Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Francis Peng 8 March 2023

Sentiment in Online Car Auctions

By

Francis Peng

Dr. Kyungmin Kim

Advisor

Economics

Dr. Kyungmin Kim

Advisor

Dr. Mike carr

Committee Member

Dr. Jinho Choi

Committee Member

2023

Sentiment in Online Car Auctions

By

Francis Peng

Dr. Kyungmin Kim

Advisor

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Economics

2023

Abstract

[[TITLE]

By [[STUDENT]]

[[ABSTRACT TEXT ABSTRACT TEXT]]

[[TITLE]]

By

[[STUDENT NAME]]

[[PRIMARY ADVISOR]]

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Economics

2023

Acknowledgements

[[ACKNOWLEDGMENTS TEXT ACKNOWLEDGMENTS TEXT ACKNOWLEDGMENTS TEXT ACKNOWLEDGMENTS TEXT ACKNOWLEDGMENTS TEXT ACKNOWLEDGMENTS TEXT ACKNOWLEDGMENTS TEXT ACKNOWLEDGMENTS TEXT ACKNOWLEDGMENTS TEXT ACKNOWLEDGMENTS TEXT]]

Contents

Sentiment in Online Car Auctions

Francis Peng

March 2023

1 Introduction

In the past three decades, online transactions of goods, or e-commerce, has become increasingly prevalent in peoples' lives, and this is evidenced by the prominence of platforms such as Ebay, Amazon, and Facebook Marketplace among others. Not only may an increased understanding of these markets have design or policy implications in e-commerce settings, but by their online nature, the data of such markets is concrete and easily collectible, making them an accessible and robust empirical study case. However, as Lewis (2011) points out, online platforms may have a greater prevalence of adverse selection effects since a buyer usually purchases goods sight unseen. Lewis (2011) argues that certain "institutional features" in Ebay car auctions—such as the ability to upload photos or describe the vehicle with text—allow sellers to reveal their private information, mitigating any impact of information asymmetry on the performance of the market.

Similarly, I seek to estimate the impact of one such institutional feature on a similar car auction website, carsandbids.com. Specifically, each auction provides a place where the seller and bidders can publicly comment, question, or generally converse about the vehicle being auctioned. And intuitively, this is a feature which may also allow agents to reveal private information. Alternatively, it is also a feature which allows the seller and bidder to express their thoughts, opinions, or more generally, their "sentiments."

The idea of sentiment having some bearing on markets is not unique to the present empirical setting where sentiments of buyers and sellers are explicitly stated in the form of an online comment. Rather, it has been theorized in other avenues of economics such as in Angelitos and La'O (2013) where a model is built to explain the co-movement of market expectations and market outcomes as a result of the heterogeneity of agents' market expectations. In this model, agents "rationalize random, and seemingly inexplicable, shifts in the optimism or pessimism that economic agents may hold about one another's choices and thereby about future market conditions."

Empirically, however, discussion of sentiment in economics often falls within the realm of finance or macroeconomic literature where sentiment is used in an asset pricing model. For example, Baker and Wurgler (2007) provide empirical evidence that "investor sentiment" does indeed affect stock prices, but the "sentiment index" they describe is constructed of six, more measurable, proxies such as trading volume and dividend premium, among others. This is an indirect way of measuring sentiment, although the justification for such a method is understandable as sentiment itself is abstract and difficult to measure in most settings.

Some work published since Baker and Wurgler (2007) has attempted to use text data, as I will be doing in this current inquiry, in order to measure sentiment. For example, Yen et al. (2021) shows that sentiment analysis of online news media as well as stock forums could predict future financial performance of companies on the Taiwanese stock market, and similarly, Xu and Hsu (2022) shows that sentiment analysis of news could be used to more accurately predict agricultural product prices.

This current inquiry is a departure from the above in several ways. First, I intend to apply sentiment as a predictor of the price in an auction rather than in an asset market.

Second, the empirical setting studied here is unique in that sentiments are expressed directly on the marketplace. Thus, the sentiment of the market is accessible to all buyers or sellers participating in the market. Furthermore, this particular setting where both the auction and the sentiment of the auction are located on the same platform in the same place provides potential for market design implications.

