

# Understanding & Quantifying How Paintings are “Valued”

*Jessie Bierschenk*

## Introduction

The analyzation of paintings has appeared across various conversations in art, politics, history, and economics, but a question still exists: How do we put a price (as a proxy for value) on such a personal and arbitrary good? Business professionals might say that [appraisals are based largely on what the art has already sold for or typically sells](#); art historians might say that [value is determined by its cultural significance within a given time period](#). Yet the question still remains of what objective qualities inherent to a work add value? In the hopes of further exploring such question, this analysis focuses on art auction sales and, by using anonymous works, attempts to strip paintings of their intangible ties to fame or reputation in order to isolate and identify the objective qualities that are associated with art prices.

[Previous research](#) has identified evidence of a strong association between formal naming conventions used by art historians and final auction prices (known as hammer prices). Today, artists are easily identifiable as it has become common practice to sign artwork. In the past (1350-1850), close attention was typically paid to the provisional name of the creator and its long-term recognition, typology, etc. Art markets are characterized by strong asymmetry of information and high uncertainty surrounding quality, leading to difficulty in accurate pricing and therefore emphasizing the importance of the reputation of the creator. To circumvent this reliance on reputation, this analysis will focus on sales of anonymous paintings in an adaption of a dataset from [Anne-Sophie Radermecker’s work on naming conventions and anonymous art](#). The analysis aims to present findings that are easily interpretable for players in the art markets not familiar with highly complex statistical method. The primary research question is as follows:

**(1) What objective, indisputable qualities of a painting are associated with it’s value (price) when the connection to its creator is not applicable or controlled for?**

The dataset in this analysis includes 13,211 observations of individual painting sales of 15th-17th century Flemish work and captures sales occurring as early as 1946 to as late as 2015. The data include extensive information for each piece including the naming convention, the working title, the hammer price in 2015 U.S. dollar amount, information about the medium and content of the piece, and information about the date, place, and type of the auction sale. Rademecker describes the full dataset being [mainly collected from the Hislop Art’s Sales Index, and have been systematically compared with data recorded by Artprice and the World Collectors Annuary to provide a comprehensive corpus of sources](#).

The dataset includes 1,581 sales of purely anonymous paintings and 11,630 sales of paintings with either a distinct author or some attribution that is expert-verified. [Previous research](#) suggests that “the artist’s name as a brand influences the general assessment of a work and, its economic value and what an art collector is willing to pay for it.” Therefore, as mentioned in [Rademecker’s research](#), market players wanting to resolve the issue of anonymity in art employ naming strategies developed by art historians to give alternative names to unrecorded artist (attributions). Listed below are conventional naming mechanisms widely used and accepted in contemporary art markets:

- “By” or “By and studio” are used in conjunction with a name when there are records, material, and formal and stylistic evidence of a piece’s creator. This is the most specific level of identification and the most “profitable”.
- “After” is considered to be a replication of a known artist’s work.
- “Attributed to”, “Studio of”, and “Circle of” are the next most specific attributes used to indirectly link an anonymous painting to a known artist, style, or school through a verifiable method.
- “Manner of”, “School of”, “Style of”, and “Follower of” are all thought to be the least specific verified attributions types before simply deemed “anonymous”.

The more specific the attribution, the more valuable a piece is to a buyer. Art that cannot be given some sort of attribute is considered anonymous. According to Rademecker, there is little work done in art history or economic research that explores this relationship of attributes with price. And, given the guarded and secretive nature art markets, it is difficult to access auction/price data and even more difficult to access information on anonymous, obscure paintings not tied to any creator (attribute); because of this, this analysis will use the unique opportunity to analyze anonymous work (as well as attributed work) to form a more comprehensive understanding of art valuations. The main response variable of interest in this analysis is the median hammer price (To allow for inflation, prices are deflated using the US Consumer Price Index, and expressed in 2015 US dollars). The median is the intended interest because [meticulous rounding up by auctioneers and theatrical tactics](#) can create highly skewed hammer prices; thus, the median is thought to be a more reflective summary statistic.

**Table 1** shows the frequency of attribute types in our data set. 12% of the pieces in this dataset are considered anonymous while the rest have some attribute.

Table 1: Frequency of Attributes in Data

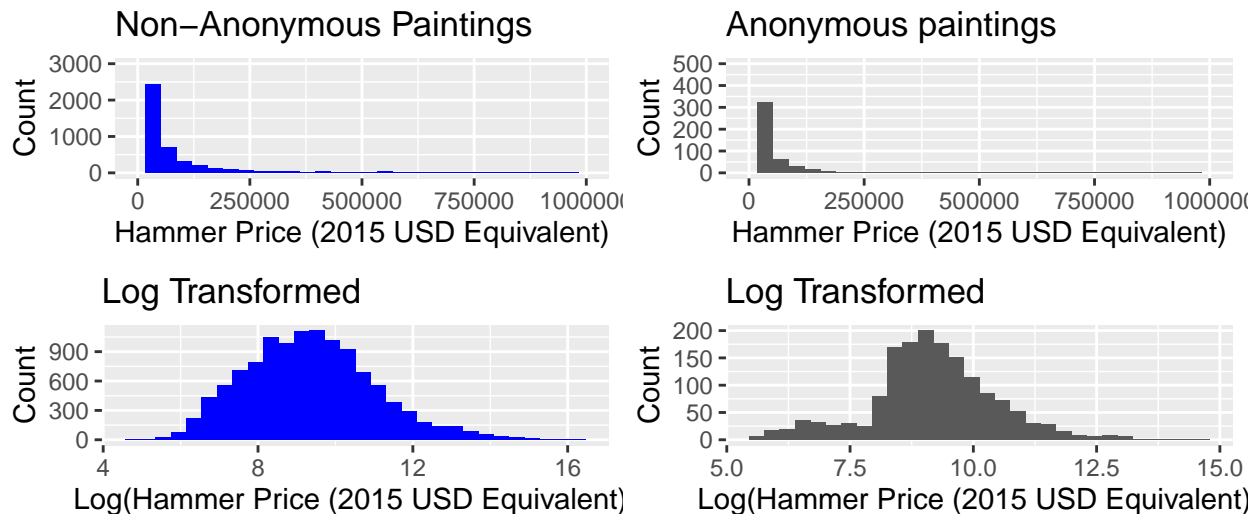
	Anonymous	After	Attributed to	By	Circle of	Follower of	School of	Studio of	Style of	Total
Count	1581	316	1489	5490	1558	187	312	597	1673	13211
Percent	12.0%	2.4%	11.3%	41.6%	11.8%	1.4%	2.4%	4.5%	12.7%	100%

## Exploratory Data Analysis

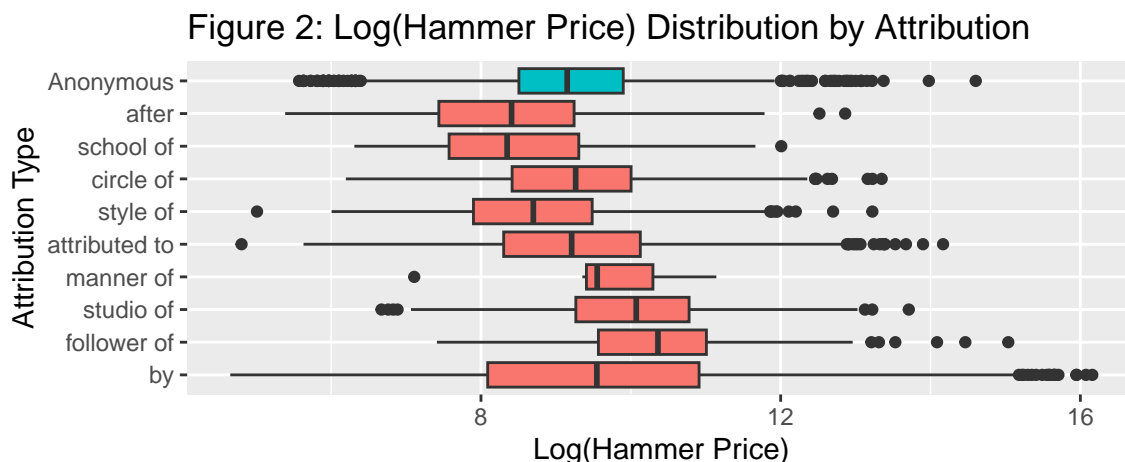
Our response variable of interest is the median hammer price in 2015 US dollars. As shown in **Figure 1**, both anonymous and non-anonymous paintings had a relatively similar shape of price distributions with significant right skew. For non-anonymous paintings, the minimum hammer price in 2015 US dollar equivalent was \$105 and the maximum was \$10,423,009; the median hammer price was \$10,589. For anonymous paintings, the minimum hammer price in 2015 US dollar equivalent was \$263 (higher than non-anonymous) and the maximum was \$2,200,000 (significantly lower than non-anonymous); the median hammer price was \$9,420 (>\$1,000 less than non-anonymous). This general trend is as expected as the price of anonymous art is not driven up by any particular artist/attribution the way that non-anonymous art is. For both categories, most hammer prices lie in the \$10k - \$50k range.

