premium (Models 1–5 in Table 4). This suggests that the provision of provenance creates trust and is reflected in the hammer prices. In particular, Model 6 shows that information on pedigree, exhibitions, and certification affect hammer prices and the certification, which comprises the physical and logged oral authenticity confirmation, has the largest price impact. An average sale with mandatory certification reporting sold through Christie’s since 2013 enjoys a premium of

37.1% ( $\exp(0.3157) - 1$ ) compared with the average sale through Sotheby’s.

We follow Roberts and Whited (2013) by running a series of placebo tests to confirm the validity of the quasi-natural experiment on a subsample of sales sold via Christie’s and Sotheby’s during the pre-policy change period (namely, from the beginning of 2007 to the end of 2011) and set the placebo policy change in the beginning of the years 2008, 2009, 2010, and 2011. We then repeat the regressions of Table 4 and find that *Treated* interacted with *Pedigree*, *Literature*, *Exhibition*, and *Certification* are not statistically significant (not tabulated).

**3.3.3. Addressing Reverse Causality Between Past Prices and Provenance Changes.** To address potential endogeneity induced by reverse causality between *Past Prices* (capturing higher quality) and *Provenance Changes*,

**Table 4.** Christie's Provenance Policy Change

Dependent variable	(1) Ln(Price)	(2) Ln(Price)	(3) Ln(Price)	(4) Ln(Price)	(5) Ln(Price)	(6) Ln(Price)
<i>Treated</i>	−0.0165 (0.0132)	−0.0026 (0.0128)	0.0547*** (0.0102)	0.0260*** (0.0100)	0.0478*** (0.0098)	−0.0019 (0.0126)
<i>Treated</i> × <i>Provenance</i>	0.1179*** (0.0125)					
<i>Treated</i> × <i>Pedigree</i>		0.1017*** (0.0122)				0.0805*** (0.0125)
<i>Treated</i> × <i>Exhibition</i>			0.0904*** (0.0136)			0.0481*** (0.0151)
<i>Treated</i> × <i>Literature</i>				0.0753*** (0.0133)		0.0145 (0.0150)
<i>Treated</i> × <i>Certification</i>					0.3581*** (0.0805)	0.3157*** (0.0781)
<i>Provenance</i>	0.2897*** (0.0061)					
<i>Pedigree</i>		0.2574*** (0.0061)				0.1729*** (0.0061)
<i>Exhibition</i>			0.4454*** (0.0070)			0.3061*** (0.0073)
<i>Literature</i>				0.5407*** (0.0072)		0.4185*** (0.0075)
<i>Certification</i>					−0.1225*** (0.0253)	−0.0743*** (0.0246)
Artist fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Hedonic controls	Yes	Yes	Yes	Yes	Yes	Yes
Auction house branch fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	198,076	198,076	198,076	198,076	198,076	198,076
R <sup>2</sup>	0.7715	0.7708	0.7745	0.7765	0.7677	0.7812

Notes. This table presents the DiD results exploiting the Christie's Provenance Policy Shock in 2012. The dependent variable is  $\ln(\text{Price})$ , the natural logarithm of deflated hammer price in US\$. The sample consists of all sales by Christie's and Sotheby's. *Treated* is a dummy that equals one when a sale takes place through Christie's since 2013. *Provenance*, *Pedigree*, *Exhibition*, *Literature*, and *Certification* equal one if the catalog provides information on provenance and its dimensions. All regressions include hedonic controls (see Section 2.1.1), and artist, year, month, and auction house branch level (AH) fixed effects. Standard errors (S.E.) are reported in parentheses.

\*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

we run a two-stage model on the repeat sales sample.<sup>30</sup> Repeat sales are identified by matching the exact artist's name, a painting's measurements (length and width), title, medium, and presence of a signature and date. After eliminating paintings from the same artists and of about the same size but with indiscriminate titles ("Landscape," "View of the Sea," "Portrait of a Lady"), we obtain 6,647 repeat sales pairs, which we are certain are real repeat sale transactions.

First, we regress the changes in provenance,  $\Delta \text{Provenance}_{a,i,(1-\delta,2-\delta)}$  ( $a$  representing four provenance dimensions: *Pedigree*, *Exhibition*, *Literature*, and *Certification*) on the past price of the painting ( $\ln(\text{Price}_{i,1})$ ), the log of the first sale's hammer price in deflated US\$. The control variables are *Time Span* (the number of days between the two sales), the initial provenance predating the first sale  $\text{Provenance}_{a,i,1-\delta}$  (with provenance characteristic  $a$  of item  $i$  at time  $1 - \delta$ ), and all the hedonic control variables included in Equation (1) (e.g., *Artist*, *Year*, *Month*, and *Auction House Branch Level* (AH) fixed effects). The subscripts 1, 2, (1,2), and  $\delta$  refer to the time of the first sale,

second sale, holding period, and a short time lag of 4–6 weeks, respectively, which are consistent with Equation (3). Hence, we estimate

$$\begin{aligned} \Delta \text{Provenance}_{a,i,(1-\delta,2-\delta)} &= \alpha_{i,(1,2)} + \beta_x \ln(\text{Price}_{i,1}) \\ &+ \beta_s \text{Time Span}_{i,(1,2)} \\ &+ \sum_{a=1}^A \beta_a \text{Provenance}_{a,i,1-\delta} \\ &+ \sum_{m=1}^M \beta_m X_{m,i,t} + \sum_{t=1}^T \gamma_t D_{i,t} \\ &+ \epsilon_{i,(1,2)}. \end{aligned} \quad (5)$$

In a second model, we regress the dependent variable  $\ln(\text{Price}_2)$  of the second sale on the residuals from provenance dimensions  $\epsilon(\Delta \text{Provenance})$ , which represents  $\epsilon(\Delta \text{Pedigree})$ ,  $\epsilon(\Delta \text{Exhibition})$ ,  $\epsilon(\Delta \text{Literature})$ , and

$\varepsilon(\Delta Certification)$  of Equation (5), with the same control variables as in Equation (3):

$$\begin{aligned} \ln(\text{Price}_{i,2}) = & \alpha_{i,2} + \sum_{a=1}^A \beta_a \varepsilon(\Delta \text{Provenance}_{a,i,(1-\delta,2-\delta)}) \\ & + \beta_x \ln(\text{Price}_{i,1}) + \beta_s \text{Time Span}_{i,(1,2)} \\ & + \sum_{a=1}^A \beta_a \text{Provenance}_{a,i,1-\delta} + \sum_{m=1}^M \beta_m X_{m,i,2} \\ & + \sum_{t=1}^T \gamma_t D_{i,2} + \varepsilon_{i,2}. \end{aligned} \quad (6)$$

We report the results of the estimations of Equations (5) and (6) in panels A and B of Table 5, respectively. For the changes in the four dimensions of provenance, we use two specifications: (i) dummy variables capturing the change and (ii) changes in the textual length of this provenance dimension. In the first specification, the indicator variable equals one when new provenance information is provided over the holding period and otherwise is zero. For the changes in provenance length, we take the natural logarithms of the difference between corresponding text character length of the two sales. Panel A of Table 5 shows that the past price  $\ln(\text{Price}_1)$  has no statistically significant impact on the decision to offer more *Pedigree* information (Models 1 and 5), but a higher past price induces the provision of more information on *Exhibition*, *Literature*, and *Certification* (Models 2–4) and increases the amount of information (text length) offered on *Literature* and *Certification* in between the two sales (Models 7 and 8). However, the economic effects are minimal. When, for instance, we take  $\Delta \text{Literature}$  in Models 3 and 7, we observe that, when a painting's past price has doubled, we observe an increase of merely 2% in the probability that *Literature* information is provided in the period between the first auction transaction and the publication of the catalog of the second auction and by 11.6% in the *Literature* text length, *ceteris paribus*. As the average character length of *Literature* amounts to about 242 characters, a doubling of the past price only leads to a four-word increase in the literature information ( $242 \times 11.59\% = 27$  characters). These results attenuate reverse causality concerns as the possibility that past prices drive the effort to do new provenance research seems rather limited. It should be noted that we control for many fixed effects (including auction house branch) and also for the time span between the two transactions because a longer holding period gives a greater opportunity to generate more provenance information as well as to collect additional information).

To further alleviate concerns about reverse causality, we include in the pricing model of the second transaction the residuals from Equation (5), which capture provenance information that is not predicted by past

prices as independent variables in Equation (6) (panel B, Table 5). The unpredicted information of *Exhibition*, *Literature*, and *Certification* is significantly positively related to the price of the second sale in all four models. The specification with changes in text length show positive correlations for *Pedigree*, *Exhibition*, and *Literature*. In summary, the results from panel B imply that art prices are affected by the provenance information but not by the provenance information that is provided or augmented following high past prices in previous transactions (as the provenance information unpredicted by past prices can predict future prices).

**3.3.4. Exogenous Sales Decisions.** We explore a subsample of what we expect to be “exogenous” sales. These sales might not be affected (or at least be less affected) by past price trends as we select sales related to the four “D”s (divorce, debt, death, and disaster). To do so, we search in the auction session titles: “estate,” “property of,” “legacy,” “bequest,” “heritage,” “gift,” “endowment,” “charity” in singular and plural forms and “late” plus a person's name. An example of such an auction is titled “The property of the late M.H.D. McAlpine: Paintings, Ceramics, Silver, Works of Art and Furniture.” In this way, we obtain 37,851 paintings of which 25,904 (68.4%) were sold. Admittedly, this subsample choice cannot completely exclude endogeneity because, whereas death may be exogenous, the sales decision by the heir may still be stalled. Still, we expect that endogeneity concerns are somewhat lower for this subsample.