In order to measure the impact of sentiment on the price of online car auctions, I first calculate a sentiment score for each auction. This sentiment score is defined by the average of the sentiment score of each message as determined by a sentiment analyzer. Then, the sentiment score and other features of the auction web page as well as various controls are used as explanatory variables in a linear regression to explain the price of the auction. Robustness checks with respect to different sentiment analysis algorithms as well as various model specifications and regression methods are conducted. Ultimately, the sentiment score of an auction is statistically significant and of the expected sign in explaining the price of an online car auction.

The remainder of the paper is laid out as follows: Section 2 describes the dataset used as well as the collection of the dataset. Section 3 describes the methodology of the empirical analysis including the sentiment analysis method as well as the regression techniques. Section 4 describes my findings, and finally section 5 discusses the economic implications of my results.

2 Data

carsandbids.com describes itself as an "online auction marketplace to buy and sell modern enthusiast cars..." It facilitates online English auctions of vehicles from model year 1980 to present, and since its launch in June of 2020, it has sold over 10,000 cars (as of March, 2023).

The home page lists currently active auctions in a gallery view in order of time left in the auction with the least time left first (by default). Clicking on any auction brings you to a page specifically for that auction where details of the vehicle and auction can be seen. Basic information such as make, model, milage, VIN number, title status, location, engine type, color, and more is listed in a table below a gallery of images of the vehicle. In addition, several sections of text in bullet points describe the "Highlights," "Equipment," "Modifications," "Known Flaws," "Recent Service History," "Other Items Included in Sale," "Ownership History," and "Seller Notes" of the vehicle. Videos consisting of walking in and around the vehicle, driving, as well as starting the vehicle may also be included on the web page.

Finally, there is a section dedicated to Q&A directed at the seller as well as a "Comments & Bids" section where any registered user on the website can comment, question, or discuss the vehicle and/or auction. These comments are shown as a scrolling chat with the latest messages appearing at the top. Bids are also shown in the same chat log with the username and price shown.

Data of 8,405 auctions which ended from April 26, 2021 to January 6, 2023 was collected from the website. A full list of the variables collected can be found in the appendix along with the Python script through which the data was collected.

Due to the nature of the detailed descriptive information sections as text data, extracting information from this text which may pertain to the value of the vehicle is difficult. Furthermore, whether the information contained in the text of these sections may increase or decrease a bidder's valuation of the vehicle is largely subjective and random to individual bidders. This is especially true for the "Modifications" section as some bidders may see the listed modifications as positive, and others may see them as negative. This also depends on what the modifications are. Other sections such as "Equipment" and "Known Flaws" will always inform either positive or negative aspects of the vehicle. Still, there may be some variability, as bidders may view some flaws as minor, and others as major or deal-breaking flaws.