**Figure 1** shows the log-transformations of the hammer price distribution. Because these distribution are originally right skewed, a log transformation can reduce skewness in the response and the response more closely reflects a normal distribution shape.

Figure 1: Comparison of Hammer Price Distributions



As discussed previously, research has presented evidence of differences in hammer prices based on attribution status. In **Figure 2**  $\log(\text{hammer price})$  distributions differ by associated attribution for paintings. The log transformed version is shown to more easily visualize distributions. Paintings that are replications or “copies” of original work appear to have the lowest interquartile range. It’s interesting to note that even anonymous paintings have a higher interquartile range than paintings known to not be original. The hammer price distribution associated with by has the largest range; this is likely because the direct link to a particular artist is likely also a direct link to the clout of a particular artist which may significantly vary on a case-by-case basis.



For some objective characteristics, there are distinct differences in between non-anonymous and anonymous paintings. As shown in **Figure 1A**, non-anonymous paintings tend to have more diversity in the paintings’ subject while most of the anonymous work is concentrated in religious scences. This is, in part, due to the [growing demand for religious and decorative pictures at the time of the 15th -17th century](#) but still an interesting distinction to make. In terms of similarities, nativity scenes were the most common subject for both anonymous and non-anonymous art followed by portraits and haigographic scenes (scenes depicting saints).

**Figure 2A** illustrates how the median hammer price according to the paintings’ subject do not have the same ordinal pattern across anonymous and non-anonymous paintings. For example, the subject matter with the lowest relative median hammer price in non-anonymous art is portraits whereas, among anonymous paintings, portraits have one of the highest relative median hammer prices. Thus, subject matter may be associated with hammer price differently depending on whether a piece is anonymous or not.

**Figure 3A** illustrates how the  $\log(2015 \text{ USD equivalent hammer prices})$  increased with the date of the recorded sale. As these hammer prices are adjusted for inflation using U.S. CPI, this trend suggests a greater demand/willingness to pay for paintings as the art economy continued to evolve and became more and more sophisticated.

An important thing to note is that not all characteristics differ depending on the anonymous status of a piece. Some relationships are consistent across art regardless of attribution. Through this analysis, we ensure that these characteristics are controlled for and included as we begin our modeling process. **Figure 4A** illustrates that despite the anonymous status of pieces, paintings in specified collection types tend to have higher price distributions than paintings with no information of collection type, and sale records indicate that evening sales tend to have higher price distributions than day sales.

These findings are all informative for our selection of variables in the model (described in the next section).

# Methodology

In order to distill art valuations to the objective aspects of the art itself, we will create and analyze two models: one model fit to the data only concerning anonymous paintings, and the other fit to the data only concerning paintings with some attribution. These models will (mostly) include the same independent variables. With the response variable (hammer price) for both models being skewed, a log transformation will be done so the patterns in the data are more interpretable and the regression model better satisfies linear regression assumptions (**See Appendix for Model Assumptions**). This method was used as results can be easily interpreted by art historians/economists unfamiliar with more sophisticated methods. Given that hammer prices operate on many different orders of magnitude, quantile regression was an additional candidate model for this analysis. The relationships between the predictors on high, medium, and low quantiles, might be different and quantile regression can identify this. However, given the nature of the data and the goal of this analysis to explore multiple different objective qualities associated with hammer price, this analysis proceeded with the log-transformed model. Many of the categorical variables in the dataset have a higher concentration at a single level, and many of the categorical variables contain levels that can not be combined (i.e. binary variables, distinct school of the work, medium of the work, etc.) as each level is completely unique and there is no logical combination. The log-transformed model allows for analysis on more variables—a key point of investigation for this analysis. However, it is noted that quantile regression would be an interesting point of exploration for other datasets of the like in the future.

## (Manipulated) Variables Included

### Shared by Both Models

- The type of sale (day or evening) was included as, traditionally, compared to day sales, evening sales tend to [have fewer works being sold, and these pieces tend to sell at much higher price points](#).
- The associated surface area of the piece recorded in the sale is included because this characteristic is a purely objective fact of a work. Because this analysis is interested in objective factors (not tied to the clout or reputation of the creator) that may be associated with the value of art, we will include this. Furthermore, previous [research](#) shows evidence of the size of paintings having a positive relationship with price. We divided this variable by 500 and centered it at the mean surface area of the data set (4,956.9 sqm cm) to make for more interpretable results
- The adjunct century of the piece is included as an objective characteristic. This dataset includes 15th - 17th century work with 2,854 missing values for this variable initially. Upon closer examination, all but 3 observations were associated with a school from the 16th-17th century (mentioned in the details of a painting's associated school). With this, an additional category was made to account for 16th - 17th c. work and another category to reflect "missing".
- The associated artistic School which refers to a common [style, inspiration, training, or aims](#) was included. This, again, is an objective quality of a piece that may be associated with median hammer price.
- The subject or topic of the piece in a sale and the year of the sale: The subject of a piece is objective and an important distinction to be made when comparing against other pieces. The year of the sale is important in understanding the context of the purchase and the nature of the art economy at the time. Furthermore, as the [taste of art markets change](#) the prices reflect changes in demand. Therefore, we will also include an interaction term with date purchased and subject to reflect the evolution of preference and trends in the art economy. The date variable was centered at 1946 (the earliest observation in our dataset).
- The currency used in the sale: Currency may indirectly link to the buyer's cultural taste or country of origin. We will include this variable to control for potential preference.
- The collection type associated with the piece sold: Most of the sales do not include this information, but whether a piece was in a private collection (owned by a single entity), or in a museum collection (shown at a museum), etc, is an important factor to control for when understanding price as more well-known pieces may have been showcased more.

- The reputation of the auction house in the sale: This is to control for potential confounding factors driving hammer prices. The status of the auction house is often linked to the quality of works and also buyers.
- A binary variable was created to indicate if there are publications on the painting or not: Having more information or recognition of a piece may be associated with its value.
- A binary variable indicating if the painting was dated or not. Having more information about the exact time period of a piece may be associated with its median hammer price.
- A variable indicating the material the painting was on and another variable for the medium of the painting (i.e. oil or tempera): These are material and objective characteristics of a piece and are relevant to look at when understanding the value of a piece independent of its associated authorship status.
- The Consumer Price Index (CPI) is included as it gives context to the state of the (U.S.) economy during the time of the sale. As the U.S. economy is closely intertwined with other developed countries, this proxy is considered adequate for reflecting economic trends during the time of the sale. This variable was centered at the CPI during 1946 (8.3).
- A binary variable indicating if the painting had been in an exhibit or not. Such indication may control for status of a piece and thus what a buyer is willing to bid.

### Unique to Model for Anonymous Paintings

The attribute of “Anonymous” was created to account for pieces in which there was no formal attribution given. This distinction was used when differentiating data that was associated with anonymous and non-anonymous work. \* A binary variable was created to determine if the piece had been reattributed—meaning the original was deemed inaccurate by a relevant expert. For anonymous art, this would mean that the piece had been attributed to an artist or group, and then later determined that such attribution was inaccurate and now it is deemed an “anonymous” work.

### Unique to Model for Non-Anonymous Paintings

For our model fit to non-anonymous data, we will include the attribution type. As mentioned previously, these attributions (naming mechanisms) are important distinctions when considering the valuation of art. For example, a painting that is verified as being “by” a particular artist is more informative for a buyer than “attributed to”, “School of”, etc., and thus, [the buyer is likely willing to pay more](#). Pieces with the attribute “by & studio” were included in the “by” attribute since both can be associated with a specific artist. \* Indication of the type of reattribution a painting may have had: A painting may have had no reattribution, or a negative reattribution meaning that the piece used to have a more profitable attribution than the one it bears for the sale, or a positive reattribution meaning that the piece used to have a less profitable attribution than the one it bears for the sale. The anonymous paintings do not have positive reattributions because being “anonymous” is regarded as the least profitable attribution. \* Binary indication of whether a painting was examined via scientific technology for authentication purposes: This information further strengthens Only one painting in the anonymous data was associated with this so for the sake of simplicity, it will only be included in the non-anonymous data. \* Binary indication of whether a piece had a signature or not: This allows for greater authentication and information for a buyer, potential associated with median hammer price. Because anonymous works do not have a known author, this variable is unique to the non-anonymous model.

### Removed Variables

With such a large dataset initially with over 500 variables, variables that were redundant or not directly relevant to this analysis were omitted. See appendix for more explanation.