The results in Table 6 demonstrate that the provenance dimensions *Pedigree* and *Exhibition* are significantly positively correlated with the probability of being sold and, respectively, indicate a 3.7% and 7.9% higher probability of being sold than for paintings lacking this kind of documentation. Including the reserve price proxy (lowest estimate) shows that a high reserve makes a sale more difficult. Turning to the price regressions in Table 6, we observe that the provenance dimensions *Pedigree*, *Exhibition*, and *Literature* are significantly positively associated with a higher price level of paintings with a price impact of 18.1% ( $\exp(0.1659) - 1$ ), 51.0%, and 66.3%, respectively. By exploiting (what we expect to be more) exogenous sales decisions that may be less affected by unobservable price-trend related motivations of the sellers, we show that the impact of provenance factors on the probability of being sold and price level is upheld. These findings are consistent with our full sample results (Table 2) and show even larger economic magnitudes.

### 3.4. Provenance Effects and Returns for Repeat Sales Transactions

To investigate the relation between provenance and returns, we turn to the return model of Section 2.2.3 (Equation (3)) and the repeat sales sample used in

**Table 5.** Addressing Endogeneity in the Relationship Between Past Prices and Provenance Changes

Panel A: The impact of first sale price on the provenance changes								
Dependent	Changes in provenance provision (dummy variables)				Changes in the text length of provenance information			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\Delta$ Pedigree	$\Delta$ Exhibition	$\Delta$ Literature	$\Delta$ Certification	$\Delta$ Pedigree	$\Delta$ Exhibition	$\Delta$ Literature	$\Delta$ Certification
$\ln(\text{Price}_1)$	−0.0001 (0.0084)	0.0109* (0.0056)	0.0226*** (0.0071)	0.0098** (0.0050)	0.0407 (0.0443)	0.0479 (0.0349)	0.1159*** (0.0392)	0.0407* (0.0220)
<i>Time Span</i>	0.0001** (0.0001)	−0.0001 (0.0001)	0.0001 (0.0001)	0.0001** (0.0001)	0.0002*** (0.0001)	0.0001* (0.0001)	0.0001*** (0.0001)	0.0002** (0.0001)
<i>Pedigree</i> <sub>1</sub>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>Exhibition</i> <sub>1</sub>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>Literature</i> <sub>1</sub>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>Certification</i> <sub>1</sub>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Artist fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hedonic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Auction house branch fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	6,612	6,612	6,612	6,612	6,612	6,612	6,612	6,612
$R^2$	0.6578	0.4311	0.3773	0.4291	0.6327	0.3286	0.3267	0.3764

  

Panel B: Second sale price and unpredicted provenance changes				
Dependent variable	Changes in provenance (dummies)		Changes in text length	
	(1)	(2)	(3)	(4)
	$\ln(\text{Price}_2)$	$\ln(\text{Price}_2)$	$\ln(\text{Price}_2)$	$\ln(\text{Price}_2)$
$\varepsilon(\Delta \text{Pedigree})$	0.0329 (0.0410)	0.0322 (0.0413)	0.0213** (0.0083)	0.0213** (0.0084)
$\varepsilon(\Delta \text{Exhibition})$	0.1050** (0.0512)	0.1050** (0.0508)	0.0206* (0.0110)	0.0203** (0.0101)
$\varepsilon(\Delta \text{Literature})$	0.1479*** (0.0273)	0.1462*** (0.0270)	0.0270*** (0.0058)	0.0266*** (0.0056)
$\varepsilon(\Delta \text{Certification})$	0.0504** (0.0235)	0.0502** (0.0242)	0.0011 (0.0054)	0.0012 (0.0054)
$\ln(\text{Price}_1)$	0.6233*** (0.0556)	0.6183*** (0.0559)	0.6229*** (0.0557)	0.6164*** (0.0561)
<i>Time Span</i>	−0.0001*** (0.0001)	−0.0001*** (0.0001)	−0.0001*** (0.0001)	−0.0001*** (0.0001)
<i>Pedigree</i> <sub>1</sub>	No	Yes	No	Yes
<i>Exhibition</i> <sub>1</sub>	No	Yes	No	Yes
<i>Literature</i> <sub>1</sub>	No	Yes	No	Yes
<i>Certification</i> <sub>1</sub>	No	Yes	No	Yes
Artist fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes
Hedonic controls	Yes	Yes	Yes	Yes
Auct. house branch fixed effects	Yes	Yes	Yes	Yes
Number of observations	6,346	6,346	6,346	6,346
$R^2$	0.9504	0.9506	0.9505	0.9506

Notes. This table shows models addressing endogeneity concerns in the relation *Past Price and Provenance Changes*. The dependent variables  $\Delta$ Pedigree,  $\Delta$ Exhibition,  $\Delta$ Literature, and  $\Delta$ Certification in panel A are changes in the information related to these provenance dimensions between the first and second sales of each repeat sales pair. The changes are captured by (i) dummy variables equal to one when new provenance information arises between the first and second sale and zero otherwise or (ii) changes in the textual length of each dimension.  $\ln(\text{Price}_1)$  is the natural log of the first sale hammer prices in real US\$. *Time Span* is the number of days between the two sales. All regressions include initial pedigree, exhibition, literature, and certification information provided just prior to the first transaction of the repeat sales pair. In panel B,  $\varepsilon(\Delta \text{Pedigree})$ ,  $\varepsilon(\Delta \text{Exhibition})$ ,  $\varepsilon(\Delta \text{Literature})$ , and  $\varepsilon(\Delta \text{Certification})$  are the corresponding residuals from the regressions in panel A. All regressions include hedonic controls (see Section 2.1.1), and artist, year, month, and auction house branch level (AH) fixed effects. Standard errors (S.E.) are reported in parentheses and clustered at the auction branch level.

\*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

**Table 6.** Exogenous Sales Decisions

Dependent variable	Sold[0,1]		Ln(Price)	
	(1) Indicator	(2) Text length	(3) Indicator	(4) Text length
<i>Pedigree</i>	0.0368** (0.0180)	0.0104** (0.0044)	0.1659*** (0.0513)	0.0439*** (0.0125)
<i>Exhibition</i>	0.0791*** (0.0277)	0.0158*** (0.0055)	0.4118*** (0.0910)	0.0866*** (0.0180)
<i>Literature</i>	0.0347 (0.0232)	0.0073 (0.0049)	0.5084*** (0.1180)	0.1082*** (0.0274)
<i>Certification</i>	0.0470 (0.0302)	0.0091 (0.0116)	0.1366 (0.1016)	0.0211 (0.0164)
<i>Ln(Low Price Estimate)</i>	-0.1221*** (0.0146)	-0.1231*** (0.0145)		
Artist fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes
Hedonic controls	Yes	Yes	Yes	Yes
Auction house branch fixed effects	Yes	Yes	Yes	Yes
Number of observations	26,733	26,733	16,527	16,527
R <sup>2</sup>	0.4369	0.4371	0.8499	0.8509

*Notes.* This table presents the models relating provenance effects to the probability of being sold or hammer price for a subsample of transactions of which the sales decision may be taken more exogenously (less dependent on past prices). The sample employed includes sales retained when their auction title refers to “Estate Sale,” “Property of,” “Legacy,” “Bequest,” “Heritage,” “Gift,” “Endowment,” or “Charity” (in singular or plural) or contains “Late” plus a person’s name. The dependent variables are (i) *Sold*[0,1] that equals one if the auction lot is successfully sold and (ii) *Ln(Price)*, which is the natural log of deflated hammer price in real US\$. *Pedigree*, *Exhibition*, *Literature*, and *Certification* are defined in Appendix A. In columns (1) and (3), the provenance variables equal one if a painting has corresponding provenance information (by provenance dimension) and zero otherwise. In columns (2) and (4), the provenance variables are the natural log of text character length. *Ln(Low Price Estimate)*, the log of the low price estimate, is a proxy for the reserve price. All regressions include hedonic controls (see Section 2.1.1), and artist, year, month, and auction house branch level (AH) fixed effects. Standard errors (S.E.) are reported in parentheses and clustered at the auction branch level.

\*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Section 3.3.3. The dependent variable is the annualized geometric returns of the painting in the repeat sales transaction, which we regress on the changes in the provenance also controlling for the initial provenance predating the first sale. Columns (1)–(4) of Table 7 are based on the full repeat sales sample, whereas columns (5)–(8) are based on the subsample in which the repeated sales take place at the same auction house branch to alleviate concerns about policy discrepancies among auction houses (or branches within the same auction house) in the provision of provenance information. For changes in provenance, we have two specifications: (i) indicator variables capturing whether a change occurred from no information to an insertion of provenance information (the reverse case does not occur) and (ii) changes in the provenance text’s character length.<sup>31</sup>

Table 7 shows that changes in information related to the *Exhibition* and *Literature* dimensions have a material impact on annual *Returns*. If a painting has no exhibition history prior to the first sale but is exhibited during the holding period, the annualized returns rise by 16.5 percentage points (column (3)). Likewise, when no information on the painting in the (art history) literature was present prior to the first sale but was subsequently present prior to the second sale, the annualized return rises by 14.9% (column (3)). An extension of the catalog text on the exhibition history of the painting also positively affects

returns (columns (2) and (4)). Columns (3) and (4) show that the initial provenance dimensions (e.g., *Pedigree*<sub>1</sub>), disclosed prior to the first transaction of the repeat sale, have no statistically significant impact on the subsequent returns. This implies that this information was already priced at the first auction and that only additional information affects the second sale and, hence, returns. We also control for a price effect that runs through the “upgrade” of the auction house, which captures that the second transaction is made at a more prestigious auction house than the previous one. Higher hammer prices are reached, reflecting that more prestigious auction houses may reach a wealthier clientele and may provide a “quality stamp” for an auctioned painting. We also want to eliminate the effect of idiosyncratic provenance provision policies of auction houses by limiting the repeat sales sample to the transactions that took place in the same auction house (branch). We observe that the provision of more information highlighting the role of the painting in the literature and its exhibition history both have a stronger economic impact on returns (columns (7) and (8) of Table 7).

