Sentiment in Online Car Auctions

Francis Peng

March 31th, 2023

Motivation

- E-commerce: Ebay, Amazon, Facebook Marketplace, and others
    - Data accessibility
    - How does the online nature of the market affect how it operates?
    - Adverse selection

Motivation

- Lewis (2011) – "Asymmetric Information, Adverse Selection and Online Disclosure: The Case of eBay Motors"
  - "Institutional features" – photos, text, etc. allows sellers to reveal private information
  - Adverse selection effects disappear
- Comments, product reviews are also similar institutional features
  - Sentiment

Motivation

Sentiment in economics:

- Angelitos and La'O (2013)
- Baker and Wurgler (2007)
- Yen et al. (2021) and Xu and Hsu (2022)

Motivation

In this paper...

- Auction rather than asset market
- Sentiments are expressed directly on the marketplace
- Sentiments are accessible to all agents of the market

Motivation

Question: Does sentiment in the context of an auction affect its final price?

Data

- carsandbids.com
  - Online car auction website
  - English auctions
  - Model years 1980 and up
  - Over 10,000 cars sold

Figure 1: carsandbids.com homepage.

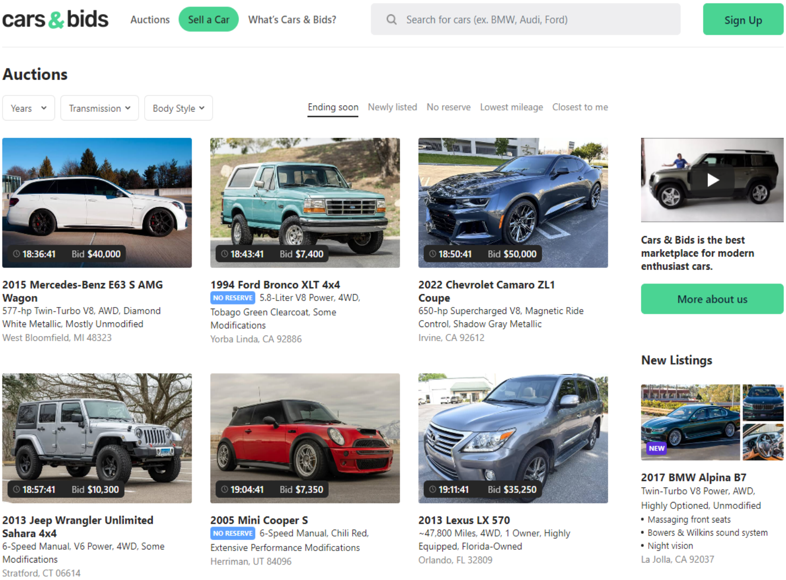


Figure 2: Basic vehicle information.

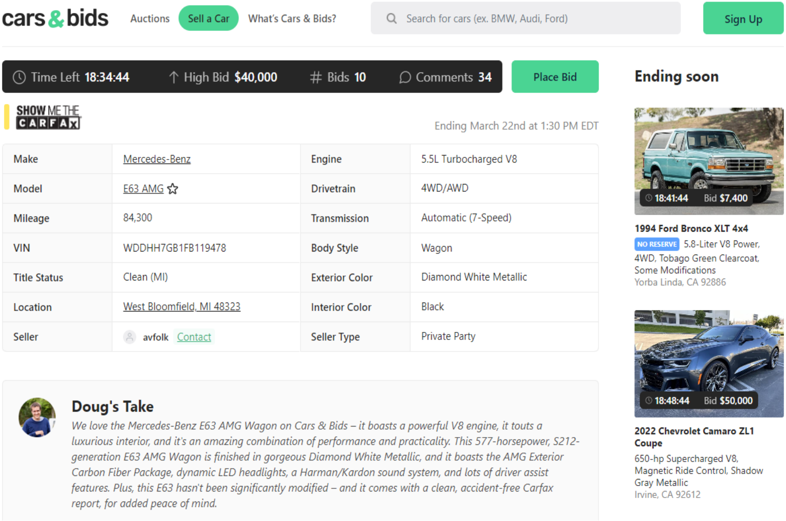
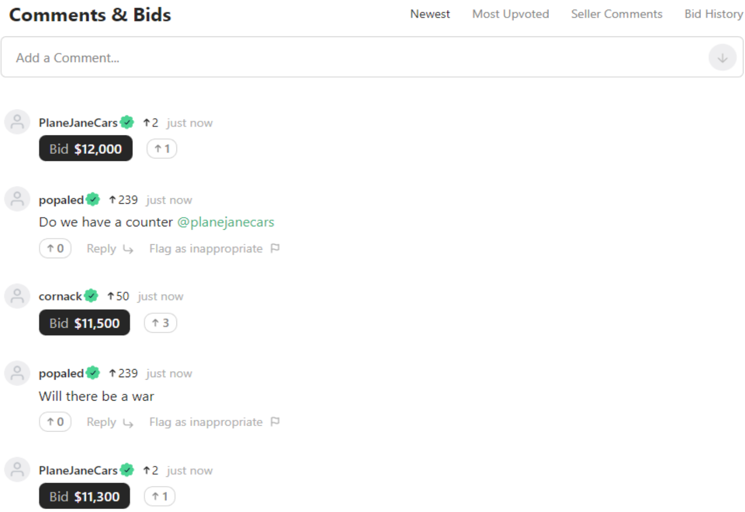