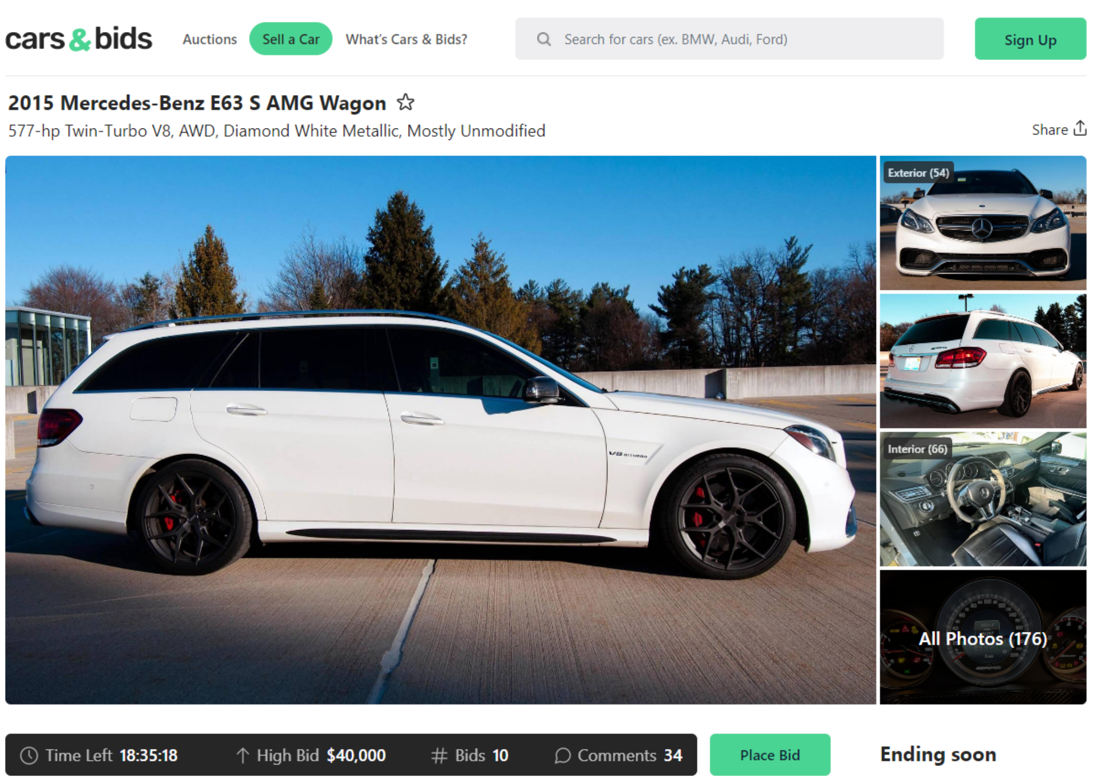

Figure 1: vehicle photo gallery.

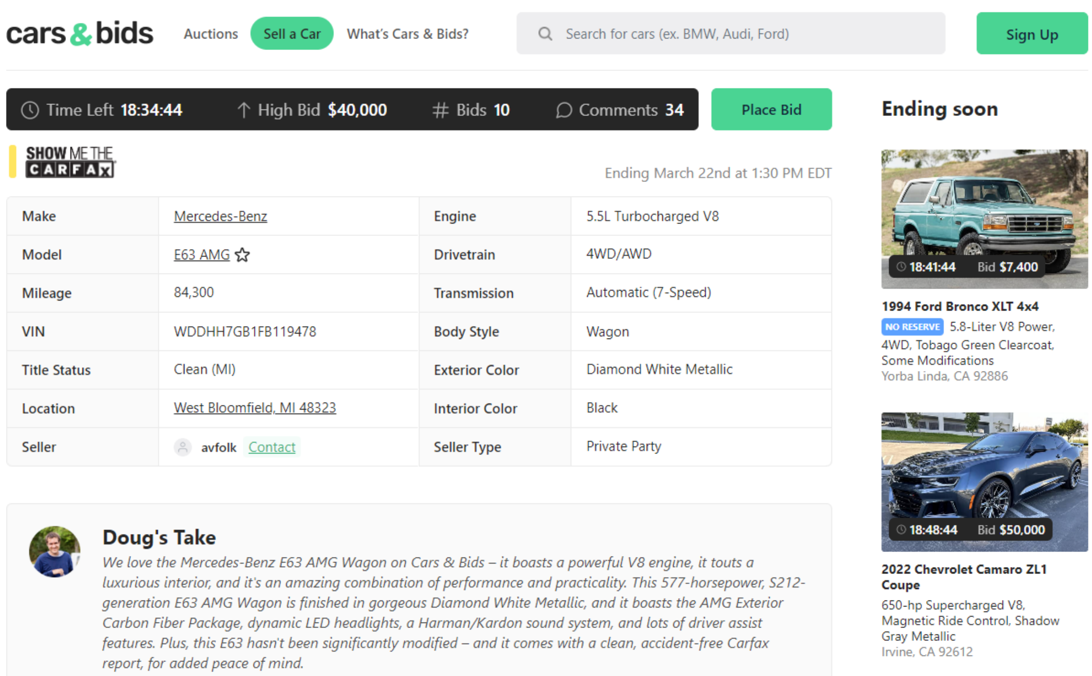

Figure 2: basic information.