## 4. Extensions and Robustness Tests

### 4.1. Subsample of Artworks with Provenance Information

A potential concern is that, for the majority of the auctioned paintings, no provenance information is provided

**Table 7.** Provenance Effects and Returns for Repeat Sales Transactions

Dependent variable: Return	Full RS sample				RS sample with transactions in the same auction house			
	(1) Indicator	(2) Text length	(3) Indicator	(4) Text length	(5) Indicator	(6) Text length	(7) Indicator	(8) Text length
$\Delta Pedigree$	-0.1123 (0.1064)	0.0032 (0.0199)	-0.0489 (0.1474)	0.0065 (0.0218)	-0.4264 (0.3403)	0.0014 (0.0634)	-0.1382 (0.4136)	0.0160 (0.0691)
$\Delta Exhibition$	0.1401** (0.0621)	0.0246** (0.0125)	0.1646** (0.0715)	0.0250** (0.0123)	0.1865* (0.0960)	0.0341 (0.0253)	0.2761* (0.1468)	0.0548** (0.0273)
$\Delta Literature$	0.1206* (0.0692)	0.0173 (0.0138)	0.1487** (0.0627)	0.0176 (0.0134)	0.2332*** (0.0748)	0.0437** (0.0214)	0.2706*** (0.0713)	0.0413* (0.0223)
$\Delta Certification$	0.0853 (0.0850)	0.0062 (0.0200)	0.0010 (0.0607)	-0.0006 (0.0185)	0.1299* (0.0673)	-0.0013 (0.0271)	0.0227 (0.0612)	-0.0054 (0.0235)
$Pedigree_1$			0.0985 (0.2082)	0.0374 (0.0407)			0.6753 (0.5931)	0.1977 (0.1282)
$Exhibition_1$			0.0555 (0.0711)	-0.0070 (0.0133)			0.1232 (0.2382)	-0.0019 (0.0408)
$Literature_1$			0.0847 (0.0700)	-0.0010 (0.0166)			0.1619 (0.1221)	0.0062 (0.0232)
$Certification_1$			-0.1658 (0.1152)	-0.0655* (0.0390)			-0.1784* (0.0943)	-0.1153** (0.0554)
<i>Auction House Upgrade</i>	0.3462*** (0.1075)	0.2931** (0.1220)	0.3888*** (0.1218)	0.3886*** (0.1181)				
Artist fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Auct. house branch fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hedonic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	6,647	6,647	6,647	6,647	4,236	4,236	4,236	4,236
$R^2$	0.3361	0.3357	0.3366	0.3364	0.3919	0.3907	0.3940	0.3947

*Notes.* The regressions are based on the repeat sales (RS) sample. The dependent variable *Return* is the geometrically annualized return; results are for the full RS sample (Models 1–4) and for the RS transactions that took place in the same auction house (branch) (Models 5–8).  $\Delta Pedigree$ ,  $\Delta Exhibition$ ,  $\Delta Literature$ , and  $\Delta Certification$  are the changes in provenance information in between the two sales. The change in provenance precedes the return as provenance information in the auction catalogue usually precedes the auction by 4–6 weeks. They are defined as either (i) a change from no provenance information (pedigree, exhibition, literature, and certification) at the first sale to available information at the second sale or (ii) a change in textual length of provenance information (by dimension) from the first sale to the second sale. *Auction House Upgrade* equals one when a second sale moves up from a small auction house to a medium or big auction house or when a second sale moves from a medium auction house to a large one. All regressions include the hedonic controls that are presented in Section 2.1.1, and artist, year, month, and auction house branch level (AH) fixed effects. Standard errors (S.E.) are reported in parentheses and clustered at the auction branch level.

\*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

in catalogs, and only traditional hedonic variables are provided (artist's name, title, measurements, medium, lot number, auction house (branch), and transaction date). We, therefore, restrict our sample to only those observations with available provenance information and replicate the baseline results for the full sample as presented in Table 2. The findings are consistent with our baseline results; *Pedigree*, *Exhibition*, *Literature*, and *Certification* are qualitatively and statistically similar in the probability of being sold and price regressions (Online Table A.VI).

#### 4.2. LASSO Estimations

A methodological concern may be that we estimated high-dimensional fixed effects models, which could result in a decrease in predictive power for each added variable. Therefore, we verify our results by means of LASSO estimation, which enables us to examine which

of the many variables (we have more than 100,000 fixed effects—about 10% of the total number of observations as we include, for instance, artists fixed effects) are the most important ones for purposes of prediction (Belloni et al. 2014). Effectively, this method chooses a simpler model with fewer variables, reduces overfitting, increases out-of-sample prediction, and generates a more efficient algorithm. The LASSO results for the hedonic probability of being sold and price regressions indicate that the statistical significance of the provenance variables remains valid (Online Table A.VII). In fact, the economic significance of the provenance variables becomes even larger relative to the baseline results of Table 2. It should be noted, however, that the LASSO method is mainly designed for prediction such that drawing inferences from model parameters could still be problematic in the sense that, when some of the parameter estimates are set to zero, an omitted variable bias could arise (which is the cost paid

for a reduction in variance). Therefore, we apply this method only as a robustness test to verify whether the provenance factors are of prime importance in relation to probability of being sold and auction pricing.

### 4.3. Nonlinearities

We also study nonlinear relations between prices and provenance. We use the text length of the provenance and the number of items mentioned in the provenance (e.g., the number of exhibitions in which a painting was shown; the number of art books in which the painting is depicted). This analysis is less suitable for the pedigree dimension (uninterrupted ownership chain, type of owners) and certification (based on presence or not). We find that the prices increase with the amount of provenance information provided in the auction catalog but only regressively so (given the negative second derivative) (see Model 2 of Online Table A.VIII). Similar patterns are presented in Models 4, 6, and 8. Thus, what matters is that a painting was, for instance, exhibited in a prestigious exhibition once or a few times, but a high frequency loses its effect as reflected in the prices.

### 4.4. Liquidity

It is possible that the provision of provenance is related to past liquidity and, hence, that the significance of provenance in the pricing models is induced by past liquidity. We estimate models with the provision of provenance and hammer prices as dependent variables and include various measures of liquidity. These measures include past sales by artist measures of varying time windows (past five through one years) and the sales ratio (sales/paintings offered for sale) by artist (over past five through one years). The former measure captures liquidity; the latter measure captures liquidity, including market sentiment. The liquidity measures may affect the subjective risk related to the quality of the painting. For each measure, we can consider global sales or sales by country depending on different assumptions about what auction information potential buyers consider when contemplating a purchase of a painting at a specific auction house branch. Would they look up information on past auction transactions in the country where a painting was currently being sold, or would they collect information on worldwide sales of an artist's oeuvre? An argument for the latter option is that it is rather easy to look up global sales when one subscribes to specialized auction sales databases, such as Artnet, Artprice, or Blouin, or when one collects basic information from the internet (which is *de facto* international). Our conclusion from the pricing models is that past liquidity (past sales and sales ratio) affects current hammer prices because buyers may be willing to pay a premium for the oeuvre of an artist that is shown to be liquid in the sense that many paintings were offered for sale and many of those paintings were successfully

auctioned (Online Table A.IX). We demonstrate that, in these models, the impacts of provenance on prices documented in previous sections are not qualitatively affected by the introduction of liquidity (Online Table A.IX). Furthermore, the interaction between liquidity and provenance does not affect prices (Online Table A.X). As the provision of provenance may be affected by liquidity and this relation may go either way—auction houses may offer more provenance for the most liquid artists given that they sell well or they may offer more provenance for a less liquid oeuvre as more effort is needed to sell the paintings—we examine whether provenance is related to lagged liquidity. We find that all models directly relating current provenance provision to past liquidity exhibit insignificant relations (Online Table A.XI). Our overall conclusion is that liquidity does not affect prices through provenance.

### 4.5. Artistic Style

Given that provenance may be more important for specific schools of art, we investigate whether the presence of a provenance premium depends on artistic style. We distinguish between 13 styles: (1) Medieval and Renaissance; (2) Baroque; (3) Rococo; (4) Neoclassicism; (5) Romanticism; (6) Realism; (7) Impressionism and Symbolism; (8) Fauvism and Expressionism; (9) Cubism, Futurism, and Constructivism; (10) Dada and Surrealism; (11) Abstract Expressionism; (12) Pop Art; and (13) Minimalism and Contemporary. We observe that provenance and its four dimensions (pedigree, literature, exhibitions, and certification) strongly and significantly affect prices for each subsample by school of art, controlling for the extensive set of hedonic controls and fixed effects (including auction house, year, seasonality). The economic effects are large for each subsample and similar; there is no evidence that younger schools of art have a different provenance premium (Appendix D).

### 4.6. Auction House Types

Some auction houses have a global reach, whereas others are smaller and focus on regional art buyers. Whereas we control for auction house branch fixed effects in the price and provenance provision regressions, we zoom in on the provenance effects by auction house type and relate this to paintings of specific price ranges. The purpose is to study whether a substitution effect exists between auction house types and the provenance information provided. We distinguish between (1) large auction houses (Christie's and Sotheby's), (2) medium-sized auction houses (Bonhams, Phillips, and other important auction houses in the United States and Europe),<sup>32</sup> and (3) other small auction houses. In panel A of Online Table A.XII, we observe that the hammer prices are strongly related to provenance (*ex ante* provided in the catalog) for each type of auction house (Models 1–3). Thus, regardless of the auction house type

(from prestigious to small), provenance information is strongly correlated with prices. The same holds for all four dimensions of provenance (Models 4–6) except certification for the largest international auction houses (Christie's and Sotheby's), for which certification may matter less given that these auction houses' transactions are insured in that they repay the purchase of paintings that could later be exposed as forgeries. In panel B, we repeat the relationship between provenance and auction house for different price ranges (quartiles). We observe that provenance matters for paintings of each price quartile and each type of auction house (the only exception is the cheapest paintings offered by the largest auction houses also because of subsample size as Christie's and Sotheby's sell relatively few paintings within this price range).