## Results

For this analysis two models are constructed. One fit to the data of non-anonymous work and the other fit to the data of anonymous work. Both will have the same response variable:  $\ln(\text{Hammer Price 2015 USD Equivalent})$ . Given the skew of the the USD hammer price, this transformation was deemed appropriated to reduce variability. The general formula (with

$$x_{ji}$$

as the respective included variable as seen above for the respective model) is as follows (**see appendix for full models**):

$$\text{Hammer.Price}_i = \exp(\beta_0)\exp(\beta_1x_{1i} + \beta_2x_{2i} + \beta_3x_{3i} + \dots)$$

Table 2: Anonymous Model

Variable	Coeff	95% CI Lower Bound	95% CI Upper Bound	P-Value	Sig
Intercept	38114.960	29982.759	48452.851	<0.001	Yes
Evening Sale (vs. Day)	5.534	3.533	8.666	<0.001	Yes
Surface Area of Painting	1.014	1.010	1.019	<0.001	Yes
16th-17th century (vs. 15th)	0.477	0.363	0.627	<0.001	Yes
Flemish School (vs. Antwerpen)	1.491	1.008	2.206	0.046	Yes
Landscape (vs. Religious topic)	0.993	0.812	1.215	0.947	No
Mythology (vs. Religious topic)	1.032	0.790	1.347	0.818	No
Purchased in Euro (vs. British Pound)	1.030	0.860	1.234	0.748	No
Purchased in US Dollars (vs. British Pound)	1.259	1.076	1.474	0.004	Yes
Museum Collection (vs. Not Specified)	1.311	0.619	2.779	0.480	No
Private Collection (vs. Not Specified)	1.485	1.261	1.749	<0.001	Yes
Year Sold	2.450	2.306	2.604	<0.001	Yes
2nd Tier Auction House (vs. Top Tier)	0.864	0.742	1.006	0.060	No
3rd Tier Auction House (vs. Top Tier)	0.666	0.548	0.810	<0.001	Yes
Publications on Painting (vs. None)	1.690	1.305	2.189	<0.001	Yes
Date on Painting (vs. None)	1.124	0.894	1.414	0.316	No
Tempera Medium (vs. Oil)	1.153	0.701	1.896	0.576	No
Canvas Material (vs. Panel)	0.742	0.639	0.862	<0.001	Yes
Copper Material (vs. Panel)	0.904	0.719	1.136	0.387	No
Had Previous Attribute (vs. Didn't)	0.996	0.748	1.327	0.979	No
Been in an Exhibit (vs. not)	0.982	0.727	1.327	0.907	No
Landscape * Year sold	1.173	0.968	1.421	0.103	No
Mythology * Year sold	1.278	1.020	1.601	0.033	Yes

**The baseline for this model can be found in the appendix** We will consider the median hammer price in our interpretations rather than the mean as unconventionally high hammer prices caused by extreme bidding distort the average. Note that “accounting for all other factors” in the following interpretations refers to the baseline metrics. A 95% confidence interval was also calculated for each coefficient estimate. From the model, some interesting interpretations from the model are:

- As expected, the median hammer price for an evening sale vs. a day sale is expected to multiply by a factor of 5.534, after accounting for all other factors.
- For every additional 500 sqr cm (just over 0.5 sqr ft) the median hammer price is expected to multiply by a factor of 1.014, after accounting for all other factors.

- The median hammer price for a purchase in USD vs. Pounds is expected to multiply by a factor of 1.259, after accounting for all other factors.
- The median hammer price for a purchase from a private collected vs. an unspecified collection is expected to multiply by a factor of 1.485, after accounting for all other factors.
- The median hammer price for an auction not at top tier auction houses is expected to be less than at a top tier, after accounting for all other factors.
- The median hammer price for a painting with publications on the work is expected to multiply by a factor of 1.690 compared to no publications, after accounting for all other factors. \*The median hammer price for painting on canvas is expected to multiply by a factor of 0.742 compared to a panel after accounting for all other factors.
- For each additional year above 1946, the median hammer price for a sale is expected to multiply by a factor of 2.450, after accounting for all other factors.
- For each additional year above 1946, the median hammer price for a mythology vs. a religious topic painting is expected to multiply by a factor of 1.278 , after accounting for all other factors.

Table 3: Non-Anonymous Model

Variable	Coeff.	95% CI Lower Bound	P-Value	95% CI Upper Bound	Sig
Intercept	991.128	849.502	<0.001	1156.364	Yes
Evening Sale (vs. Day)	3.223	2.638	<0.001	3.937	Yes
Surface Area of Painting	1.014	1.012	<0.001	1.016	Yes
16th-17th century (vs. 15th)	1.063	0.935	0.353	1.208	No
Flemish School (vs. Antwerpen)	1.454	1.122	0.005	1.885	Yes
Landscape (vs. Religious topic)	2.180	1.806	<0.001	2.632	Yes
Mythology (vs. Religious topic)	1.140	0.786	0.489	1.654	No
Purchased in Euro (vs. British Pound)	0.843	0.769	<0.001	0.924	Yes
Purchased in US Dollars (vs. British Pound)	1.343	1.254	<0.001	1.438	Yes
Museum Collection (vs. Not Specified)	0.913	0.588	0.686	1.418	No
Private Collection (vs. Not Specified)	1.205	1.111	<0.001	1.306	Yes
Year Sold	1.074	1.071	<0.001	1.077	Yes
2nd Tier Auction House (vs. Top Tier)	0.788	0.734	<0.001	0.845	Yes
3rd Tier Auction House (vs. Top Tier)	0.442	0.394	<0.001	0.495	Yes
Publications on Painting (vs. None)	1.562	1.404	<0.001	1.738	Yes
Date on Painting (vs. None)	1.234	1.121	<0.001	1.358	Yes
Tempera Medium (vs. Oil)	1.352	0.906	0.140	2.016	No
Canvas Material (vs. Panel)	0.744	0.695	<0.001	0.797	Yes
Attributed to (vs. By)	0.552	0.515	<0.001	0.592	Yes
Circle of (vs. By)	0.441	0.412	<0.001	0.472	Yes
Follower of (vs. By)	0.425	0.357	<0.001	0.505	Yes
Manner of (vs. By)	0.171	0.078	<0.001	0.374	Yes
School of (vs. By)	0.356	0.312	<0.001	0.407	Yes
Studio of (vs. By)	0.648	0.587	<0.001	0.716	Yes
Style of (vs. By)	0.304	0.284	<0.001	0.326	Yes
Examined via Scientific Tech. (vs. Not)	0.749	0.443	0.283	1.269	No
Signed (vs. Not)	1.585	1.449	<0.001	1.734	Yes
Been in an Exhibit (vs. not)	1.742	1.546	<0.001	1.963	Yes
Reattributed to a Less Profitable Attrib. (vs. not)	0.635	0.502	<0.001	0.803	Yes
Landscape * Year sold	0.988	0.984	<0.001	0.992	Yes

The baseline for this model can be found in the Appendix. Note that “accounting for all other

factors” in the following interpretations refers to the baseline metrics. From the model, some interesting interpretations are:

- Similar to the anonymous work, the median hammer price for an evening sale vs. a day sale is expected to multiply by a factor of greater than 1, after accounting for all other factors.
- For every additional 500 sqr cm (just over 0.5 sqr ft) the median hammer price is expected to multiply by a factor of 1.014, after accounting for all other factors. This is the same factor as seen in the anonymous model.
- As expected, the median hammer price for attributions considered less prestigious as “by” are all expected to multiply by a factor of less than one, with “after” and “manner of” having the lowest estimates, after accounting for all other factors.
- The median hammer price for pieces that have been in an exhibit vs. have not is expected to multiply by a factor of 1.742, after accounting for all other factors.
- Similar to the anonymous work, the median hammer price for an dated painting vs. not dated is expected to multiply by a factor of greater than 1, after accounting for all other factors.
- Similar to the anonymous work, the median hammer price for an painting on canvas vs. on a panel is expected to multiply by a factor of less than 1, after accounting for all other factors.
- For each additional year above 1946, the median hammer price for a sale is expected to multiply by a factor of 1.074, after accounting for all other factors. Like the anonymous model, this factor is greater than one.
- For each additional year above 1946, the median hammer price for a landscape vs. a religious topic painting is expected to multiply by a factor of 0.988, after accounting for all other factors. Unlike the anonymous model, this factor is less than one although the coefficient was not deemed significant in the anonymous model.

## Discussion

Some limitations to this analysis rest in the asymmetry of accessible data for anonymous vs. non-anonymous work. There was substantially more data for non-anonymous pieces potentially due to the fact that auctions, in general, prioritize the sales of attributed work over anonymous work, or potentially due to the fact that there are less anonymous Flemish paintings from the 15th - 17th c. in existence. Another limitation is that there is no evidence to believe that findings for these paintings (Flemish 15th - 17th century) can be extrapolated to other genres. The findings of this analysis are informative and enlightening, yet there is no evidence to suggest they are generalizable to other genres of work.