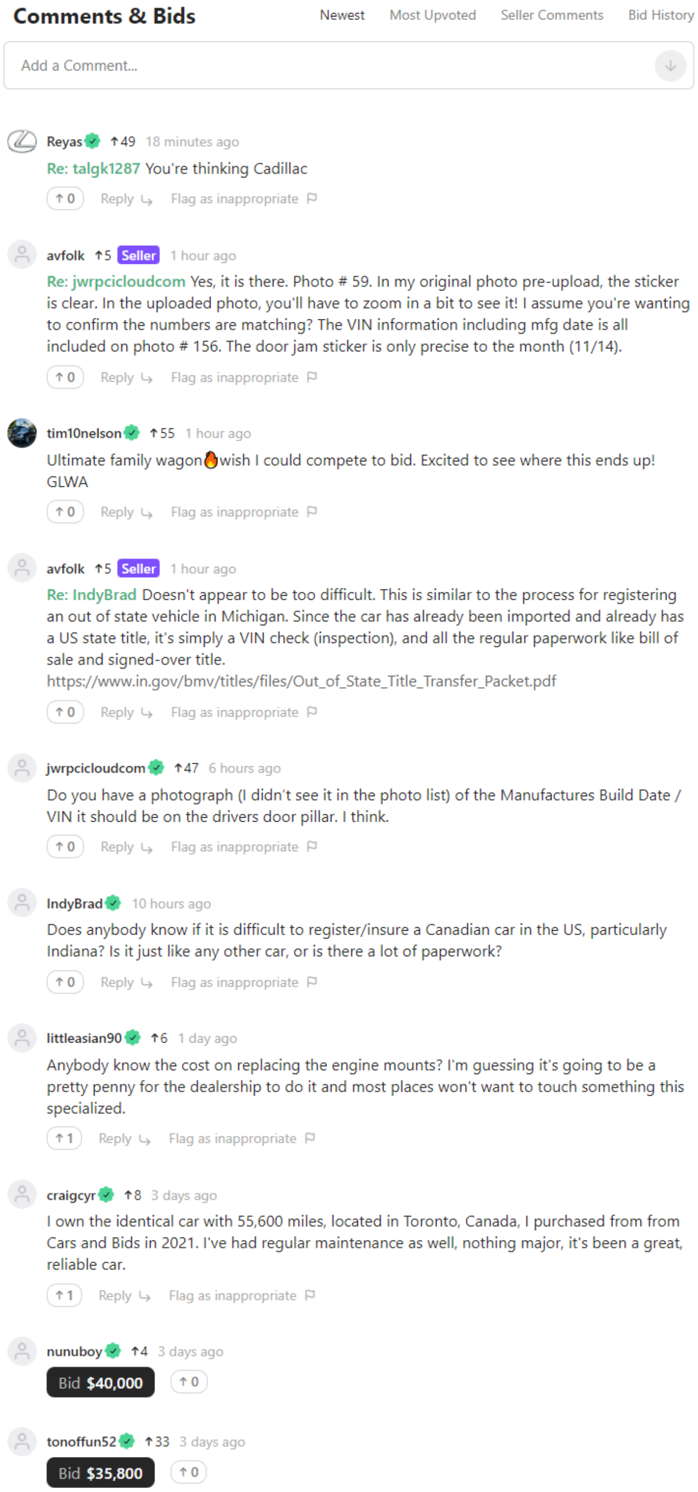

Figure 3: Comments & Bids "chatlog."

What is easily extracted from the detailed descriptive information, however, is the number of bullet points listed in each section: for each section, highlights, flaws, modifications, items, etc. are separated by bullet points. Thus, for each section we extract the number of highlights, number of flaws, etc.

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

3 Methodology

3.1 Sentiment Analysis

The goal of the research question—to investigate whether messages between bidders and sellers affects the sale price of an auction—first requires a method to quantitatively analyze the messages observed in the "Comments & Bids" chatlog section—from here forward, "comment section"—of each auction. To do this, a natural language processing (NLP) technique called "sentiment analysis" is employed.