#### 4.7. (Non)established Artists

One may wonder whether paintings by established artists (e.g., Pablo Picasso) need less provenance to sell or reach high prices. Alternatively, established artists need more provenance information because (a) artists with a more expensive oeuvre are more likely to be subject to being forged and (b) prolific artists for which there do not exist comprehensive catalogues raisonnés are also more likely to be forged. In addition, it is rumored (but difficult to prove) that about 40%–50% of the contemporary art market consists of forgeries. Consequently, for an established artist, provenance information is very important. We perform the following test based on price estimates in the auction catalogs: we take the lowest price estimate and consider artists above the 75th percentile as established artists and those below the 25th percentile as nonestablished. We take only the lowest estimate, which is close to the secret reserve price, and not the highest estimate as this may be used to create anchoring effects. We document in Online Table A.XIII that, for both established and nonestablished artists, a strong correlation exists between prices and provenance and its dimensions. The relationship between prices and provenance is even stronger for established artists. As a further illustration of the importance of provenance within the oeuvre of established artists, we focus on Picasso, Raoul Dufy, and Andy Warhol and confirm the conclusion (see Online Table A.XIV).

#### 4.8. Trust vs. Quality

Of the four provenance measures, *Pedigree* and *Certification* seem *prima facie* the dimensions that are expected to create most trust in the offered object of art, whereas exhibitions and literature might not only capture trust, but also to some degree reflect or enhance the quality of a painting. *Pedigree* captures ownership, and *Certification* captures whether there is physical/oral evidence of authenticity. In what follows, we make a few caveats about quality versus trust versus glamour.

First, with regard to *Pedigree*, ideally, there is an uninterrupted ownership chain between the current seller and the artist; a painting was, for example, in the possession of a family for decades/centuries or in the personal collection of a private collector. However, for paintings that were created decades or centuries ago, an uninterrupted ownership chain is rare. In many cases, the ownership link is severed, which may be caused by the nonavailability of intermediate ownership information or originate from the discretion that art buyers cherish about their art purchases/collections. In auctions, the current seller is hardly ever mentioned, and in the vast majority of cases, the counterparty prefers discretion (one often only learns *ex post*—if at all—who has purchased a painting). As a consequence of this preference for discretion, the average ownership chain almost inevitably has some lacunas. Missing information may arise when this information is not deemed important (e.g., when nothing is known about an individual intermediate owner apart from temporary ownership, the catalog might not offer such information). Still, even in cases of interrupted ownership chains, the number of ownership indications and the detail of this information could help enhance trust. It may be sufficient for potential buyers that there is information available on the first transaction (between an artist or gallery and a first buyer) as this may then be sufficient proof that the painting is not a forgery. Our textual analysis picks this up as well as transactions through the generations (we capture inheritance/descent—lemmatized). In addition, we identify whether specific types of owners appearing in the pedigree generate trust—either because these owners are considered knowledgeable about art (investments) or may be considered wealthy enough to hire the necessary expertise in order to verify art quality/authenticity. As such, we identify prominent collectors (we gather information on 3,885 important collections around the world) and also of nobility/royalty as art is often in such families across generations.

Thus, whereas pedigree may be more related to trust than to quality, some types of owners in the pedigree may unavoidably be related to art quality (famous collectors) or even glamour (wealthy families, CEOs, and celebrities). The presence of such past owners can be related to (a) trust in authenticity because these past owners can afford to buy in expertise, (b) quality of the art object as past owners' wealth could enable them to focus on the "best" art (e.g., by art school, period, and artist), and (c) glamour in case current buyers desire to own a piece of art that belonged to a person they admire. Hence, even in the *Pedigree* dimension, for some types of owners, trust, quality, and glamour are inextricably connected. Included in *Pedigree* is the reputation of the (historical) auction house/dealer responsible for past transactions because the quality of expertise by the research departments of auction

houses varies. Therefore, we distinguish in detailed regression models among Christie's, Sotheby's, Bonhams, Phillips, historically important auction houses (e.g., Hôtel Drouot, Dorotheum, and many others), and important current and renowned historic dealers. The fact that past transactions have passed through these auction houses/dealers can induce trust in that art might have been screened well in the past. Still, such paintings might be related to quality in that these auction houses/dealers might focus on auctioning "quality" art.

Second, *Certification* is most clearly related to trust, as here, we search for certificates, photographs, witness statements (by the artist's family, pupils, descendants), and statements by foundations and experts. The caveat here is that certificates provided by experts are not foolproof. Ample examples of errors made by experts can be cited; one of the most prominent examples is the certification of Han van Meegeren's biblical Vermeers, which were authenticated by the leading expert Abraham Bredius (whose mistake enabled van Meegeren to mislead the art scene in the first half of the 20th century as all the Biblical Vermeers were fakes). Van Meegeren also benefited from the expert Cornelis Hofstede de Groot's mistake in identifying a fake Frans Hals as authentic.

Whereas one can regard the provenance dimensions *Literature* and *Exhibition* as capturing quality (rather than trust) as one can expect that the most important paintings appear in the literature and or are exhibited, this is only partially true because both dimensions also have a strong trust aspect in the following sense. In relation to *Literature*, catalogues raisonnés strive, by definition, to list and describe an artist's oeuvre exhaustively. Moreover, art books on specific artists often cover a substantial part of their works. The *Literature* dimension also captures whether (historic) photos/images/illustrations of a particular work are included in art books and this dimension in this respect also plays a role in certification. We focus on the most authoritative press (e.g., scientific books by university presses) and the most reputable publishers, which all use refereeing committees). Thus, these books are based on art historical/art market research. Consequently, we argue that the literature dimension includes an important aspect of trust (likely even dominating the "quality" label of this dimension), but acknowledge that a complete separation between quality and trust cannot be achieved. Regarding the *Exhibition* dimension, it may indeed be the case that an artist's highest quality paintings are exhibited as part of a permanent collection of museums and included in exhibitions organized by other renowned museums/galleries. Nonetheless, exposure through the latter channel may face impediments as museums or collectors often do not let their most important works travel. In addition, exhibitions focusing on artists' impact and significance in art history exhibit

paintings from their early or late periods to offer a comprehensive picture of their careers or to show influences by or on other artists. Moreover, exhibitions often trigger new research on the exhibited work and the artist with scientific articles published in the exhibition catalog such that the dimensions *Exhibition* and *Literature* can be correlated (but not to the extent of causing multicollinearity in our models). Finally, over the past three decades, new exhibitions with loaned paintings often lead to technical examinations (infrared reflectography, noninvasive spectroscopic imaging, X-rays, chemical analysis of paint, etc.) or restoration. Consequently, exhibited works are thoroughly scrutinized such that exhibitions contribute to deeper insights affecting trust. Of all the four dimensions, the *Exhibition* dimension is most related to quality or salience.

In summary, trust and quality (and even status or glamour) are embedded in provenance. *Pedigree* and *Certification* capture trust but may reflect to some degree quality and status. As argued, *Literature* may capture trust to a much larger degree than it reflects the quality of a painting. *Exhibition* may indeed be more related to quality than trust, and the latter aspect originates from additional exhibition-induced research.

If trust were priced, one could expect a higher probability of being sold and higher price premiums for the provenance dimensions in this order: pedigree/certification, literature, and exhibitions. However, the statistical and economic significance also depends on the degree to which each of the dimensions (and certainly their constituting elements) (a) captures quality/status/glamour and (b) is present within the provenance records of the auction catalogs. The pricing models of Table 2 (panel A) report that each of the four dimensions has a strong price impact with *Literature* exhibiting the strongest and *Exhibition* the second strongest effect such that a high price premium may capture not merely trust but also some aspects of the quality of the painting (whereas it should be recognized that the models control for artist reputation, auction house reputation, and physical aspects of the painting as well as transaction characteristics). The dimensions *Pedigree* and *Certification* may capture trust to a much larger extent than quality, which may explain why the price premium is lower (but still strongly significant and economically large). When we investigate in more detail (panel B), we find that, within the *Certification* dimension, a physical certification by the artist (trust) has the strongest impact on the probability of being sold. Conditional on being sold, we find that all types of certificates affect hammer prices (with the highest impact by a physical certificate by an expert and the nonphysical testimonial by the artist).

Within the *Literature* dimension, the probability of being sold is positively affected first by the presence of the picture on the cover page (capturing a combination of trust, quality, and salience) and second by inclusion

in the catalogue raisonné (which is also one of the strongest trust measures). Again, all constituting elements of the *Literature* dimension affect prices (cover page, publication by authoritative press, etc.).

For the *Pedigree* dimension, past ownership (prominent collectors, corporate collections, nobility/royalty, wealthy families) strongly affects the probability of being sold. It is also remarkable that the category of famous sportspeople affects the probability and prices; here, the aspect of glamour (owning a piece acquired by an idol) may add to both trust and quality (as wealthy sportspeople can afford quality and pay for expertise to select art). Descent information is priced (also when a first sale can be traced directly to an artist).

Turning to the *Exhibition* dimension, we find that the most prominent exhibitions (organized by prominent museums and renowned galleries) affect the probability of being sold as well as prices. These results can be explained by selection of the highest quality paintings, but the arguments outlined explain why there may also be a strong trust aspect.

## 5. Conclusion

Trust is key for any type of market but particularly for illiquid, opaque, and largely unregulated markets, such as the art market. A lack of trust undermines sales and prices, which is worsened when fakes and forgeries are rumored to circulate in the art market. Therefore, guarantees about the authenticity of an art object are pivotal in creating trust, and the provision of provenance can be a (partial) solution by emitting a signal about the art's authenticity. Provenance comprises records of ownership or pedigree, exhibition history, literature coverage, and certification, all of which relate to the artwork's authenticity and can enhance trustworthiness of the object offered for sale. For example, if there is evidence that a painting was originally purchased from the artist or artist's family, from a first buyer whose family has held the painting for generations, or from famous collectors and there are corroborating documents, the potential buyer's caution is attenuated.