Figure 3: Comments & Bids section.



Data

- Total data scraped:
  - 8,405 auctions
    - Reduced to 6,942 after removing make-models with few observations
  - Auctions ending April 26, 2021 to January 6, 2023

Table 1: Summary Statistics

Variables	Mean	SD	Min	Max	N
Price	30,815.89	30,121.88	2,950	405,911	6,942
Sentiment score	0.62	0.11	0	1	6,942
Number bids	28.32	13.50	1	139	6,942
Number bidders	12.41	4.57	1	33	6,942
Number comments	41.29	26.25	1	381	6,942
Mileage	76,421.89	56,092.31	5	336,400	6,942
Number highlights	5.21	0.86	3	8	6,942
Number equipment	11.96	4.21	1	35	6,942
Number modifications	5.77	8.62	0	118	6,942
Number known flaws	6.45	3.84	0	28	6,942
Number service history	3.68	3.65	0	40	6,942
Number other items	4.23	2.18	0	24	6,942
Number owner history	1.00	0.07	0	1	6,942
Number videos	2.92	2.13	0	22	6,942
Number views	10,590.55	5,792.04	2,912	96,557	6,942
Number photos	117.98	47.86	37	443	6,942

Methodology: Overview

1. Define and create a sentiment score for each auction
2. Define a linear model to explain price as a function of the other variables
    - Regress ln(Price) on ln(Sentiment score), other covariates, and controls

$$\ln y_t = \beta \times \ln(\text{Sentiment Score}_t) + \lambda x_t + \alpha \gamma_t + \varepsilon_t$$

Methodology: Sentiment Analyzer

- Optimally:
  - Random subset of comments would be used to create training dataset
  - Sentiment analyzer would be trained using this dataset
  - Benefit: best performance
  - Downside: long process, time and resource constraints
- Alternative:
  - Use preexisting, “off-the-shelf” sentiment analysis model
  - “Twitter-roBERTa”
    - Trained on 124M tweets from January 2018 to December 2021

Methodology: Sentiment Score

- Mean:

$$sentimentscore_n = \frac{\sum_i sentimentscore_{n,i}}{I_n}$$

Where

- $sentimentscore_n$ is the overall sentiment score for an auction n
- $sentimentscore_{n,i}$ is the sentiment score for a message i in an auction n
- I is the total number of messages in an auction n
- All sentiment scores min-max scaled to [0, 1]

Methodology: Regression Analysis

$$\ln y_t = \beta \times \ln(\text{Sentiment Score}_t) + \lambda x_t + \alpha\gamma_t + \varepsilon_t$$

Where

- y_t is the price or ending bid of an auction t
- Sentiment score$_t$ is the sentiment score of auction t
- x_t is a vector of other covariates (next slide)
- γ_t is a vector of control variables which contains the year as well as dummy variables for the make, model, and color of vehicle in order to control for the intrinsic value of the vehicle of auction t
- ε_t represents the idiosyncratic preferences of the highest bidder of an auction t which may affect their bid

Table 2: Variables in vector x

Variables	Type	Description
Number bids	discrete	Number of bids of the auction.
Number bidders	discrete	Number of bidders in the auction.
Number comments	discrete	Number of comments in the "Comments & Bids" section.
Mileage	continuous	Milage of vehicle on auction.
Private seller	binary	1 if seller of vehicle is a private seller, i.e. not a dealership. 0 otherwise.
Number highlights	discrete	Number of highlights listed.
Number equipment	discrete	Number of equipment listed.
Number modifications	discrete	Number of modifications listed.
Number known flaws	discrete	Number of known flaws listed.
Number service history	discrete	Number of service history listed.
Number other items	discrete	Number of other items listed.
Number owner history	discrete	Number of owner history listed.
Number videos	discrete	Number of videos on auction page.
Title status	binary	1 if vehicle has a clean title. 0 otherwise.
Number views	discrete	Number of times auction was viewed.
Number photos	discrete	Number of photos of vehicle.