As seen in the interpreted results, the two fitted model for anonymous and non-anonymous works had both similar and dissimilar associations between median hammer price and independent variables after adjusting for baseline metrics. The purpose of creating two models was to be able to isolate factors associated with median hammer price that were not associated with the subjective reputation or clout of a painting’s creator. Even still, both the anonymous and non-anonymous model (at the baseline of highest attribution “by”) presented evidence for some similar associations between the response (median hammer price) and purely objective independent variables such as type of auction (evening or day), painting surface area, whether the painting was dated, material of the painting, and year of the sale. From the non-anonymous model, there is evidence to suggest that the attribution title is significantly associated with the median hammer price. These findings help art historians and auctioneers to contextualize how art is valued in contemporary markets and how, if stripped away of material prestige, there are inherent qualities of art tied to its worth. Analyzing anonymous art sales serves as a way to distill artwork (in this case Flemish paintings from the 15th-17th century) down to the pure qualities that are associated with it’s value.

Future analyses may explore the use of quantile regression for art prices given that hammer prices operate on many different orders of magnitude. The relationships between the predictors on high, medium, and low quantiles, might be different and quantile regression can identify this. While this analysis, for the sake of simplicity and digestibility, only looked at log-transformed regression analysis, further research may



explore differences in predictor-response relationships for different price quantiles (i.e. differences in predictor associations for cheap vs. expensive paintings).

Ultimately, with a significant lack of accessibility and transparency in auction houses and art markets, there remains an extensive amount of meaningful research to be done, especially in the area of anonymous work. Hopefully, analyses like this can be conducted on other areas of art and motivate art market players—who have a notorious reputation for elusive pricing and unobtainable records/data—to engage in more transparent transactions and art valuations.

## Appendix

Figure 1A: Comparison of Painting Subject  
Between Non-Anonymous and Anonymous Paintings