We investigate the impact of providing detailed provenance information, measured by hundreds of variables resulting from textual analysis applied to auction catalogs, along with the set of traditional art value determinants on the probability of being sold,

hammer prices, and returns of about two million paintings and works on paper. We find that provenance information provision increases the probability of being sold by 2%–4%, leads to a price premium of 14%–54%, and increases the annualized returns by 5–16 percentage points after controlling for artwork and transaction characteristics (e.g., topic, signature, medium, and measurements) as well as artist, time, and auction house branch fixed effects.

A first type of endogeneity is embedded in the decision to offer provenance, which may be affected by recent price increases of the painting (in case of a repeat sale) or of similar paintings (e.g., by the same artist or school) as well as by expected prices (proxied by price estimates). To address the concern of reverse causality in the relationship between past prices and changes in provenance, we study the provenance effects for artists affected by the discoveries of fakes and forgeries, and exploit the Christie's policy change in the provision of provenance information following a litigation case against the firm as a quasi-natural experiment in a DiD setting. We also run two-staged regressions for repeat sales transactions to control for changes in provenance induced by past high prices. To address the potential endogeneity concern in the sales decision, we examine subsamples of exogenous sales, which we expect to be less affected by past price trends (e.g., sales following the death of collectors). These attempts to address endogeneities yield results similar to those of our baseline models. In conclusion, provenance information is an important factor corroborating an artwork's authenticity and creating trust in art markets as reflected in sales, prices, and returns.

## Acknowledgments

The authors thank Jennie Bai, Fabio Braggion, Geraldine David, Frank de Jong, Joost Driessen, Rik Frehen, William Goetzmann, Kathy Graddy, Harrison Hong, Jun Huang, Daniel Karpati, Arthur Korteweg, Stefano Lovo, Stijn van Nieuwerburgh, Kim Oosterlinck, Christoph Schneider, Christophe Spaenjers, Oliver Spalt, Jan-Jacob Van Turnhout, Patrick Verwijmeren, Hong Yan, Chris Yung, and seminar and conference participants at Tilburg University, Columbia University, New York University, Financial Management Association 2019 European Conference, Southwestern Finance Association 2019 Annual Meeting, and European Financial Management Association 2019 Annual Meeting for valuable comments. All errors are our own.

## Appendix A. Variable Definitions

<i>Auction European</i>	The category includes important auction houses in Europe: Lyon & Turnbull (Scotland); Francis Briest/Artcurial Briest (France); Ader, Picard & Tajan/Ader & Tajan/Tajan (France); Bruun Rasmussen (Denmark); Dorotheum (Austria); Koller (Switzerland); Lempertz (Germany); Neumeister (Germany); Finarte (Italy); Bukowskis (Sweden); and Stockholms Auktionsverk (Sweden).
<i>Auction American</i>	The category includes important auction houses in the United States: Butterfields (until 2002), Swann Auction Galleries, Skinner, Doyle New York, Freeman's, and Leslie Hindman.
<i>Auction House Upgrade</i>	A dummy variable equal to one when a second sale (in a repeat sale) moves up from a small auction house to a subtop or large prominent auction house or when it moves up from a medium-sized one to a large prominent one. Large prominent auction houses include all the branches of Christie's and Sotheby's; the subtop auction houses include all branches of Bonhams and Phillips or other important European and U.S. auction houses, which we group as other important European auction houses and other important American auction houses (see the classification above). Small auction houses include all unclassified auction houses.
Pedigree (past ownership)	<i>Prominent Collectors</i> , <i>Royalty/Nobility</i> , <i>Wealthy Families</i> , <i>CEOs</i> , <i>Influential People (Time 100)</i> , <i>Celebrities</i> , and <i>Famous Sportspeople</i> are dummy variables equal to one if the painting has been in the collections of those respective types of collectors.
Pedigree (descent)	<i>Directly from Artist</i> , <i>From Artist Family</i> , and <i>From Sitter</i> are dummy variables and equal one if the artworks are acquired directly from these categories, respectively. <i>Other Descent Information</i> equals one if the artwork's pedigree information contains any descent information.
Pedigree (past sales)	<i>Sold by Sotheby's or Christie's</i> , <i>Sold by Bonhams or Phillips</i> , <i>Sold by Historic Auction Houses</i> , <i>Sold by Other Important Auction Houses</i> , and <i>Sold by Prominent Dealers</i> are dummy variables equal to one if the painting was sold in the past at the auction house (types)/dealer, respectively.
Pedigree (other)	<i>Corporate Collection</i> and <i>Private Collection (Anonymous)</i> are dummy variables equal to one if the artwork was at one point in corporate and private collections, respectively. <i>Other Pedigree Information</i> is a dummy variable equal to one if the painting includes pedigree information not in any of the preceding pedigree categories.
Exhibition	<i>Prominent Exhibition</i> , <i>Prominent Art Fair</i> , <i>Prominent Museum</i> , <i>Other Museum</i> , <i>Cultural City</i> , and <i>Gallery Exhibition</i> are dummy variables equal to one if the artwork was at one point exhibited in these types of exhibitions/fairs/museums/cities/galleries, respectively.
Literature	<i>Catalogue Raisonné</i> , <i>Cover Page</i> (of an art history book), <i>Illustration</i> (in an art history book), and <i>Authoritative Press</i> are dummy variables equal to one if the artworks are illustrated in these ways, respectively. <i>Other Literature</i> is a dummy variable and equals one if the artwork's catalog information refers to information in other types of publication.
Certification	<i>Artist (Physical)</i> , <i>Artist Family (Physical)</i> , <i>(artists') Association (Physical)</i> , <i>Expert (Physical)</i> , and <i>Other People (Physical)</i> are dummy variables equal to one if the artwork is auctioned with physical certification issued by these sources, respectively. <i>Artist (Nonphysical)</i> , <i>Artist Family (Nonphysical)</i> , <i>Association (Nonphysical)</i> , <i>Expert (Nonphysical)</i> , and <i>Other People (Nonphysical)</i> are dummy variables and equal one if the artworks are auctioned with nonphysical certification issued by these sources, respectively.

## Appendix B. String Searches and Sources

### Panel A: Pedigree

#### Past Ownership

##### - Prominent Collectors

Sources: various lists from *Artnet World's Top Art Collectors 1990–2017*; *Artnet 20 of the World's Most Innovative Art Collectors*; *Forbes Top Billionaire Art Collectors*; *Grove Art Online*; art collectors from 18th century to 21st century in Wikipedia; 3,885 names<sup>33</sup> in total.

##### - Royalty/Nobility

Sources: textual analysis by searching the royal and noble ranks in seven languages (English, Latin, Dutch, French, German, Italian, and Spanish), including imperial titles; high royal titles; royal titles; princely, ducal, and other sovereign titles; tribal titles; religious titles; other sovereigns, royalty, peers, and major nobility; minor nobility, gentry, and other aristocracy from various areas, cultures, and countries in history from Wikipedia,<sup>34</sup> 364 ranks and titles in total.

##### - Wealthy Families

Sources: *Forbes World's Billionaires 1987–2017*; *Contemporary Wealthiest Family List* from Wikipedia,<sup>35</sup> 8,479 names in total.

##### - CEOs

Sources: various sources, including *Chief Executive CEOs of the year 1986–2017*; *Chief Executive CEO1000 tracker full list*; *Forbes Most Powerful People 2007–2016* (CEO, founder, cofounder, chairman, executive vice president, co-chief investment officer, chief investment officer, director-general, etc.); *Forbes America's top 20 favorite bosses*; *Forbes world's 10 most powerful CEOs 2016*; *Industry Week 10 most popular manufacturing CEOs*; *Industry Week CEO of the year 2004*; *Industry Week CEO of the century*; *Time Magazine person of the year (1991, 1997, 1999, 2005, 2010)*; *Cable News Network (CNN) top 25 influential business leaders 2005*; *Ernst & Young entrepreneur of the year award 2001–2017*; *Atlantic Business CEO of the year 2005–2017*; *Finance Monthly CEO awards 2016–2017*; *Harvard Business Review best-performing ceos in the world 2010–2017*; *The New York Times Equilar 200 2016*; *Barron's world's best CEOs 2016*; *Fortune 25 most powerful people in business*; *Wikipedia CEOs of notable companies*,<sup>36</sup> etc.; 2,703 names in total.

**Appendix B.** (Continued)- Influential People (*Time* 100)

Sources: *Time* 100 lists of titans, pioneers, artists, leaders, and icons 2004–2017 and *Time* 100 persons of the 20th century; 3,519 names in total.

## - Celebrities

Sources: IMDb Top 1,000 actors/actresses/directors/producers; *Forbes* world's highest paid celebrities 1999–2017 (including actors, actresses, comedians, models, musicians, vocalists, directors, producers, filmmaker, TV personalities); *Forbes* world's highest-paid TV show hosts 2016; British television personalities in Wikipedia;<sup>37</sup> American television talk show hosts in Wikipedia,<sup>38</sup> 6,255 names in total.

## - Famous Sportspeople:

Sources: *Forbes* world's highest-paid sportspeople 2012–2017 (including boxing, golf, basketball, tennis, soccer, football, baseball, racing, motorcycle, cricket, track, auto racing, mixed martial arts, etc.); world champions and superstars of sports including golf, basketball, tennis, soccer/football, baseball, motorsport, cricket, and hockey; 4872 names in total.

Golf: official world golf ranking top 100 2003–2017; ESPN golf world rankings top 100 2017; U.S. Open champions 1895–2017; The Open Championship 1860–2017; Masters Tournament champions 1934–2017; PGA Championship 1916–2017.

Basketball: NBA all-stars; hall of fame.