Optimally, to employ sentiment analysis on the messages exchanged in the comment section of carsandbids.com auctions, a sentiment analysis model would be developed which uses the messages from carsandbids.com as its training corpus. A subset of all comment sections of all auctions would be used to create a dataset to train a sentiment analysis model. To create this dataset, the messages data would be "annotated" by humans, that is, messages would be labeled, e.g. "negative," "neutral," or "positive," according to the emotional sentiment of the message. Using this method would likely yield a model with the greatest performance measured in terms of accuracy of the sentiment analysis on the messages. However, given time and resource constraints, creating a sentiment analysis model in this way was infeasible.

Rather than constructing a training dataset through an annotation process, a next-best alternative is to use a pre-trained sentiment analysis model. Specifically, it is intuitive to use a model whose training corpus is similar to the corpus of online car auction messages. For this reason, a roBERTa based NLP model fine tuned on Twitter tweets for the purpose of sentiment analysis—from here forward "Twitter-roBERTa"—was used. Although certainly not exactly the same, both Twitter and carsandbids.com are online platforms on which commonly-used "internet slang" may be similar. Robustness of the results to different sentiment-analysis methods is discussed later.

For each auction, only relevant messages, that is, messages which were communicated prior to the end of the auction, were analyzed. carsandbids.com allows for messages to continue to be exchanged after the auction has ended, and such messages are not relevant. For each relevant message, a sentiment score—0, 1, or 2 to indicate negative, neutral, and positive, respectively—was obtained for each message using the Twitter-roBERTa model. Then, a mean sentiment score was calculated by taking the sum of the sentiment scores for the relevant messages divided by the number of relevant messages. This returns an "overall" sentiment score for the auction.

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

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, and I is the total number of messages in an auction n.

Finally, the sentiment scores were min-max scaled such that the auction with the lowest sentiment score had a score of 0 and the auction with the highest sentiment score had a score of 1.

3.2 Regression Analysis

The goal of this regression analysis is to determine whether the price, i.e. final bid, of an auction can be explained by the sentiment score of that auction. Thus, we seek to estimate β in the following model specification for an auction t:

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

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 as described in Table 2. γ_t, as described in Table 3, 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.

The intuition behind this log-log—rather than linear—specification is that vehicle prices may vary widely. Although γ_t controls for the intrinsic value of the vehicle of the auction t, a linear specification would presume that the covariates have a dollar effect on the price which would not change depending on the price of the vehicle. Intuitively, though, the empirical significance of that dollar effect is much higher for a lower priced vehicle than a higher priced vehicle and this should not necessarily be the case. Specification (1) accounts

Table 2: Variables in vector x

Variables	Type	Description
Ending bid	continuous	Ending bid or price at auction expiration.
Sentiment score	continuous	Sentiment score of the auction.
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.

Table 3: 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).

for this intuition such that β can be interpreted as a *percentage* effect on the price of the vehicle per 1% change in β.

4 Results

Regression results are found in Table 4 and we find an R-squared value of 0.874 meaning that 87% of the variation in price can be explained by the model. If the model is specified correctly, then, the remaining unexplained 13% of variation is due to the idiosyncratic tastes and preferences of the winning bidder.

We find that ln(sentiment score) is positive and statistically significant in explaining ln(price). All else constant, a 1% increase in sentiment score causes a 0.199% increase in price. Alternatively, we can more intuitively interpret the magnitude of β in the following manner: From Table 1 we see that the average price of a vehicle in our data is around \$31,000. As an example, let us say that one such average vehicle has a completely neutral sentiment score of 0.5. All else constant, if the sentiment score were to be increased by 100%, i.e. from 0.5 to 1, we would have an increase of

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

Thus, it would take our average vehicle from \$31,000 to \$35,588, an order of magnitude which seems reasonable.

