

Predicting Prices in the Art Auction Market

IDS 702 Final Project Presentation

Xinyi Pan

What an auction is like?



Key Research Questions

- Which features drive up the prices of paintings at auction?
- Does the auction house where the painting is auctioned affect the sale price?
- Are the presence of the artist's signature on the painting and the medium used important factors affecting the price of the painting?



What an auction lot is like?



How auction lots are recorded?



Art images copyright© of artist or assignee

LADY BESIDE A TREE

Yi Ren

Sales Price: \$14,345 (HKD 112,500) **(Fees to buyer included)
Estimate: \$10,200 - \$15,300 (HKD 80,000 - HKD 120,000)
Price/Square Inch: \$16.95

Size: 55.12" x 15.35"
Year Created: 1884
Signature: Signed and Dated
Medium/Ground: Watercolor and ink / Paper

Auction Date(s): [10/06/2019 \(all lots\)](#)
Auction: Fine Classical Chinese Paintings 10/6/2019
By: [Sotheby's Hong Kong](#)
Lot: 2506

Add to Report ☐

[Copy Lot](#)

[Lot URL](#)

[More details and large image](#)

Scope of Study

- Genre: Contemporary and modern art
- Period: 2016-2020
- Auction House: Christie's vs Sotheby's
- Auction Location: Hong Kong

Data

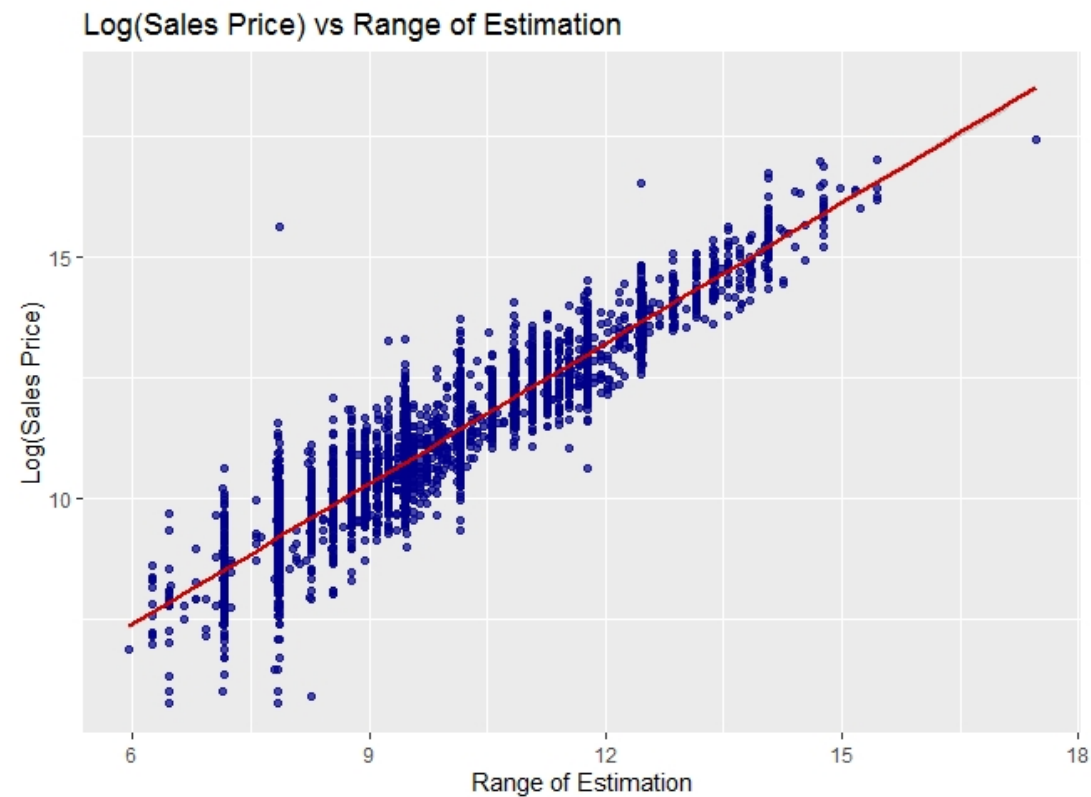
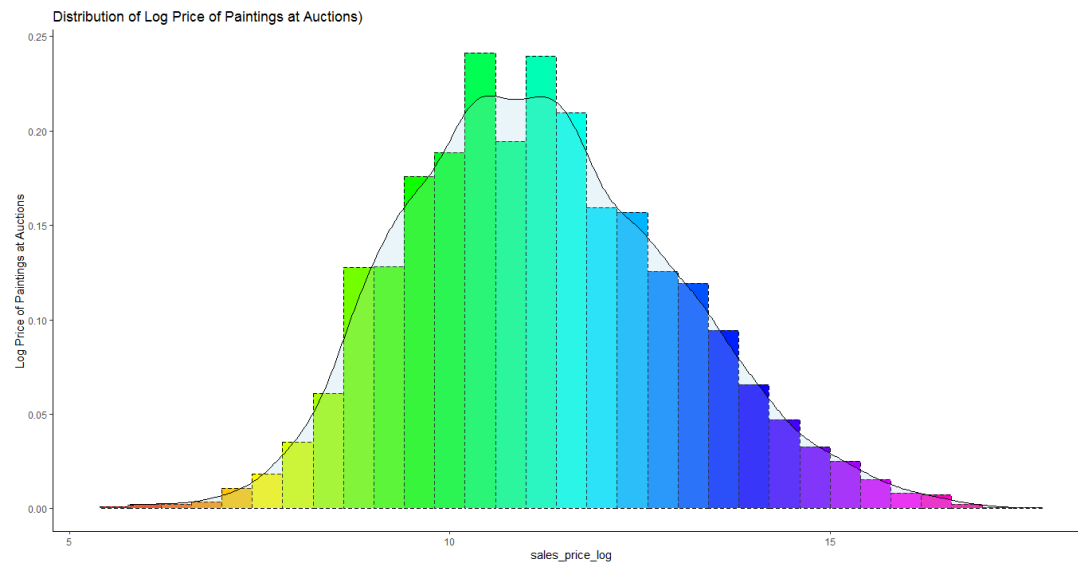
- Numeric variables