Tennis: Association of Tennis Professionals rankings top 100; Women's Tennis Association rankings top 100; Australian Open champions 1969–2017; French Open champions 1891–2017; Wimbledon champions 1877–2017; US Open champions 1881–2017; Grand Slam related tennis records.

Soccer: FIFA 100; *English International Football Magazine* the greatest players of the 20th century; *World Soccer Magazine* world player of the year 1982–2016; *Guardian* top 100 footballers 2013–2016; *Guardian* World Cup top 100 footballers of all time.

Baseball: Baseball hall of fame 1936–2017.

Motorsport: Formula One World Drivers' Champions 1950–2017; Formula Two Champions 1967–2012; 500cc/MotoGP Motorcycle World Champions 1949–2017; Motorsport Drivers Current Standings Top 32.

Cricket: ICC ODI championship batsmen top 100; historical test cricket rankings 1877–2016; current test rankings top 10; current ODI rankings top 10; current T20I rankings top 10.

Hockey: *The Hockey News* Top 100 National Hockey League players of all-time; NHL top 200 rankings 2017–2018; ESPN top 300 fantasy hockey rankings 2017–2018.

## - Corporate Collection

String search: corporate collection.

## - Private Collection (Anonymous)

String search examples: private collection.

## Descent

## - Directly from Artist

String search examples: from artist, from the artist, directly from artist, directly from the artist, by artist, by the artist, gift(s) (courtesy/donation(s)/goodwill(s)/bequest(s)/endowment(s)/present(s)) of (the) artist.

## - From Artist's Family:

String search examples: by descent (by inheritance/estate/legacy/inherited/descended/collection) from artist (the artist/by artist/by the artist/of artists/of the artist/from painter/from the painters), artist's (artists/artist's/artist's) + family (son/daughter/wife/husband/partner/spouse/girlfriend/boyfriend/widow/brother/sister/sibling/cousin/grandson/granddaughter/uncle/aunt/nephew/niece/heirs/heir/grandnephew/grandniece).

## - From Sitter:

String search examples: sitters, sitter, from sitter, from the sitter, from sitters, from the sitters.

## - Other Descent Information:

String search examples: descent, descended, inheritance, inherited.

## Past Sales Channel

## - Sold by Sotheby's or Christie's

String search examples: Christie; Sotheby.

## - Sold by Bonhams or Phillips

String search examples: Bonhams; Phillips.

## - Sold by Historic Auction Houses

Sources: Getty Provenance Index.

String search examples: Achenbach; Anderson & Garland; Thomas Dodd; F. Dörling; Dorotheum; Dowell's; Hôtel Drouot; Galerie Fischer; Edward Foster & Son; Messrs Foster; Frederik Muller & Co.; John Gerard; Gerard-Tasset-Juge; Gilhofer & Ranschburg; Goesin-Verhaeghe; Pierre François; Paul Graupe; Heinrich Hahn; Hugo Helbing; Galerie Helbing; Internationales Kunst Auktionshaus; George Jones; Albert Kende; S. Kende; Thomas King; August Klipstein; Galerie Kornfeld; Knight Frank & Rutley; W. S. Kündig; Hans W. Lange; Langford; Mathias Lempertz; Heinrich Lempertz; Gallery Lempertz Contemporaria; Venator & Hanstein; Kunsthaus Lempertz; Leo Spik; Rudolph Lepke; Bignell Marle; P. L. Mastraeten; Franz A. Menna; Corneille Moor; Morrison Mcchlerly; Max Perl; Thomas Philipe; Harry Phillips; Mr. Prestage; Puttick & Simpson; William Richardson; George Henry Robins; Henry J. Robins; Robinson & Foster; Robert Saunders; Hodgson & Co; Saunders & Hodgson; Philippus Van Der Schley; James Webber Southgate; George Squibb; Squibb & Son; Rushworth, Abbott & Co; George Stanley; J. A. Stargardt; William Stewart; E. J. Terlinck; De Vries; Adolf Weinmüller; Munich Auction House; Benjamin Wheatley; Willis's Rooms; Winstanley & Sons; Puttick & Simpson; Stewart, Wheatley & Adlard; Wheatley & Adlard.

## - Sold by Other Important Auction Houses

String search examples: Butterfields; Lyon & Turnbull; Francis Briest; Artcurial Briest; Tajan; Bruun Rasmussen; Dorotheum; Koller; Lempertz; Neumeister; Finarte; Bukowskis; Stockholms Auktionsverk; Swann Auction Galleries; Swann Galleries; Skinner; Doyle New York; Freeman's; Freeman's; Freeman's; Leslie Hindman.

---

## Appendix B. (Continued)

---

- Sold by Prominent Dealers

Sources: *Artnet* 100 best galleries, *Forbes* contemporary dealers, and *Grove Art Online* famous historic dealers; 233 names in total.

---

### Panel B: Exhibition

---

- Prominent Exhibition

String search examples: retrospective, rétrospective, anniversary, anniversaire, biennale, triennale, biannual, biennial, triannual, triennial.

- Prominent Art Fair

Sources: *Artnet*; *Artprice*

String search examples: ARCO Madrid, Armory Show New York, Art Basel, Art Basel HK, Art Basel Miami Beach, Art Cologne, Art Miami, Art Santa Fe, ARTISSIMA, Documenta Kassel, FIAC Paris, Frieze London, Frieze New York, India Art Fair, PAN Amsterdam, TEFAF Maastricht, TEFAF New York, Venice Biennale, BRAFA Brussels.

- Prominent Museum

Sources: most important museums of paintings in important art cities from *National Geographic*, *Wikipedia*, *Reuters*, and *The Telegraph*; 517 museums in total.

- Other Museum:

String search examples: museum, musée, museo, museu, museums, musea, museen, musées, museos, museus, musei.

- Cultural City

Sources: European Capital of Culture,<sup>39</sup> UN City 2016, City Mayors EU 500, City Mayors World 300, and other cultural cities (defined by locations with a considerable number of museums, galleries, and auction houses) around the world; city names in English, French, and original languages; 236 cities in total.

---

### Panel C: Literature

---

- Catalogue Raisonné

String search examples: catalog(ue)/catalog(ue) raisonne, catalogue/catalogue raisonné.

- Cover Page

String search examples: cover.

- Illustration

String search examples: illustration, illustrated, cover, images, image, photos, photo.

- Authoritative Press

Sources: 280 notable university presses from *Wikipedia*<sup>40</sup> and World's 57 largest book publishers from *Publishers Weekly Magazine*.

---

### Panel D: Certification

---

- Certification

String search examples: echtheitsbestätigung, gutachten, essay(s), assessment(s), opinion(s), appraisal(s), expert(s), expertise(s), report(s), mail(s), photo certificate(s), photocopy, photocopies, issued, verified, witnessed, authenticity, authentication.

- Forms physical

String search examples: photocertificate(s), report, written, handwritten, photocopy, photocopies, photo(s), photography, photographic, photograph, foto(s), foto's, photography, fotografische, photographie, fotografie, fotografien, photography, photographique, photographie(s).

- Issuers artist

String search examples: issued (verified/witnessed/certificates/certificate/certificate + signed/certified/authenticity/authenticity signed/authentication/authentication signed/authenticated/identified/identification/confirmed/confirmation/confirmatory information/registered/registration/registration card/registered/recorded/documentation/letter(s)/photo(s)/photo(s) signed/photograph(s)/photograph(s) signed) + by artist (by the artist/from artist/from the artist/of artist/of the artist).

- Issuers artist's family

String search examples: son, daughter, wife, husband, partner, spouse, girlfriend, boyfriend, widow, brother, sister, sibling, cousin, grandson, granddaughter, uncle, aunt, nephew, niece, family, descendants, descendant, biographer, pupils, pupil, students, student.

- Issuers association

String search examples: authentication, board, estate, foundation(s), fundament, stiftung, fondation, fundación, fundação, fondazione, association, vereniging, verband, asociación, associação, associazione, committee, commissie, ausschuss.

- Issuers expert

String search examples: Dr, Prof, curator(s), custodian(s), professor(s), doctor(s), director(s), expert(s), expertise(s), professoren, professore, professoressa, professeur(s), professore(s), professori, profesor.

---

## Appendix C. Impact on Probability of Being Sold After Discoveries of Fakes

Dependent variable	(1) Sold[0,1]	(2) Sold[0,1]	(3) Sold[0,1]	(4) Sold[0,1]	(5) Sold[0,1]	(6) Sold[0,1]
<i>Treated</i>	−0.0319 (0.0262)	−0.0247 (0.0248)	−0.0094 (0.0205)	−0.0204 (0.0208)	0.0004 (0.0191)	−0.0233 (0.0249)
<i>Treated</i> × <i>Provenance</i>	0.0514** (0.0234)					
<i>Treated</i> × <i>Pedigree</i>		0.0466** (0.0212)				0.0089 (0.0219)
<i>Treated</i> × <i>Exhibition</i>			0.0364** (0.0145)			0.0017 (0.0120)
<i>Treated</i> × <i>Literature</i>				0.0666*** (0.0182)		0.0617*** (0.0187)
<i>Treated</i> × <i>Certification</i>					−0.0214 (0.0331)	−0.0021 (0.0358)
<i>Provenance</i>	0.0235*** (0.0055)					
<i>Pedigree</i>		0.0242*** (0.0063)				0.0172*** (0.0062)
<i>Exhibition</i>			0.0496*** (0.0053)			0.0379*** (0.0041)
<i>Literature</i>				0.0402*** (0.0065)		0.0236*** (0.0062)
<i>Certification</i>					0.0085 (0.0135)	0.0104 (0.0134)
<i>Ln(Low Price Estimate)</i>	−0.0650*** (0.0036)	−0.0647*** (0.0036)	−0.0652*** (0.0036)	−0.0652*** (0.0036)	−0.0641*** (0.0035)	−0.0662*** (0.0036)
Artist fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Hedonic controls	Yes	Yes	Yes	Yes	Yes	Yes
Auction house branch fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	1,707,136	1,707,136	1,707,136	1,707,136	1,707,136	1,707,136
R <sup>2</sup>	0.1747	0.1747	0.1748	0.1747	0.1745	0.1750

Notes. This table presents the DiD estimators for the discoveries of fakes cases. The dependent variable is the outcome of the sale (sold or unsold). *Treated* equals one if the auction date falls after the date of discovery of fakes for artists whose paintings have been forged. *Provenance* (*Pedigree*, *Exhibition*, *Literature*, and *Certification*) is the dummy variable equaling one if the paintings contain any of these types of provenance information. As controls, all hedonic variables discussed in Section 1 are included. In columns (1)–(5), the interaction terms of *Treated* with each provenance dimension (*Provenance*, *Pedigree*, *Exhibition*, *Literature*, and *Certification*) are presented separately, whereas in column (6), all these interactions are combined. All regressions include hedonic controls (see Section 2.1.1), and artist, year, month, and auction house branch level (AH) fixed effects. Standard errors (S.E.) are reported in parentheses and clustered at the auction branch level.