Although initially the coefficients on number of bids and number of bidders may seem in conflict as one is positive and the other negative, we must remember that the interpretation of the coefficients is done holding all else equal/constant. If we hold the number of bidders constant and increase the number of bids, then we are effectively increasing the number of bids per bidder. Then, the positive coefficient on the number of bids makes sense and is consistent with results found in auction theory where more bids raises the price. We find

Table 4: 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

Note: Robust standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

that an additional bid raises the price by 0.7%. Conversely, if we hold the number of bids constant and increase the number of bidders, then we are effectively decreasing the number of bids per bidder, thus the negative coefficient on the number of bidders makes sense as the bidders participating are bidding fewer times. We find that an additional bidder may cause a 1.3% decrease in price.

The coefficient on reserve is positive and statistically significant. Economically, its interpretation is interesting as it suggests that all else equal, an auction which has a reserve has a price 17% higher than that of an equivalent auction with no reserve. Furthermore reserve has low correlation with the number of bids or the number of bidders, dissolving the intuitive hypothesis that an auction with no reserve may simply attract more bids, raising the price. Simply, it may be that bidders choose to place higher bids in auctions with a reserve knowing that in order to win the auction, the reserve price must be met.

The interpretations of the other coefficients are rather straightforward. Milage has the expected sign showing that higher milage decreases price, although the empirical effect is very small (magnitude of one hundredth of a percent). The positive coefficient on the number of photos is consistent with Lewis (2011), although again the effect is small. Coefficients on the number of known flaws and the number of modifications make intuitive sense as well. For every additional flaw, price may be 2.6% lower and for every additional modification, 0.4% higher.

4.1 Robustness and Sparsity

Due to the high dimensionality of the model caused by the large number of control dummy variables as seen in Table 3, we may be experiencing a loss in efficiency of the estimators in our OLS model specification (1). It may be the case that not every single control dummy in γ is needed. For example, in a case where two models of vehicle have the same intrinsic value, they could be represented in a single category or control dummy which would effectively

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)
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.

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

reduce the dimensionality of the model. A typical approach in empirical settings is for the researcher to use their domain knowledge in order to group many categories into fewer categories. However, this requires that the researcher be necessarily correct and unbiased.

Instead, we can use lasso inferential regression methods in order to robustly select controls such that the dimensionality of the model may be reduced. Such methods use the data present, removing the need for the researcher to be necessarily correct or unbiased in their specification. Appendix A shows results of double-selection (Belloni, Chernozhukov, and Hansen, 2014), partialling-out (Belloni et al., 2012), and cross-fit partialling-out (Chernozhukov et al, 2018) lasso linear regressions. Hyperparameters for the lassos are chosen using the plugin method developed by Bickel, Ritov, and Tsybakov (2009) and Belloni et al. (2012) with these lasso inferential methods in mind. These regression methods estimate β and λ but will use lassos to select a subset of the controls, γ, of the model specified in 1.

The results of the lasso linear estimation is found in Table 5. We find that the coefficients are largely similar indicating that the initial results are robust to these regression methods.

5 Discussion

Although I have shown that the sentiment score of an online car auction has a positive and significant effect on the selling price of the auction, what remains to be discussed is the way in which sentiment score actually affects the selling price. In other words, the sentiment score of an auction is simply a metric or an indicator that certain aspects of the messages have an effect on price, but we should explore what its economic meaning is, especially in terms of causality.

First, it may be that positive sentiments expressed in the message evoke an emotional response in other bidders such that they are enticed to bid. If it does indeed cause more bidding, this may cause the price of the auction to increase. In other words, bidders may be encouraged to bid more because of the positive sentiments expressed on the auction page. However, correlation between the number of bidders and the sentiment score of the auction is very low in absolute value. Furthermore, attempting to explain the number of bidders using sentiment score as well as the other covariates yields a model which does not explain a large portion of the variability in the number of bidders as shown by the low R-squared value.

Alternatively, it may be that the sentiment score is highly correlated with other regressors, indicating that although we observe that sentiment has an effect on price, it is instead the case that the effect of what we believe to be sentiment is disguised in the other variables. Again, this is not the case as sentiment score has low correlation in absolute value with the other regressors. Additionally, attempting to explain sentiment score by regressing it on the other covariates yields a model which does not explain a large portion of the variability in the number of bidders as shown by the low R-squared value.