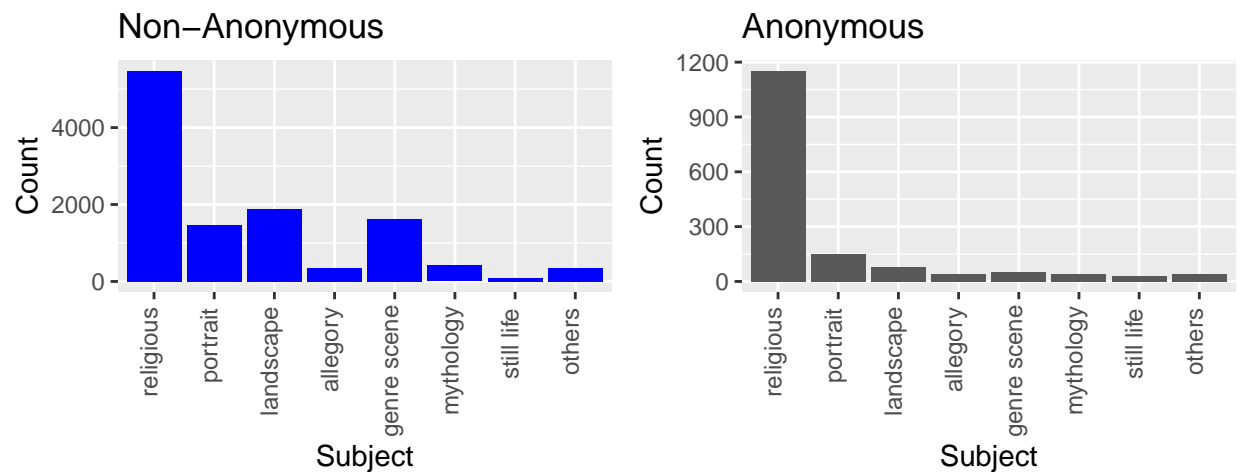


Figure 2A: Hammer Price Medians by Subject

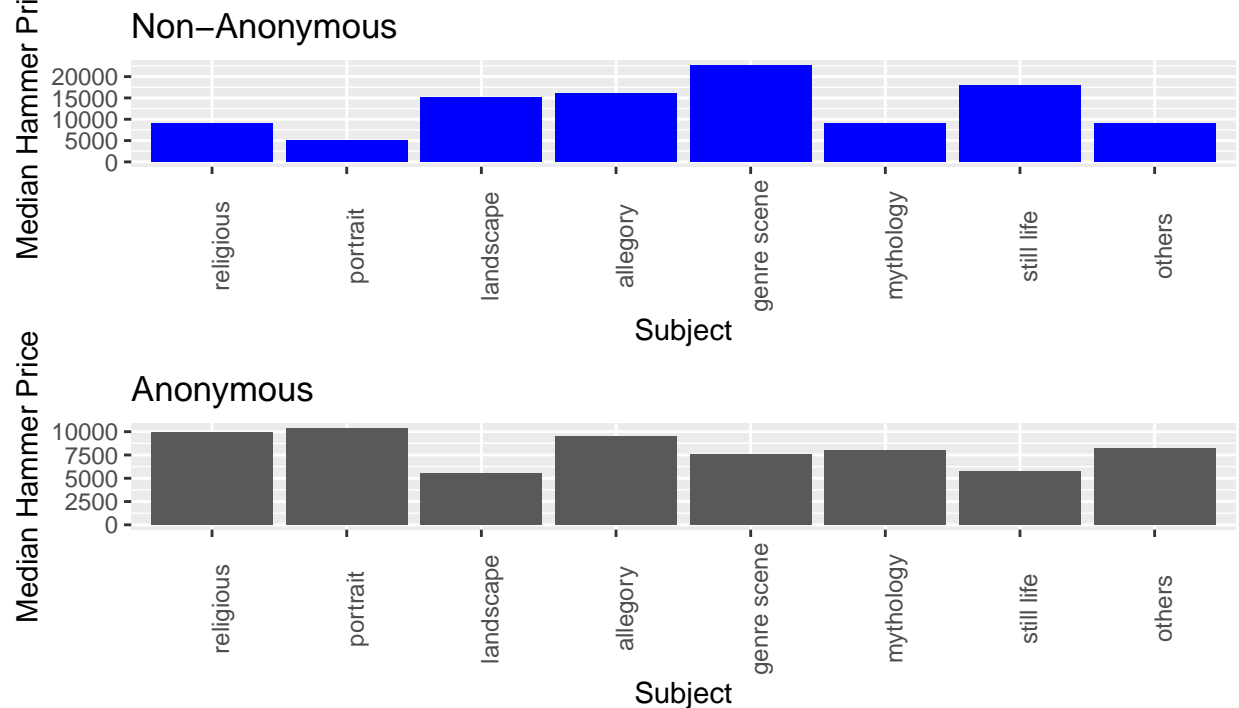


Figure 3A: Log(Hammer Price in 2015 USD) vs. Date of Sale

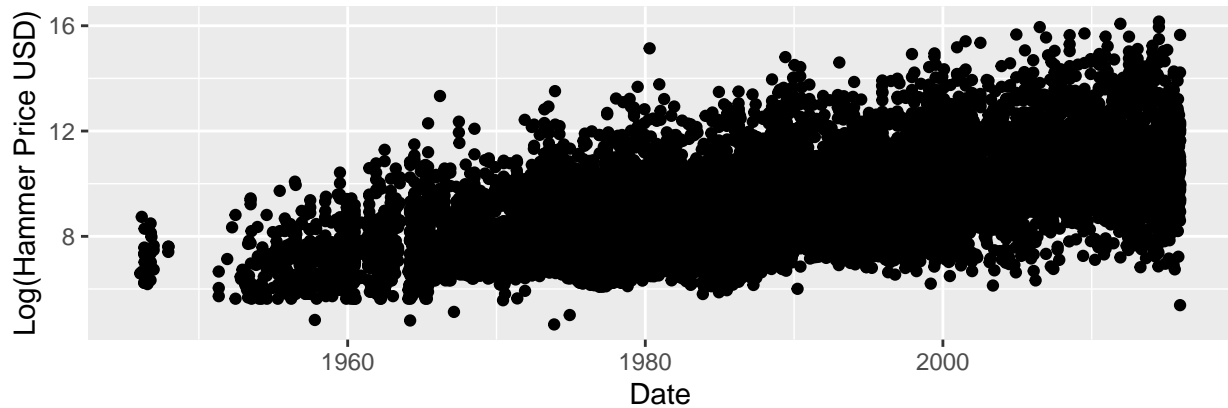
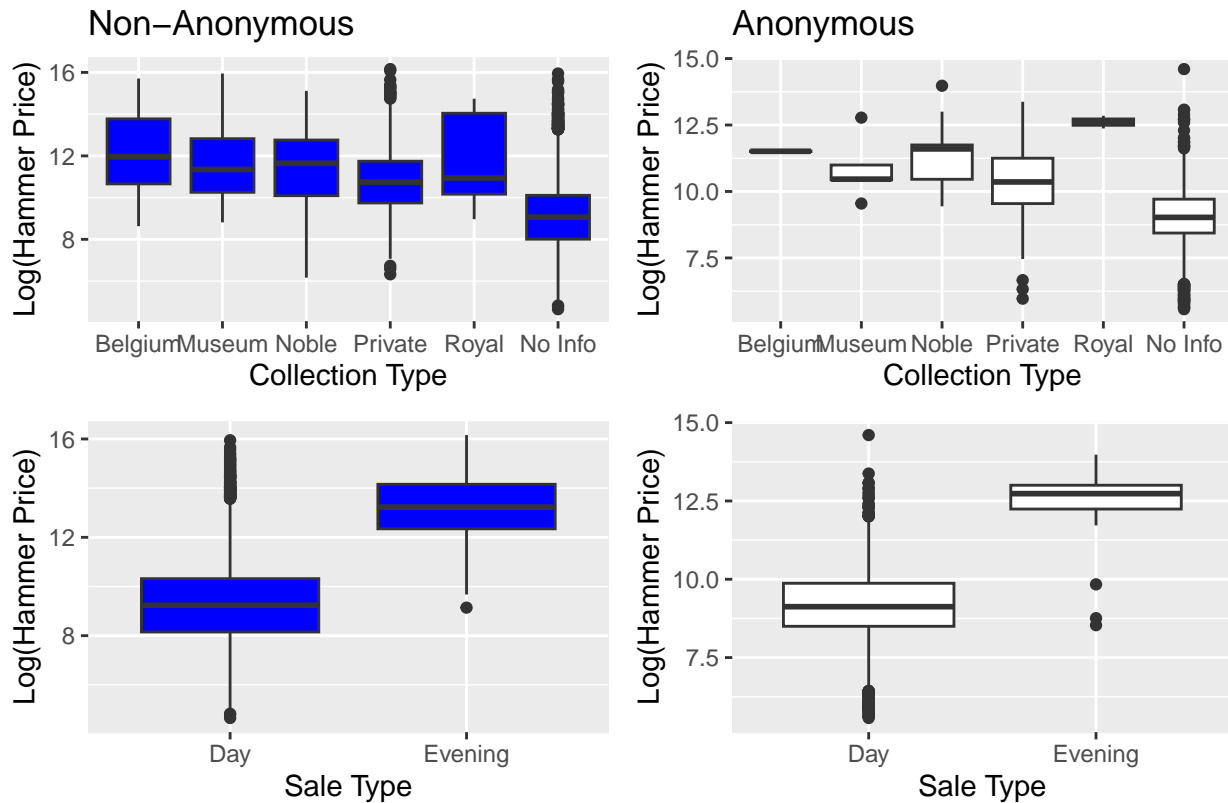


Figure 4A: Consistent Trends Seen in Sale Type and Collection Type



## Removed Variables

Firstly, variables removed from this analysis were redundant variables with different formatting or variable type (i.e. binary vs. categorical). Additionally, information about the class of “old Masters” (a name for a individual or type/school of art) was not used in this analysis. While Rademecker’s research was focused on associations between different Old Masters paintings, this analysis was more concerned with the distinction of purely anonymous versus formally attributed paintings. Lastly, variables that contained more specific details about a given auction were not used as it was believed that this could be adequately captured in the ordinal ranking (“tier”) of the auction house.

## Full Model Formulas

### Anonymous Work Model

$$\begin{aligned} \text{Hammer.Price}_i = & \exp(\beta_0) \exp(\beta_1 \text{Sales.Type}_i + \beta_2 \text{Surface.Area}_i + \\ & \beta_3 \text{Century.Made}_i + \beta_4 \text{School}_i + \beta_5 \text{Subject}_i + \beta_6 \text{Currency}_i + \\ & \beta_7 \text{Collection.Type}_i + \beta_8 \text{Year.Purchased}_i + \\ & \beta_9 \text{Auction.House.Prestige}_i) + \beta_{10} \text{Literature}_i + \beta_{11} \text{Dated}_i + \\ & \beta_{12} \text{Medium}_i + \beta_{13} \text{Material}_i + \beta_{14} \text{US.CPI}_i + \\ & \beta_{15} \text{Previously.Attributed}_i + \beta_{16} \text{Exhibit}_i + \\ & \beta_{17} (\text{Year.Purchased}_i * \text{Subject}_i) \end{aligned}$$

### Non-Anonymous Work Model

$$\begin{aligned} \text{Hammer.Price}_i = & \exp(\beta_0) \exp(\beta_1 \text{Sales.Type}_i + \beta_2 \text{Surface.Area}_i + \\ & \beta_3 \text{Century.Made}_i + \beta_4 \text{School}_i + \beta_5 \text{Subject}_i + \beta_6 \text{Currency}_i + \\ & \beta_7 \text{Collection.Type}_i + \beta_8 \text{Year.Purchased}_i + \\ & \beta_9 \text{Auction.House.Prestige}_i) + \beta_{10} \text{Literature}_i + \beta_{11} \text{Dated}_i + \\ & \beta_{12} \text{Medium}_i + \beta_{13} \text{Material}_i + \beta_{14} \text{US.CPI}_i + \\ & \beta_{15} \text{Reattribution.Type}_i + \beta_{16} \text{Exhibit}_i + \beta_{17} \text{Signature}_i + \\ & \beta_{18} \text{Scientific.Investigation}_i + \beta_{19} (\text{Year.Purchased}_i * \text{Subject}_i) \end{aligned}$$

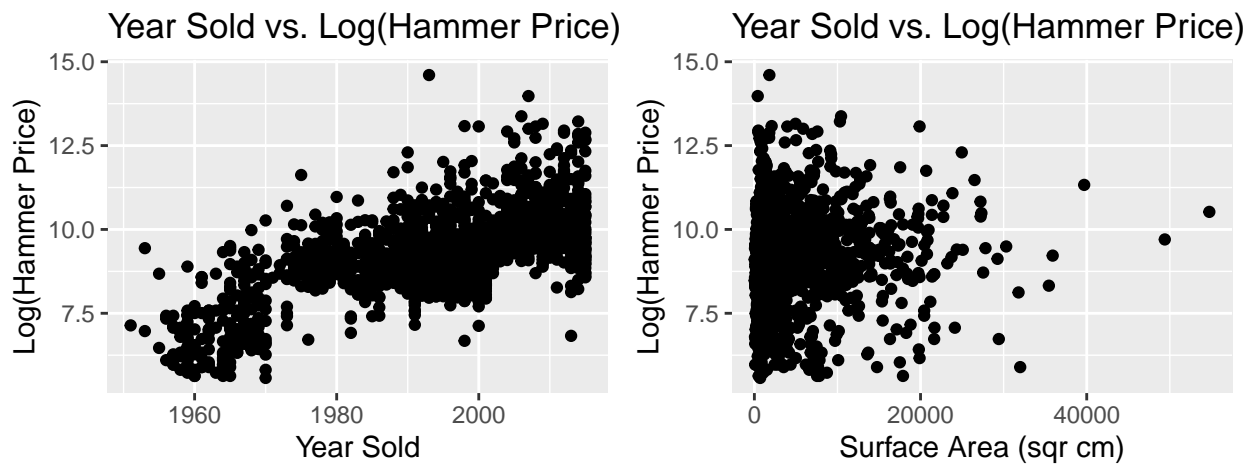
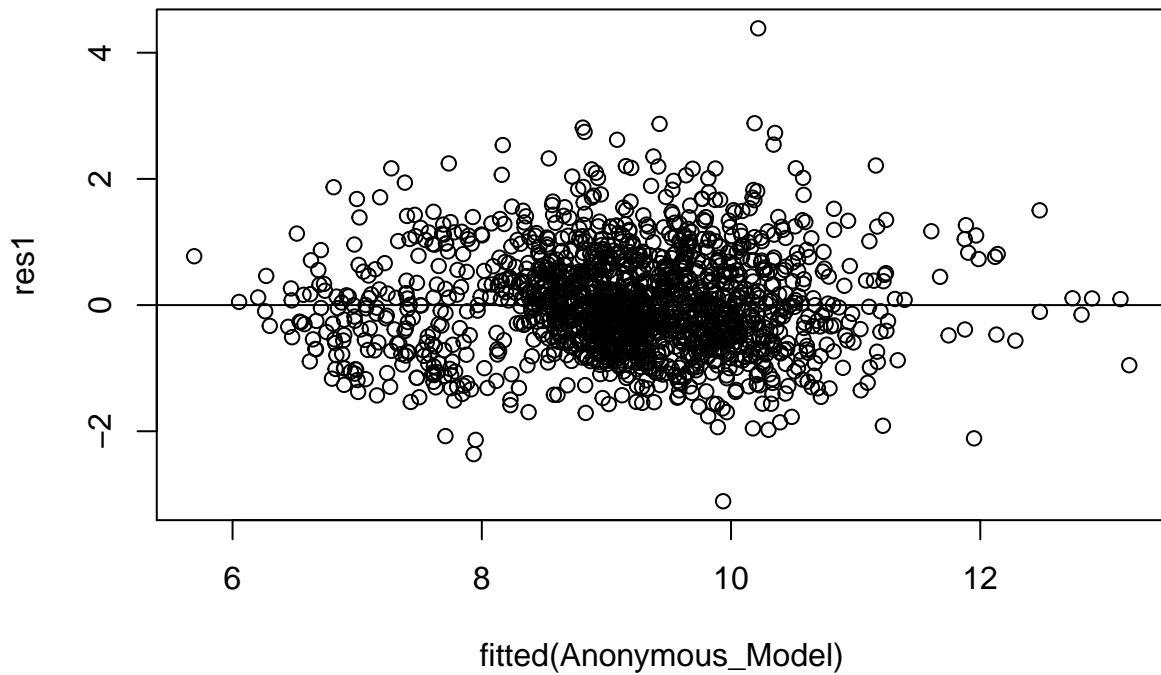
## Model Assumptions

The assumptions for this model are as follows: 1. Linearity: There is a linear relationship between the response ( $\log(\text{Hammer Price})$ ) and each predictor variable. 2. Constant Variance: The variability of the errors is equal for all values of the predictor variable. 3. Normality: The errors follow a normal distribution. 4. Independence: The errors are independent from each other.