\*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

**Appendix D. Hedonic Pricing Regressions for Subsamples of Schools and Movements**

Panel A: Hedonic pricing regressions with provenance dummy for subsamples of schools and movements													
Sample	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Dependent variable	School 1 Ln(Price)	School 2 Ln(Price)	School 3 Ln(Price)	School 4 Ln(Price)	School 5 Ln(Price)	School 6 Ln(Price)	School 7 Ln(Price)	School 8 Ln(Price)	School 9 Ln(Price)	School 10 Ln(Price)	School 11 Ln(Price)	School 12 Ln(Price)	School 13 Ln(Price)
Provenance	0.2751*** (0.0487)	0.3505*** (0.0319)	0.3733*** (0.0532)	0.1849*** (0.0573)	0.3923*** (0.0439)	0.3569*** (0.0514)	0.3401*** (0.0304)	0.2460*** (0.0341)	0.2212*** (0.0389)	0.2123*** (0.0537)	0.2055*** (0.0345)	0.6051*** (0.1296)	0.2417*** (0.0596)
Artist fixed effects	Yes	Yes	Yes	Yes									
Year fixed effects	Yes	Yes	Yes	Yes									
Month fixed effects	Yes	Yes	Yes	Yes									
Auction house branch fixed effects	Yes	Yes	Yes	Yes									
Hedonic controls	Yes	Yes	Yes	Yes									
Number of observations	4,297	17,290	4,463	2,502	8,762	9,626	22,181	17,180	10,959	13,761	15,538	8,691	12,739
R <sup>2</sup>	0.6733	0.6426	0.6548	0.6453	0.6935	0.7149	0.7756	0.7765	0.7772	0.7663	0.7575	0.7628	0.7668
Panel B: Hedonic pricing regressions with four provenance dimensions for subsamples of schools and movements													
Sample	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Dept. Var.:	School 1 Ln(Price)	School 2 Ln(Price)	School 3 Ln(Price)	School 4 Ln(Price)	School 5 Ln(Price)	School 6 Ln(Price)	School 7 Ln(Price)	School 8 Ln(Price)	School 9 Ln(Price)	School 10 Ln(Price)	School 11 Ln(Price)	School 12 Ln(Price)	School 13 Ln(Price)
Pedigree	0.1439*** (0.0434)	0.1672*** (0.0281)	0.1243** (0.0523)	0.0291 (0.0598)	0.2457*** (0.0535)	0.1639*** (0.0475)	0.1801*** (0.0322)	0.1004** (0.0393)	0.1039*** (0.0349)	0.1188* (0.0648)	0.1534*** (0.0343)	0.5391*** (0.1171)	0.1180 (0.0726)
Exhibition	0.3504*** (0.0811)	0.4265*** (0.0804)	0.4145*** (0.0874)	0.1884* (0.1017)	0.3433*** (0.0851)	0.3707*** (0.0554)	0.4456*** (0.0304)	0.3088*** (0.0381)	0.3040*** (0.0420)	0.3771*** (0.0397)	0.1927*** (0.0299)	0.4790*** (0.0480)	0.3483*** (0.0205)
Literature	0.4245*** (0.0777)	0.3719*** (0.0380)	0.4613*** (0.0683)	0.4719*** (0.1126)	0.3782*** (0.0884)	0.4247*** (0.0438)	0.2947*** (0.0404)	0.3124*** (0.0432)	0.4161*** (0.0623)	0.4683*** (0.0530)	0.5163*** (0.0401)	0.4366*** (0.0729)	0.4414*** (0.0842)
Certification	-0.2583* (0.1363)	0.2022*** (0.0494)	0.2767** (0.1332)	-0.0652 (0.1986)	0.4481*** (0.0898)	0.2122*** (0.0719)	0.1423*** (0.0417)	0.1311*** (0.0411)	0.1882*** (0.0505)	0.0991** (0.0455)	-0.0065 (0.0446)	0.1042* (0.0629)	0.2566*** (0.0647)
Artist fixed effects	Yes	Yes	Yes	Yes									
Year fixed effects	Yes	Yes	Yes	Yes									
Month fixed effects	Yes	Yes	Yes	Yes									
Auction house branch fixed effects	Yes	Yes	Yes	Yes									
Hedonic controls	Yes	Yes	Yes	Yes									
Number of observations	4,297	17,290	4,463	2,502	8,762	9,626	22,181	17,180	10,959	13,761	15,538	8,691	12,739
R <sup>2</sup>	0.6834	0.6513	0.6651	0.6508	0.6998	0.7223	0.7820	0.7807	0.7836	0.7767	0.7663	0.7757	0.7754

*Notes.* We match and classify artists into 13 major movements in art history based on the biography information from Oxford Art Online: (1) Medieval and Renaissance; (2) Baroque; (3) Rococo; (4) Neoclassicism; (5) Romanticism; (6) Realism; (7) Impressionism and Symbolism; (8) Fauvism and Expressionism; (9) Cubism, Futurism, and Constructivism; (10) Dada and Surrealism; (11) Abstract Expressionism; (12) Pop Art; and (13) Minimalism and Contemporary. The dependent variable is  $Ln(Price)$ , the natural logarithm of the real hammer price in real (2007) US\$. In panel A, *Provenance* is a dummy variable capturing if the artwork's catalogue comprises any provenance information. In panel B, *Pedigree*, *Exhibition*, *Literature*, and *Certification* are dummy variables if the artworks have any information on this provenance dimension, respectively. All regressions include hedonic controls (see Section 2.1.1), and artist, year, month, and auction house branch level (AH) fixed effects. Standard errors (S.E.) are reported in parentheses and clustered at the auction branch level. \*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

## Endnotes

<sup>1</sup> One of the art market's greatest challenges for paintings by dead artists is to verify their authenticity and provenance. The Fine Arts Expert Institute in Geneva claims that more than 50% of the artworks circulating on the market are either forged or cannot be attributed to the correct artist. Source: <https://news.artnet.com/market/over-50-percent-of-art-is-fake-130821>.

<sup>2</sup> The descriptive statistics for the hedonic variables are presented in Online Table A.I.

<sup>3</sup> It is difficult to separate the effects of glamour and artistic quality. On the one hand, ownership by a celebrity may induce a sales premium, but the art object itself might be of high quality in that the celebrity might have been well-advised (by art experts) upon original purchase of the work of art. For example, the fact that a painting has been owned by Oprah Winfrey or Elton John at one point in the painting's history might be a selling point to potential buyers who feel sympathetic toward a star. However, it is unclear whether, if a premium is paid, the premium reflects stardom in past ownership or the possibility that a star is able to pick quality art with high value growth potential as they can obtain advice from the best art consultants.

<sup>4</sup> The string searches and sources are reported in Appendix B.

<sup>5</sup> Examples are as follows: "acquired directly from the artist by the present owner in 2002," "courtesy of the artist," and "gift from the artist."

<sup>6</sup> Examples are as follows: "by descent in the family of the artist until the late 1980s," "descended within the family of the artist," and "purchased from the artist's family."

<sup>7</sup> Examples are as follows: "from the sitter, by descent to the present owner," "by descent in the sitter's family until 2010," and "by descent through the sitter's grandson, Montague Peregrine Almarle, 12th Earl of Lindsey (1861–1938) to his daughter, the late Lady Muriel Barclay-Harvey (1893–1980)."

<sup>8</sup> Examples are as follows: "by family descent for three generations," and "by descent in the family."

<sup>9</sup> Examples are as follows: "Christie's, London, 25 January 1991, lot 20," and "Sotheby's London, Russian Pictures, Icons and Russian Works of Art, 15 February 1984, Lot 106."

<sup>10</sup> Examples are as follows: "Bonhams, London, 14 June 2005, lot 109," and "Phillips, London, 14 June 2000, lot 60."

<sup>11</sup> See examples in Appendix B.

<sup>12</sup> Examples are Georges Petit, Gagosian Gallery, Pierre Matisse, Sidney Janis, and Leo Castelli.

<sup>13</sup> Precise definitions of these ownership categories (e.g., for nobility, celebrities, and influential people) can be found in Appendix B.

<sup>14</sup> Prominent exhibitions include retrospectives, anniversary exhibitions (birth/death year of artists), biannual, and triannual exhibitions. Prominent art fairs include TEFAF, Art Basel, Art Miami, Biennale, and Frieze, whereas prominent museums include Getty, Louvre, and Museo del Prado. Cultural cities include New York City, Paris, and London. The detailed keywords and sources are given in the Appendix B.

<sup>15</sup> Examples are Oxford University Press and Cambridge University Press. The sources are in Appendix B.

<sup>16</sup> Examples are the artist's foundation and registry. For detailed definitions, see Appendix B.