1. Sales price (outcome)
2. Range of estimation
3. Surface (height * width)

- Categorical variables

1. Auction House (Sotheby's vs Christie's)
2. Year of Auction
3. Month of Auction
4. Weekday of Auction
5. Dummy variables related to signature & medium

EDA

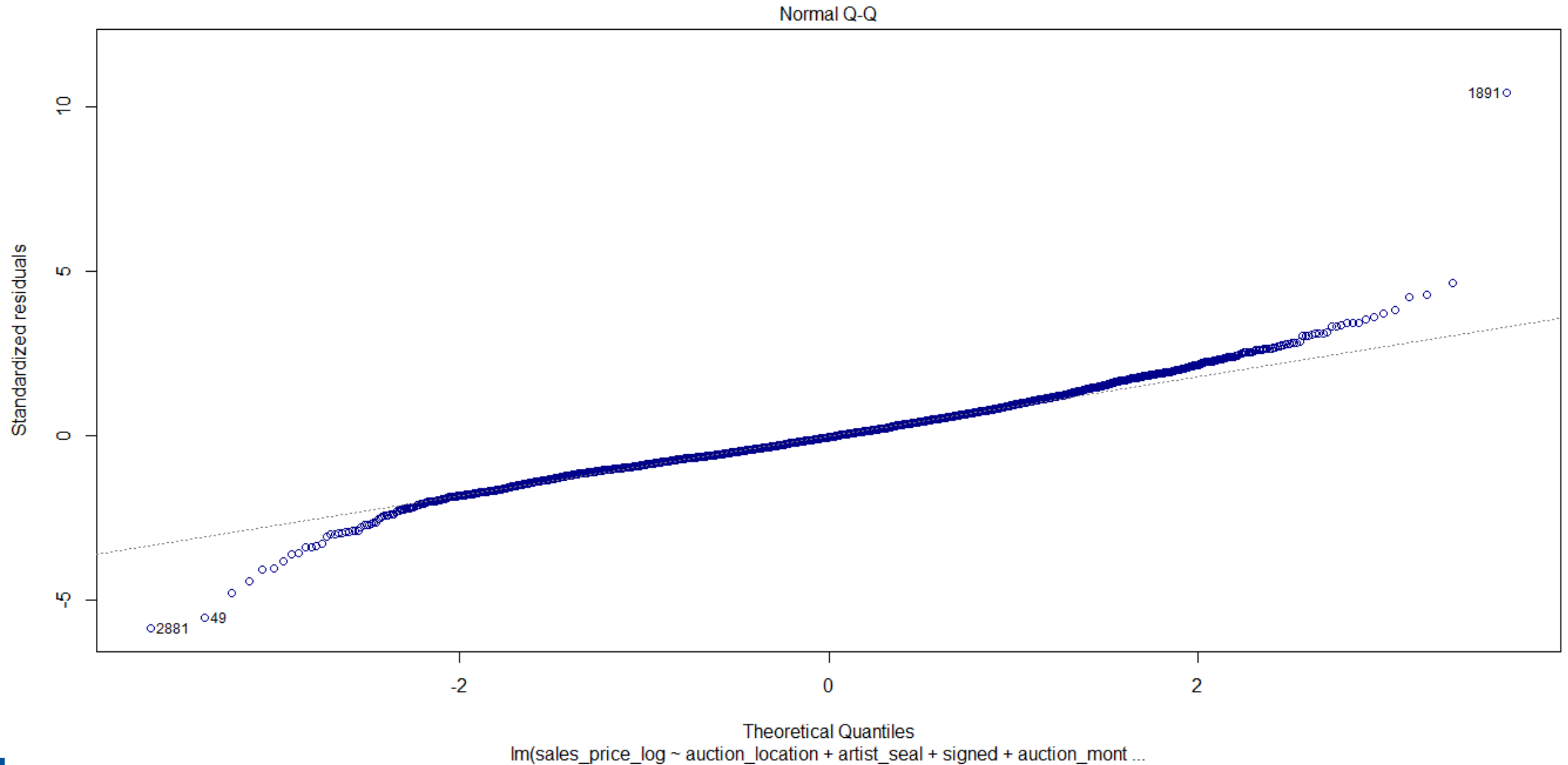


Model

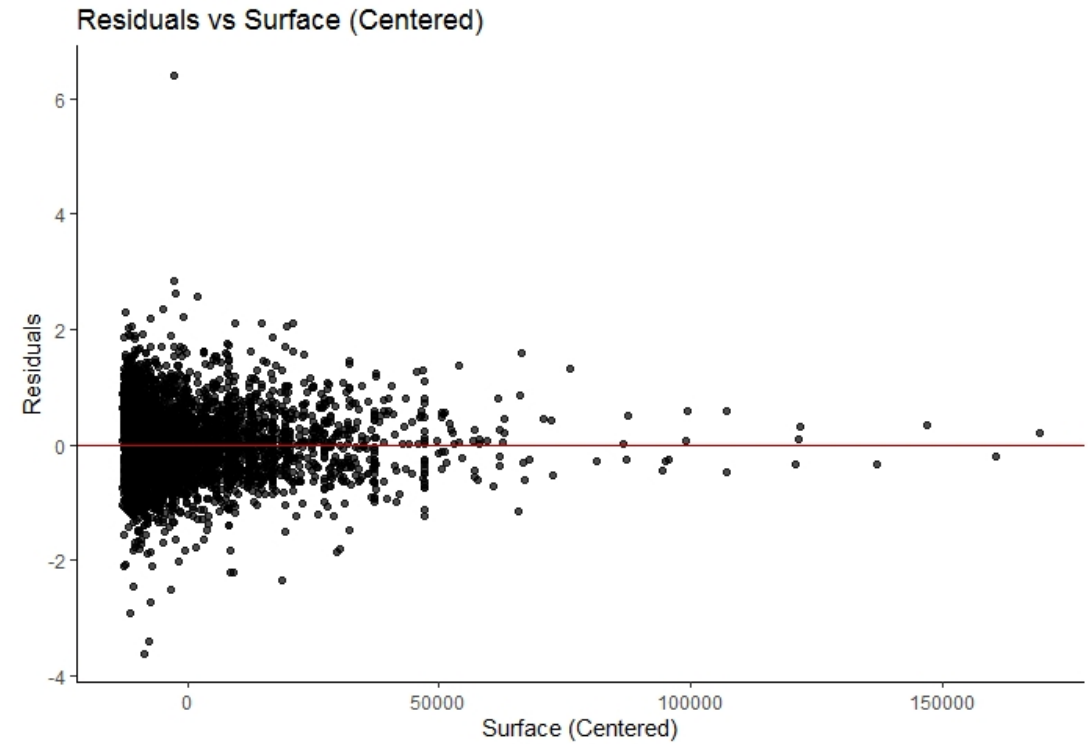
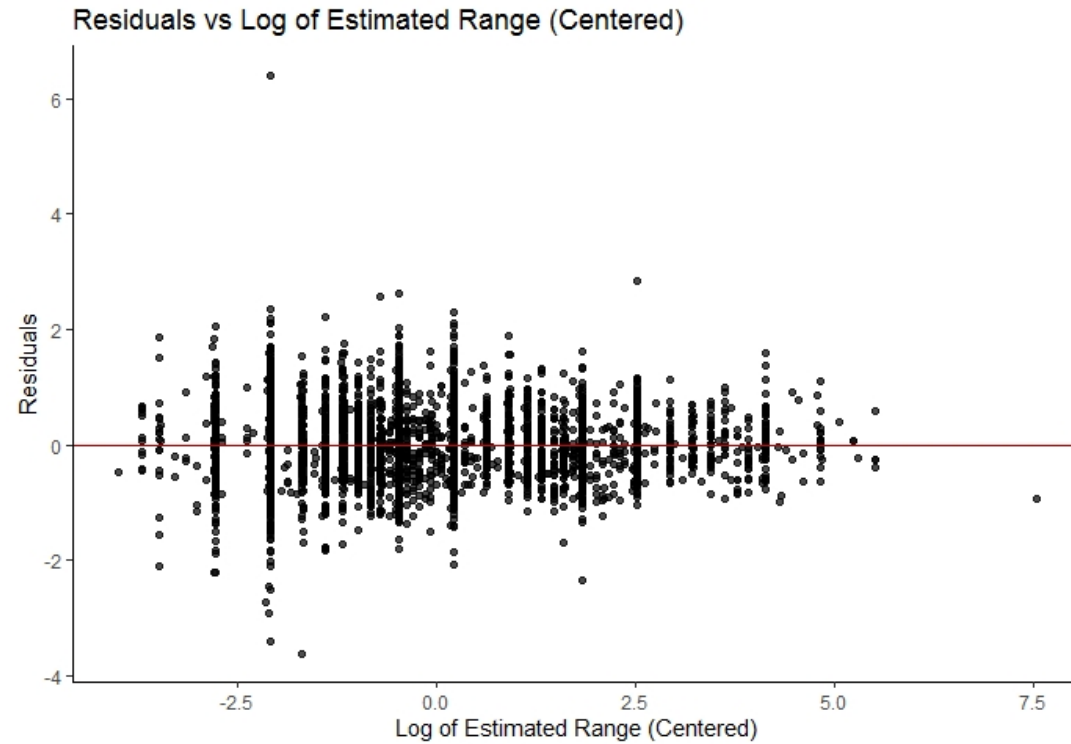
```
lm(formula = sales_price_log ~ artist_seal + signed + auction_month +
    auction_year + estimate_range_log_c + canvas * (oil + acrylic),
    data = mydata)
```