Methodology: Specification

$$\ln y_t = \beta \times \ln(\text{Sentiment Score}_t) + \lambda x_t + \alpha\gamma_t + \varepsilon_t \qquad (1)$$

versus

$$y_t = \beta \times (\text{Sentiment Score}_t) + \lambda x_t + \alpha\gamma_t + \varepsilon_t \qquad (2)$$

- (2) presumes that covariates have a dollar effect on price
    - Does not change depending on the price of vehicle

however...

- A dollar effect is higher for a lower priced vehicle than higher priced vehicle
- Intuitively, this should not necessarily be the case
- (1) with log allows for *percentage* effect on price per 1% change in β

Results

Table 3: OLS Estimation Results

	(1)
	ln(Price)
ln(Sentiment score)	0.199***
	(0.019)
Reserve	0.170***
	(0.010)
Number bids	0.007***
	(0.000)
Number bidders	-0.013***
	(0.001)
Number comments	0.000
	(0.000)
Milage	-0.000***
	(0.000)
Number modifications	0.004***
	(0.001)
Number known flaws	-0.026***
	(0.001)
Number views	0.000***
	(0.000)
Number photos	0.000***
	(0.000)
R-squared	0.874
Observations	6941

Robust standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Results

- R-squared: 87% of variation in ln(price) explained by model
  - Remaining unexplained 13% variation due to idiosyncratic tastes of winning bidder
- Sentiment score: 1% increase in sentiment score causes 0.199% increase in price
- Number of bids and bidders: interpretable as bids per bidder
- Reserve: higher price for an auction with reserve. Why?

Results: Sentiment Score

- Average vehicle price = \$31,000
- Suppose we have one such vehicle with sentiment score of 0.5. All else constant, an increase of sentiment score from 0.5 to 1, i.e. a 100% increase is given by:

$$(2^{\beta} - 1) \times 100 = (2^{0.199} - 1) \times 100 = 14.8\%.$$

Why? See next slide...

Remember:

$$\ln y = \beta \times \ln(\text{Sentiment score}) + \lambda x + \alpha\gamma$$

Suppose we increase Sentiment score by 100%, then

$$\text{Sentiment score}_{new} = 2 \times \text{Sentiment score}.$$

Then

$$\begin{aligned}
\ln(y_{new}) &= \beta \ln(2\text{Sentiment score}) + \lambda x + \alpha\gamma \\
&= \beta \ln(\text{Sentiment score}) + \lambda x + \alpha\gamma + \beta \ln(2) \\
&= \ln(y) + \beta \ln(2) \\
\ln(y_{new}) - \ln(y) &= \beta \ln(2) \\
e^{\ln(y_{new}) - \ln(y)} &= e^{\beta \ln(2)} \\
\frac{y_{new}}{y} &= 2^{\beta} \\
100 \times \left(\frac{y_{new} - y}{y}\right) &= (2^{\beta} - 1) \times 100
\end{aligned}$$

Results: Robustness and Sparsity

Potential problem: Loss of estimator efficiency due to high dimensionality

Table 4: Variables in vector γ

Variables	Type	Description
Make-model	nominal	Unique make and model of vehicle (encoded as 264 dummy variables).
Year	discrete	Model year of vehicle.
Interior color	nominal	Interior color of vehicle (encoded as 13 dummy variables).
Exterior color	nominal	Exterior color of vehicle (encoded as 13 dummy variables).