<sup>17</sup> Examples are "a certification by the Picasso Administration will be given to the buyer," "accompanied by a certificate of authenticity from the artist," "accompanied with Solomon Gallery exhibition catalogue as well as letter written by the artist in 2002 following the purchase of the work," and "Registered in the artist's archive in Paris, no, 86170 SWF."

<sup>18</sup> Examples are "verbally authenticated by Dr. Paul Vogt, Essen" and "verbally authenticated by the Ludwig von Hofmann Archive, Zurich."

<sup>19</sup> Even over a long time period spanning decades, the number of repeat sales is limited to about merely 3%–5% of the total sales as the average holding periods are extensive (e.g., by collectors) and some art objects never return to the market (e.g., art bought by museums or private collectors) (Renneboog and Spaenjers 2013).

<sup>20</sup> See Appendix A for details.

<sup>21</sup> In Equations (1) and (2), the  $\delta$  lag also applies in that all provenance information is available prior to the auction. By inserting  $\delta$  in Equation (3), we emphasize that the change information affects changes in prices and, hence, returns.

<sup>22</sup> The parameter estimates of the full regression are provided in Online Table A.II.

<sup>23</sup> If the painting was at one point directly purchased from the sitter or the sitter's family, the probability of being sold is 8.9% lower, which may be affected by the scarcity of such paintings in auctions. Furthermore, such paintings (usually commissioned by the sitter) are of personal value for the sitter or the family and may be less appealing to people not related to the sitter's family.

<sup>24</sup> The sample size used to estimate the models of columns (3) and (5) of Table 2 is smaller than that of columns (1) and (2) because the sample for the latter models includes all paintings offered for sale (including those who were bought in).

<sup>25</sup> The models presented in panels A and B of Table 2 show only the parameter estimates of the variables capturing provenance details. The coefficients of the hedonic variables and fixed effects (included in Equation (2)) are given in Online Table A.III. Those results are in line with past research (e.g., Renneboog and Spaenjers 2013). Artworks with the attributions "style," "after," "school," "circle," "studio," and "attributed" are priced at large discounts (relative to the price of an authenticated artist), whereas signed, dated, or inscribed works tend to have higher prices. Oil paintings and watercolors command higher prices than drawings. Furthermore, prices increase with size (measurements), up to the point that the work becomes too large, which is captured by the negative coefficients on the squared terms. In addition, portraits and studies are sold at a discount. Unsurprisingly, Sotheby's London and Christie's London sell artworks with the highest prices, *ceteris paribus*.

<sup>26</sup> In some countries (e.g., the United Kingdom), it is possible that the fiscal authorities accept the donation of art to public museums as payment of inheritance tax.

<sup>27</sup> Bocart and Oosterlinck (2011) show that fraud discoveries shift the market toward intermediaries with higher reputation (e.g., Christie's and Sotheby's). Given that high-end auction houses provide insurance in that they offer to repurchase an art object if it is shown after the auction to be a counterfeit, they have a strong interest to collect detailed and reliable provenance information. Thus, the provision of provenance not only protects an auction house against claims and court cases, but also helps the intermediary to build a reputation of trustworthiness.

<sup>28</sup> The sample of fake and forgery cases can be found in Online Table A.IV.

<sup>29</sup> See opinion and order of *Waren v. Christie's Inc.* 16cv1386 signed by Judge William H. Pauley, III on May 31, 2018, at the Southern District of New York: U.S. Christie's amended its provenance policy in 2012 such that "... consignors were required to submit verifiable documentation to substantiate the Provenance and Country of Origin (PCOO) Form. Verifiable documentation could include receipts, invoices, inheritance documents, insurance listings, photographs, letters, or any other source that the clients can provide above and beyond their own testimony." Source: <https://law.justia.com/cases/federal/district-courts/new-york/nysdce/1:2016cv01386/453955/112/>.

<sup>30</sup> We exclude between-sales periods of less than 180 days to avoid speculative transactions. As pointed out by Pénasse et al. (2021), a short holding period usually indicates that a “flipper” (speculator) is able to purchase at a low price and quickly offers the object at a higher price. For this type of speculator, the selling decision may be more endogenous and is driven by recent prices for similar paintings.

<sup>31</sup> We present only the analysis on the four provenance dimensions in Table 7 and also provide an analysis with detailed provenance elements in Online Table A.V.

<sup>32</sup> For auction house details, see Appendix A.

<sup>33</sup> We create various name patterns for all the names in our lists, for example, with fully spelled out first name, with initials, and with and without middle names.

<sup>34</sup> See [https://en.wikipedia.org/wiki/Royal\\_and\\_noble\\_ranks](https://en.wikipedia.org/wiki/Royal_and_noble_ranks).

<sup>35</sup> See [https://en.wikipedia.org/wiki/List\\_of\\_wealthiest\\_families](https://en.wikipedia.org/wiki/List_of_wealthiest_families).

<sup>36</sup> See [https://en.wikipedia.org/wiki/List\\_of\\_chief\\_executive\\_officers](https://en.wikipedia.org/wiki/List_of_chief_executive_officers).

<sup>37</sup> See [https://en.wikipedia.org/wiki/Category:British\\_television\\_personalities](https://en.wikipedia.org/wiki/Category:British_television_personalities).

<sup>38</sup> See [https://en.wikipedia.org/wiki/Category:American\\_television\\_talk\\_show\\_hosts](https://en.wikipedia.org/wiki/Category:American_television_talk_show_hosts).

<sup>39</sup> See [https://en.wikipedia.org/wiki/European\\_Capital\\_of\\_Culture](https://en.wikipedia.org/wiki/European_Capital_of_Culture).

<sup>40</sup> See [https://en.wikipedia.org/wiki/List\\_of\\_university\\_presses](https://en.wikipedia.org/wiki/List_of_university_presses).

## References

- Adams RB, Kräussl R, Navone M, Verwijmeren P (2021) Gendered prices. *Rev. Financial Stud.* 34(8):3789–3839.
- Beggs A, Graddy K (2009) Anchoring effects: Evidence from art auctions. *Amer. Econom. Rev.* 99(3):1027–1039.
- Belloni A, Chernozhukov V, Hansen C (2014) High-dimensional methods and inference on structural and treatment effects. *J. Econom. Perspect.* 28(2):29–50.
- Bocart F, Oosterlinck K (2011) Discoveries of fakes: Their impact on the art market. *Econom. Lett.* 113(2):124–126.
- Cameron L, Goetzmann WN, Nozari M (2019) Art and gender: Market bias or selection bias? *J. Cultural Econom.* 43(2):279–307.
- Chohan UW (2021) Non-fungible tokens: Blockchains, scarcity, and value. Critical Blockchain Research Initiative (CBRI) Working Paper, Islamabad, Pakistan.
- Deloitte Luxembourg and ArtTactic (2014) Art & finance report 2014. Accessed February 7, 2022, [https://www2.deloitte.com/content/dam/Deloitte/es/Documents/acerca-de-deloitte/Deloitte-ES-Opera\\_Europa\\_Deloitte\\_Art\\_Finance\\_Report2014.pdf](https://www2.deloitte.com/content/dam/Deloitte/es/Documents/acerca-de-deloitte/Deloitte-ES-Opera_Europa_Deloitte_Art_Finance_Report2014.pdf).
- Deloitte Luxembourg and ArtTactic (2016) Art & finance report 2016. Accessed February 7, 2022, <https://www2.deloitte.com/content/dam/Deloitte/lu/Documents/financial-services/artandfinance/lu-en-artandfinancereport-21042016.pdf>.
- Fairfield J (2022) Tokenized: The law of non-fungible tokens and unique digital property. *Indiana Law J.* 97(4):1261–1313.
- Goetzmann WN, Renneboog L, Spaenjers C (2011) Art and money. *Amer. Econom. Rev.* 101(3):222–226.
- Graddy K, Loewenstein L, Mei JP, Moses M, Pownall RAJ (2023) Empirical evidence of anchoring and loss aversion from art auctions. *J. Cultural Econom.* 47:279–301.
- Guiso L, Sapienza P, Zingales L (2008) Trusting the stock market. *J. Finance* 63(6):2557–2600.
- Korteweg A, Kräussl R, Verwijmeren P (2016) Does it pay to invest in art? A selection-corrected returns perspective. *Rev. Financial Stud.* 29(4):1007–1038.
- Li Y, Ma M, Renneboog L (2022) Pricing art and the art of pricing: On returns and risk in the art auction markets. *Eur. Financial Management* 28(5):1139–1198.
- Lovo S, Spaenjers C (2018) A model of trading in the art market. *Amer. Econom. Rev.* 108(3):744–774.
- Mei J, Moses M (2002) Art as an investment and the underperformance of masterpieces. *Amer. Econom. Rev.* 92(5):1656–1668.
- Pénasse J, Renneboog L (2022) Speculative trading and bubbles: Evidence from the art market. *Management Sci.* 68(7):4939–4963.
- Pénasse J, Renneboog L, Scheinkman J (2021) When a master dies: Speculation and asset float. *Rev. Financial Stud.* 34(8):3840–3879.
- Pénasse J, Renneboog L, Spaenjers C (2014) Sentiment and art prices. *Econom. Lett.* 122(3):432–434.
- Pownall R (2017) TEFAF Art Market annual report 2017. The European Fine Art Foundation. Accessed February 7, 2022, <http://made2measure.org/tefaf/amr2017/>.
- Renneboog L, Spaenjers C (2013) Buying beauty: On prices and returns in the art market. *Management Sci.* 59(1):36–53.
- Roberts MR, Whited TM (2013) Endogeneity in empirical corporate finance. Constantinides G, Harris M, Stulz R, eds. *Handbook of the Economics of Finance*, vol. 2 (Elsevier, Amsterdam), 493–572.
- Thompson D (2010) *The \$12 Million Stuffed Shark: The Curious Economics of Contemporary Art* (Aurum Press, London).