sales_price_log							
Predictors	Estimates	CI	p				
(Intercept)	11.36	11.15 – 11.56	<0.001	auction_month [11]	-0.06	-0.28 – 0.15	0.558
auction_location [Christie's]	-0.03	-0.16 – 0.10	0.644	auction_month [12]	-0.30	-0.61 – 0.01	0.058
artist_seal [Artist's Seal]	-0.19	-0.39 – 0.01	0.068	auction_year [2017]	-0.02	-0.08 – 0.03	0.387
signed [Signed]	0.07	0.00 – 0.14	0.047	auction_year [2018]	-0.04	-0.10 – 0.01	0.127
auction_month [3]	0.06	-0.14 – 0.27	0.539	auction_year [2019]	-0.04	-0.10 – 0.01	0.149
auction_month [4]	-0.03	-0.20 – 0.14	0.725	auction_year [2020]	-0.36	-0.55 – -0.17	<0.001
auction_month [5]	-0.03	-0.25 – 0.19	0.771	estimate_range_log_c	0.96	0.95 – 0.97	<0.001
auction_month [6]	0.23	-0.03 – 0.49	0.090	canvas [Canvas]	-0.16	-0.31 – -0.01	0.036
auction_month [7]	0.62	0.27 – 0.98	0.001	oil [Oil]	-0.16	-0.26 – -0.05	0.005
auction_month [9]	0.14	-0.07 – 0.35	0.187	acrylic [Acrylic]	-0.13	-0.25 – -0.00	0.041
auction_month [10]	-0.07	-0.25 – 0.10	0.412	canvas [Canvas] * oil [Oil]	0.16	-0.00 – 0.32	0.052
				canvas [Canvas] * acrylic [Acrylic]	0.17	0.01 – 0.34	0.043
Observations				4194			
R ² / R ² adjusted				0.881 / 0.880			