291 control variables, 309 total covariates against $\sim$7,000 observations

Robustness and Sparsity: Solution(s)

- Use domain knowledge
  - Relies on accuracy and correctness of researcher
- Use lasso-inferential regression method
  - robustly selects controls, uses the data to decide

1. Double-selection — Belloni, Chernozhukov, and Hansen (2014)
2. Partialling-out — Belloni et al. (2012)
3. Cross-fit partialling-out — Chernozhukov et al. (2018)

- All 3 methods will estimate β and λ but will use lassos to select a subset of γ. α is not estimated.

$$\ln y_t = \beta \times \ln(\text{Sentiment score}_t) + \lambda x_t + \alpha\gamma_t + \varepsilon_t$$

Table 5: Lasso Linear Estimation Results: ln(Price)

	(1) Double-Selection	(2) Partialing-Out	(3) Cross-Fit Partialing-Out
ln(Sentiment score)	0.161***	0.137***	0.127***
	(0.022)	(0.023)	(0.024)
Reserve	0.250***	0.275***	0.288***
	(0.011)	(0.011)	(0.012)
Number bids	0.007***	0.007***	0.007***
	(0.001)	(0.001)	(0.001)
Number bidders	-0.013***	-0.013***	-0.014***
	(0.001)	(0.002)	(0.002)
Number comments	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Milage	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
Number modifications	0.003***	0.002***	0.003***
	(0.001)	(0.001)	(0.001)
Number known flaws	-0.030***	-0.031***	-0.033***
	(0.002)	(0.002)	(0.002)

Table 6: Lasso Linear Estimation Results: Continued

	(1) Double-Selection	(2) Partialing-Out	(3) Cross-Fit Partialing-Out
Number views	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)
Number photos	0.000***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)
Observations	6941	6941	6941
Number potential controls	289	289	289
Number controls selected	162	162	171

Note: Robust standard errors in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Conclusion

- What does sentiment score represent?
  - Implications for causality
- Does positive sentiment cause people to bid more?
  - corr(Sentiment score, Number bids) = -0.084
  - corr(Sentiment score, Number bidders) = -0.080
- Can we explain sentiment score?
  - R-squared = 0.14

Conclusion: Sentiment vs Useful Information

There is both irrational emotional information, i.e. sentiment, and also useful private information expressed in the comments.

c.f.

> *There are pictures of the car literally underneath a lift so it seems like a very small effort to put the car on the lift to get undercarriage pictures. It is unfathomable to me the seller isn't willing to put in that small of an effort. If the undercarriage is as clean as they claim it is a difference of easily multiple 10's of thousands of dollars would be had. Combine that with the caginess answering questions and erroneously listing it as a V-Spec originally and this deal smells real bad.*

Conclusion: Sentiment vs Useful Information

...versus...

> *Just checked out your instagram. Great Collection you got going on!*

and

> *Makes sense, Thanks for the response brother!*

Conclusion: Extensions

- Sentiment analyzer unable to differentiate between pure sentiment and useful information
- How could one do so?
  - Aspect based sentiment analyzer for this corpus
  - Topic model to categorize comments

Thank You!

References I

Cars & Bids: Auctions of cool modern cars, trucks, and SUVs.

G.-M. Angeletos and J. La'O.
Sentiments.
Econometrica, 81(2):739–779, 2013.
_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA10008.

M. Baker and J. Wurgler.
Investor Sentiment in the Stock Market.
Journal of Economic Perspectives, 21(2):129–151, Apr. 2007.

A. Belloni, D. Chen, V. Chernozhukov, and C. Hansen.
Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain.
Econometrica, 80(6):2369–2429, 2012.
_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA9626.

References II

A. Belloni, V. Chernozhukov, and C. Hansen.
Inference on Treatment Effects after Selection among High-Dimensional Controls†.
The Review of Economic Studies, 81(2):608–650, Apr. 2014.

V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins.
Double/debiased machine learning for treatment and structural parameters.
The Econometrics Journal, 21(1):C1–C68, Feb. 2018.

Jia-Lang Xu and Ying-Lin Hsu.
The Impact of News Sentiment Indicators on Agricultural Product Prices.
Computational Economics, Sept. 2021.
MAG ID: 3196414579.

References III

G. Lewis.
Asymmetric Information, Adverse Selection and Online Disclosure: The Case of eBay Motors.
The American Economic Review, 101(4):1535–1546, 2011.
Publisher: American Economic Association.

M. Yen, Y. Huang, L. Yu, and Y. Chen.
A Two-Dimensional Sentiment Analysis of Online Public Opinion and Future Financial Performance of Publicly Listed Companies.
Computational Economics, 59(4):1677–1698, Apr. 2022.