### Anonymous Model

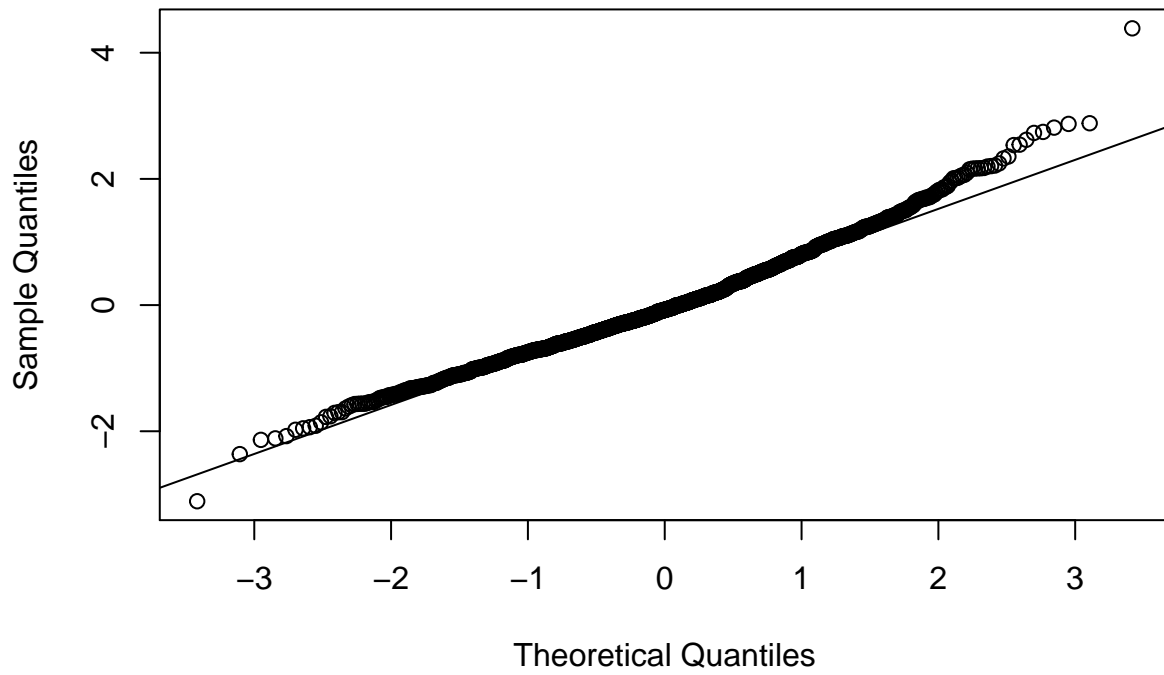
1. For linearity: The plot of standardized residuals vs. predicted shows no distinguishable pattern and for continuous, numeric variables there appears to be a linear relation with year sold and the response. For the painting's surface area, there does not appear to be a linear relationship, however because there is no distinguishable pattern, we will continue with this model although the linearity assumption is not entirely met.

### standardized residuals vs. predicted



2. For Constant Variance, the vertical spread of the residuals is relatively constant across the plot. The constant variance condition is satisfied
3. For normality, the errors follow approximately a normal distribution. Although not perfectly satisfied, we will continue with the model.

### Normal Q-Q Plot

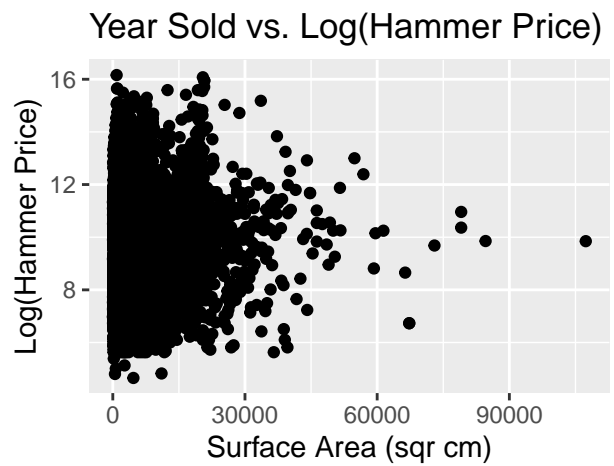
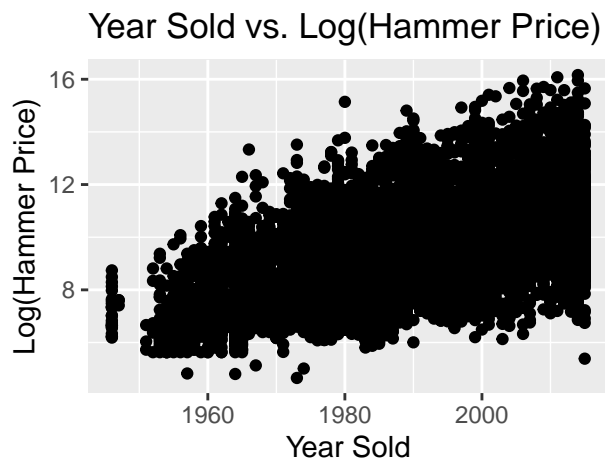
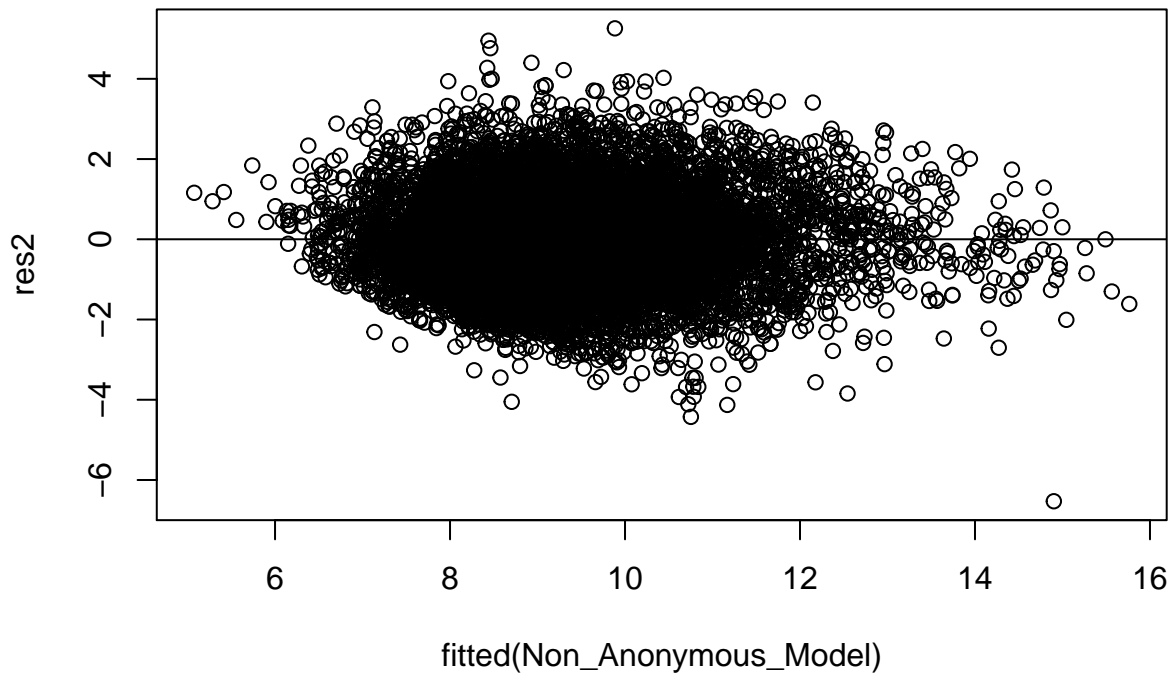


4. Independence: We have no reason to believe that each observation is not independent of one another. Each observation is a unique sale not related to other pieces of art.