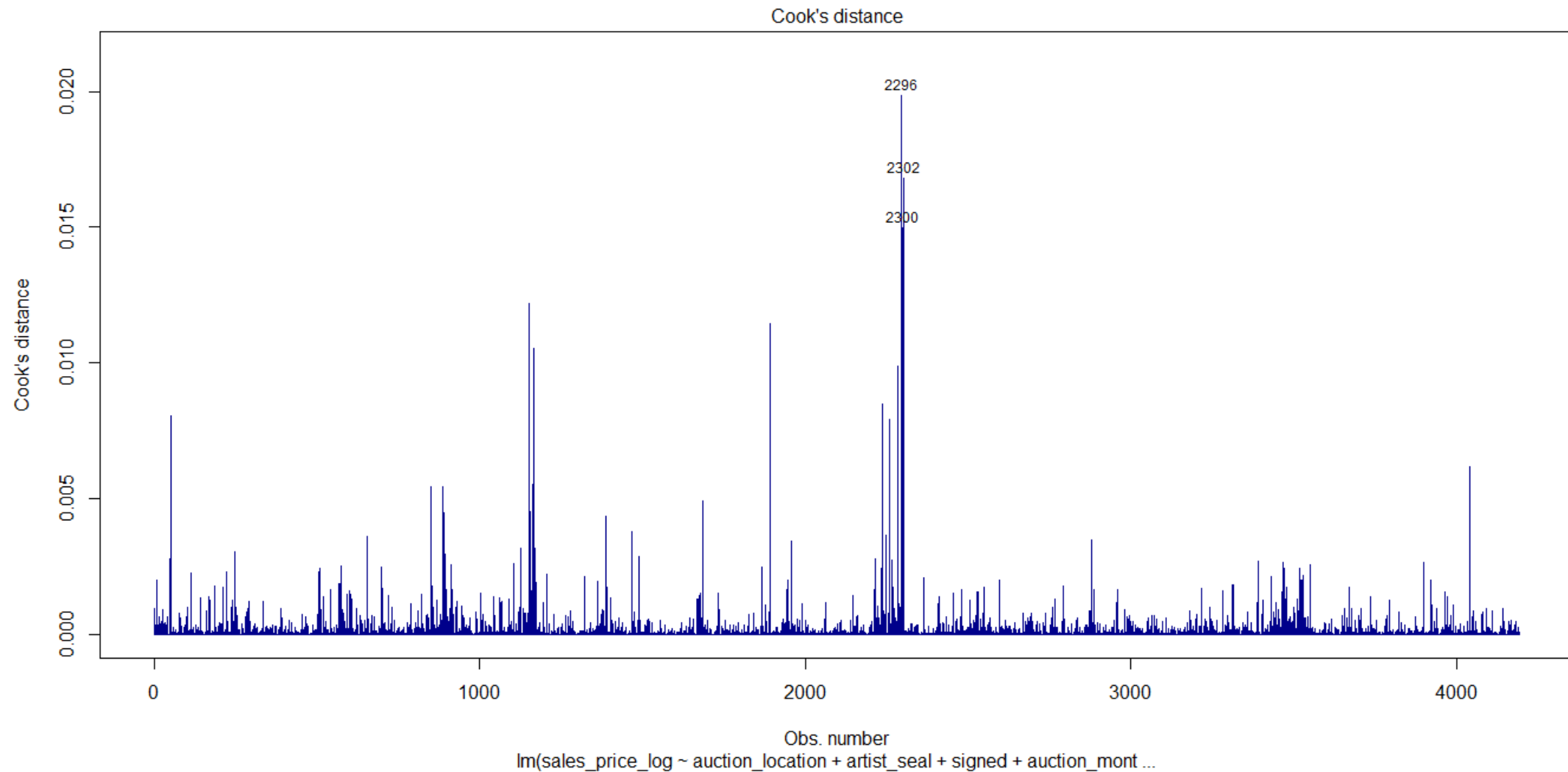
Model Assessments



Model Assessments



Model Assessments



Limitations & Future Work

- Residuals with high leverage;
Outliers exist
- Multicollinearity
- Missing Data Imputation
- Prediction yet to finish
- Non-parametric Models
- Unbalanced observations
→ weighted effect coding

Thanks for watching!