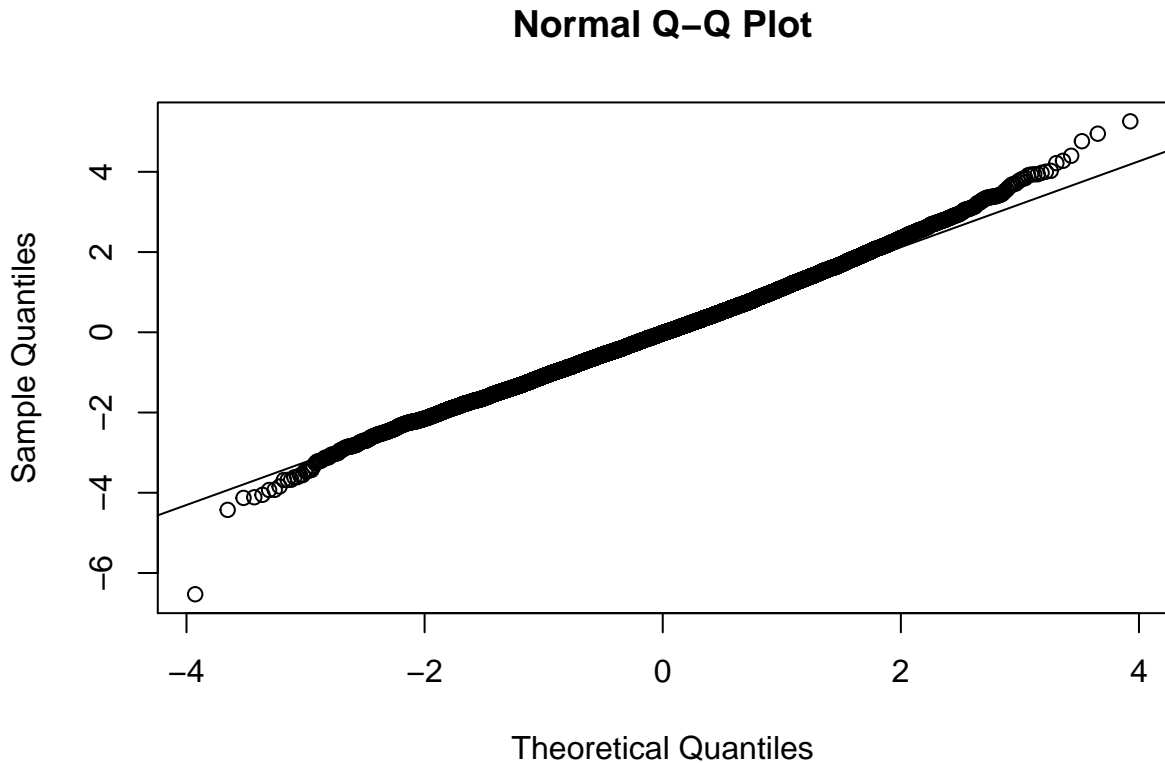
### Non-Anonymous Model

1. Same interpretaion as (1) anonymous model.

### standardized residuals vs. predicted



2. Same interpretaion as (2) anonymous model.
3. Same interpretaion as (3) anonymous model.



4. Same interpretation as (4) anonymous model.

### Releveling/Baseline Assumptions For Both Models

- Antwerpen School was made the baseline for the school of a painting because it has the most observations.
- 15th century was made the baseline for painting's century because it is the earliest time period and it is most logical to interpret coefficients in chronological order.
- The Religious Topic was made the baseline for painting subject matter because it has the most observations.
- Great Britain's Pound was made the baseline for currency used in sale because has the most observations.
- The un-mentioned collection type was made the baseline for collection type as it has the most observations and it has the lowest distribution in price.
- The most top tier auction house was made the baseline for auction house prestige as it has the most observations and is logical to compare from top organizations
- Oil medium was made the baseline for painting medium as it has the most observations and, during the 15-17th century, most work was done in oil.
- Panel was made the baseline for material as it has the most observations.

### Baseline for Anonymous Model

An evening sale; The dataset's average painting surface area of 4,956 sqm; 15th century painting; Antwerpen School; Paid for with British Pounds; Collection type not specified; Sold in the first year recorded in the dataset (1946); Of a religious topic; Sold at a top tier auction house; No publications on the painting; No date on the painting; Painted in oil on a panel; CPI at time of sale was 8.3 (recorded for 1946); Was not Reattributed; And has not been in an exhibit.

## Baseline for Non-Anonymous Model

An evening sale; The dataset's average painting surface area of 4956 sqr cm; 15th century painting; Antwerpen School; Paid for with British Pounds; Collection type not specified; Sold in the first year recorded in the dataset (1946); Of a religious topic; Sold at a top tier auction house; No publications on the painting; No date on the painting; Painted in oil on a panel; CPI at time of sale was 8.3 (recorded for 1946); Was not Reattributed; Has not been in an exhibit; Was not examined via scientific technology; Was not signed.

## Full Anonymous Model

Table 4: Non-Anonymous Model

Variable	Coeff	95% CI Lower Bound	95% CI Upper Bound	P-Value	Sig
Intercept	38114.960	29982.759	48452.851	<0.001	Yes
Evening Sale (vs. Day)	5.534	3.533	8.666	<0.001	Yes
Surface Area of Painting	1.014	1.010	1.019	<0.001	Yes
15th-16th century (vs. 15th)	0.734	0.508	1.060	0.100	No
16th century (vs. 15th)	0.563	0.471	0.673	<0.001	Yes
16th-17th century (vs. 15th)	0.477	0.363	0.627	<0.001	Yes
Century not mentioned (vs. 15th)	0.473	0.188	1.191	0.112	No
Bruges School (vs. Antwerpen)	1.217	1.002	1.477	0.047	Yes
Bruxelles School (vs. Antwerpen)	1.074	0.705	1.637	0.739	No
Flemish School (vs. Antwerpen)	1.491	1.008	2.206	0.046	Yes
Ghent School (vs. Antwerpen)	1.738	0.676	4.468	0.252	No
Leuven School (vs. Antwerpen)	4.542	0.891	23.146	0.069	No
Netherlandish School (vs. Antwerpen)	0.797	0.714	0.890	<0.001	Yes
Portrait (vs. Religious topic)	1.019	0.877	1.186	0.803	No
Landscape (vs. Religious topic)	0.993	0.812	1.215	0.947	No
Allegory (vs. Religious topic)	0.941	0.722	1.227	0.655	No
Genre Scene (vs. Religious topic)	1.081	0.847	1.380	0.533	No
Mythology (vs. Religious topic)	1.032	0.790	1.347	0.818	No
Still Life (vs. Religious topic)	1.020	0.728	1.429	0.908	No
Other Topic (vs. Religious topic)	1.084	0.820	1.432	0.571	No
Purchased in Schillings (vs. British Pound)	0.827	0.559	1.223	0.341	No
Purchased in Austrian Dollars (vs. British Pound)	0.889	0.385	2.054	0.783	No
Purchased in Belgian Francs (vs. British Pound)	0.799	0.619	1.030	0.084	No
Purchased in Canadian Dollars (vs. British Pound)	5.250	1.001	27.538	0.050	No
Purchased in Swiss Francs (vs. British Pound)	1.168	0.908	1.502	0.226	No
Purchased in Deutsche Mark (vs. British Pound)	1.223	0.966	1.548	0.094	No
Purchased in Danish Krone (vs. British Pound)	0.915	0.521	1.607	0.757	No
Purchased in Spanish Peseta (vs. British Pound)	1.052	0.761	1.456	0.758	No
Purchased in Euro (vs. British Pound)	1.030	0.860	1.234	0.748	No



Variable	Coeff	95% CI Lower Bound	95% CI Upper Bound	P-Value	Sig
Purchased in French Franc (vs. British Pound)	1.134	0.937	1.372	0.198	No
Purchased in Irish Pound (vs. British Pound)	0.339	0.066	1.735	0.194	No
Purchased in Italian Lira (vs. British Pound)	0.879	0.695	1.112	0.283	No
Purchased in Dutch Guilder (vs. British Pound)	1.263	1.002	1.592	0.048	Yes
Purchased in Polish Zloty (vs. British Pound)	0.622	0.114	3.402	0.584	No
Purchased in Swedish Krona (vs. British Pound)	1.306	0.938	1.820	0.114	No
Purchased in US Dollars (vs. British Pound)	1.259	1.076	1.474	0.004	Yes
Belgium Collection (vs. Not Specified)	2.828	0.866	9.237	0.085	No
Museum Collection (vs. Not Specified)	1.311	0.619	2.779	0.480	No
Noble Collection (vs. Not Specified)	2.603	1.590	4.259	<0.001	Yes
Private Collection (vs. Not Specified)	1.485	1.261	1.749	<0.001	Yes
Royal Collection (vs. Not Specified)	0.931	0.263	3.290	0.912	No
Year Sold	2.450	2.306	2.604	<0.001	Yes
2nd Tier Auction House (vs. Top Tier)	0.864	0.742	1.006	0.060	No
3rd Tier Auction House (vs. Top Tier)	0.666	0.548	0.810	<0.001	Yes
4th Tier Auction House (vs. Top Tier)	0.694	0.568	0.848	<0.001	Yes
Publications on Painting (vs. None)	1.690	1.305	2.189	<0.001	Yes
Date on Painting (vs. None)	1.124	0.894	1.414	0.316	No
Other Medium (vs. Oil)	1.060	0.589	1.910	0.845	No
Tempera Medium (vs. Oil)	1.153	0.701	1.896	0.576	No
Canvas Material (vs. Panel)	0.742	0.639	0.862	<0.001	Yes
Copper Material (vs. Panel)	0.904	0.719	1.136	0.387	No
Other Material (vs. Panel)	0.868	0.570	1.322	0.510	No
CPI at time of sale	1.144	1.121	1.167	<0.001	Yes
Had Previous Attribute (vs. Didn't)	0.996	0.748	1.327	0.979	No
Been in an Exhibit (vs. not)	0.982	0.727	1.327	0.907	No
Portrait * Year sold	1.062	0.919	1.227	0.418	No
Landscape * Year sold	1.173	0.968	1.421	0.103	No
Allegory * Year sold	1.138	0.858	1.510	0.368	No
Genre Scene * Year sold	1.138	0.897	1.443	0.288	No
Mythology * Year sold	1.278	1.020	1.601	0.033	Yes
Still Life * Year sold	0.797	0.490	1.298	0.362	No
Other Topic * Year sold	1.109	0.829	1.482	0.487	No

## Full Non-Anonymous Model

Table 5: Non-Anonymous Model

Variable	Coeff.	95% CI Lower Bound	P-Value	95% CI Upper Bound	Sig
Intercept	991.128	849.502	<0.001	1156.364	Yes
Evening Sale (vs. Day)	3.223	2.638	<0.001	3.937	Yes

Variable	Coeff.	95% CI Lower		P-Value	95% CI Upper	
		Bound			Bound	Sig
Surface Area of Painting	1.014	1.012	<0.001		1.016	Yes
15th-16th century (vs. 15th)	0.993	0.879	0.911		1.122	No
16th century (vs. 15th)	0.610	0.542	<0.001		0.686	Yes
16th-17th century (vs. 15th)	1.063	0.935	0.353		1.208	No
Beglian School (vs. Antwerpen)	1.699	0.563	0.347		5.131	No
Bruges School (vs. Antwerpen)	1.538	1.414	<0.001		1.673	Yes
Bruxelles School (vs. Antwerpen)	1.615	1.511	<0.001		1.726	Yes
Flemish School (vs. Antwerpen)	1.454	1.122	0.005		1.885	Yes
Ghent School (vs. Antwerpen)	1.210	0.896	0.214		1.635	No
Leuven School (vs. Antwerpen)	1.327	1.082	0.007		1.627	Yes
Netherlandish School (vs. Antwerpen)	0.991	0.938	0.757		1.048	No
Portrait (vs. Religious topic)	0.680	0.558	<0.001		0.829	Yes
Landscape (vs. Religious topic)	2.180	1.806	<0.001		2.632	Yes
Allegory (vs. Religious topic)	1.346	0.870	0.182		2.084	No
Genre Scene (vs. Religious topic)	1.799	1.463	<0.001		2.211	Yes
Mythology (vs. Religious topic)	1.140	0.786	0.489		1.654	No
Still Life (vs. Religious topic)	5.771	3.101	<0.001		10.738	Yes
Other Topic (vs. Religious topic)	0.723	0.479	0.124		1.093	No
Purchased in Schillings (vs. British Pound)	1.094	0.940	0.247		1.274	No
Purchased in Austrian Dollars (vs. British Pound)	1.043	0.453	0.921		2.401	No
Purchased in Belgian Francs (vs. British Pound)	1.174	0.974	0.093		1.415	No
Purchased in Canadian Dollars (vs. British Pound)	0.975	0.465	0.945		2.041	No
Purchased in Swiss Francs (vs. British Pound)	1.147	1.014	0.030		1.299	Yes
Purchased in Czech Koruna (vs. British Pound)	1.634	0.175	0.666		15.230	No
Purchased in Deutsche Mark (vs. British Pound)	1.147	0.995	0.059		1.322	No
Purchased in Danish Krone (vs. British Pound)	1.318	0.910	0.144		1.909	No
Purchased in Spanish Peseta (vs. British Pound)	1.389	1.008	0.044		1.912	Yes
Purchased in Euro (vs. British Pound)	0.843	0.769	<0.001		0.924	Yes
Purchased in Finnish Markka (vs. British Pound)	0.248	0.052	0.079		1.175	No
Purchased in French Franc (vs. British Pound)	1.547	1.409	<0.001		1.698	Yes
Purchased in Hungarian Forint (vs. British Pound)	1.558	0.172	0.694		14.110	No
Purchased in Irish Pound (vs. British Pound)	1.248	0.570	0.580		2.730	No
Purchased in Italian Lira (vs. British Pound)	1.113	0.921	0.268		1.344	No
Purchased in Dutch Guilder (vs. British Pound)	1.005	0.919	0.913		1.100	No
Purchased in New Zealand Dollar (vs. British Pound)	0.411	0.046	0.428		3.705	No

Variable	Coeff.	95% CI Lower Bound	P-Value	95% CI Upper Bound	Sig
Purchased in Polish Zloty (vs. British Pound)	1.721	0.345	0.508	8.582	No
Purchased in Swedish Krona (vs. British Pound)	1.134	0.947	0.171	1.358	No
Purchased in US Dollars (vs. British Pound)	1.343	1.254	<0.001	1.438	Yes
Purchased in S. African Rand (vs. British Pound)	0.647	0.136	0.584	3.075	No
Belgium Collection (vs. Not Specified)	1.350	0.923	0.122	1.974	No
Museum Collection (vs. Not Specified)	0.913	0.588	0.686	1.418	No
Noble Collection (vs. Not Specified)	1.456	1.158	0.001	1.829	Yes
Private Collection (vs. Not Specified)	1.205	1.111	<0.001	1.306	Yes
Royal Collection (vs. Not Specified)	1.746	0.832	0.140	3.661	No
Year Sold	1.074	1.071	<0.001	1.077	Yes
2nd Tier Auction House (vs. Top Tier)	0.788	0.734	<0.001	0.845	Yes
3rd Tier Auction House (vs. Top Tier)	0.442	0.394	<0.001	0.495	Yes
4th Tier Auction House (vs. Top Tier)	0.593	0.528	<0.001	0.667	Yes
Publications on Painting (vs. None)	1.562	1.404	<0.001	1.738	Yes
Date on Painting (vs. None)	1.234	1.121	<0.001	1.358	Yes
Other Medium (vs. Oil)	0.763	0.556	0.094	1.047	No
Tempera Medium (vs. Oil)	1.352	0.906	0.140	2.016	No
Canvas Material (vs. Panel)	0.744	0.695	<0.001	0.797	Yes
Copper Material (vs. Panel)	1.078	0.987	0.095	1.178	No
Other Material (vs. Panel)	0.754	0.625	0.003	0.910	Yes
CPI at time of sale	1.046	1.037	<0.001	1.055	Yes
After (vs. By)	0.241	0.211	<0.001	0.276	Yes
Attributed to (vs. By)	0.552	0.515	<0.001	0.592	Yes
Circle of (vs. By)	0.441	0.412	<0.001	0.472	Yes
Follower of (vs. By)	0.425	0.357	<0.001	0.505	Yes
Manner of (vs. By)	0.171	0.078	<0.001	0.374	Yes
School of (vs. By)	0.356	0.312	<0.001	0.407	Yes
Studio of (vs. By)	0.648	0.587	<0.001	0.716	Yes
Style of (vs. By)	0.304	0.284	<0.001	0.326	Yes
Examined via Scientific Tech. (vs. Not)	0.749	0.443	0.283	1.269	No
Signed (vs. Not)	1.585	1.449	<0.001	1.734	Yes
Been in an Exhibit (vs. not)	1.742	1.546	<0.001	1.963	Yes
Reattributed to a Less Profitable Attrib. (vs. not)	0.635	0.502	<0.001	0.803	Yes
Reattributed to a More Profitable Attrib. (vs. not)	0.930	0.745	0.524	1.162	No
Portrait * Year sold	1.002	0.998	0.371	1.006	No
Landscape * Year sold	0.988	0.984	<0.001	0.992	Yes
Allegory * Year sold	1.001	0.992	0.828	1.010	No
Genre Scene * Year sold	0.999	0.995	0.614	1.003	No
Mythology * Year sold	0.995	0.987	0.186	1.003	No
Still Life * Year sold	0.979	0.967	0.002	0.993	Yes
Other Topic * Year sold	1.007	0.998	0.140	1.016	No