Therefore, we may suggest that sentiment score represents some true causal component of explaining the ending price of the auction. Although sentiment is often defined as emotion or feeling, and one of the main drivers of the analysis at hand is sentiment analysis which is aimed at extracting emotion from text, is it necessarily true that the messages exchanged publicly in the auction contain only irrational, emotional information? As an example, take the following message posted on a carsandbids auction which ended May 9, 2022:

> 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.

> Good luck to whoever winds up owning it.
>
> I wouldn't be surprised to see this being a weird flex and just failing to meet an unreasonable reserve because there could be a lot more done if there was an honest effort to sell this car.

As shown in the example above, any commenter on the auction page may reveal private information through their message, that is, information about the vehicle that may be useful to other bidders' valuations but cannot be found elsewhere on the auction page. In this instance, the commenter is effectively expressing that the condition of the undercarriage is of high improtance to the value of the vehicle, and the unwillingness from the seller to show pictures of it may be a "red flag." Alternatively, take these other two comments posted on the same auction page:

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

and

> Makes sense, Thanks for the response brother!

Fundamentally, the difference between these shorter quotes and the previous longer quote is that the longer quote contains some valuable private information wheras the two shorter quotes here are simply expressing positive emotion. The problem arises when we consider that the sentiment analyzer is unable to distinguish between comments which contain valuable private information and those that do not. By the nature of the comment, the longer negative comment also contains emotionally loaded language which the sentiment analyzer will interpret as negative.

For bidders of the auction browsing the page, however, it is clear that the longer comment shown above may have a much greater impact on their valuation of the vehicle and as a result a greater impact on their bidding. The two shorter comments, although contribute to the emotional capital of the auction page, likely have a lesser effect on a bidders valuation and bidding.

The ability to distinguish between the two, then, may be important analysis to conduct in order to discover a true causal relationship between aspects of the messages in the comment section and the ending price of the auction. If a sentiment analyzer were to be created to be used specifically in the context of online car auction messages, such a sentiment analyzer could be trained as to distinguish between various "aspects." An aspect based sentiment analyzer may be able to distinguish between and associate emotionally charged words and the objects on which the emotions are being reflected. Thus, in this sentiment analysis method, aspects which have directly to do with the physical properties of the vehicle (or other factors affecting its value) may be given more weight. Alternatively, it may be possible to do an analysis of the messages using another NLP method called "topic modeling." In topic modeling, messages can be differentiated or categorized by "topic." And in this hypothetical analysis, we may be able to distinguish messages with information of value to other bidders from messages composed of pure sentiment.

References

Cars & Bids: Auctions of cool modern cars, trucks, and SUVs. URL https://carsandbids.com/.

George-Marios Angeletos and Jennifer La'O. Sentiments. *Econometrica*, 81 (2):739–779, 2013. ISSN 1468-0262. doi: 10.3982/ECTA10008. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA10008. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA10008.

Malcolm Baker and Jeffrey Wurgler. Investor Sentiment in the Stock Market. *Journal of Economic Perspectives*, 21(2):129–151, April 2007. ISSN 0895-3309. doi: 10.1257/jep.21.2.129. URL https://pubs.aeaweb.org/doi/10.1257/jep.21.2.129.

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. ISSN 1468-0262. doi: 10.3982/ECTA9626. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA9626. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA9626.

Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. Inference on Treatment Effects after Selection among High-Dimensional Controlsâ . *The Review of Economic Studies*, 81(2):608–650, April 2014. ISSN 0034-6527. doi: 10.1093/restud/rdt044. URL https://doi.org/10.1093/restud/rdt044.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, February 2018. ISSN 1368-4221, 1368-423X. doi: 10.1111/ectj.12097. URL https://academic.oup.com/ectj/article/21/1/C1/5056401.

Jia-Lang Xu and Ying-Lin Hsu. The Impact of News Sentiment Indicators on Agricultural Product Prices. *Computational Economics*, September 2021. doi: 10.1007/s10614-021-10189-4. MAG ID: 3196414579.

Gregory Lewis. Asymmetric Information, Adverse Selection and Online Disclosure: The Case of eBay Motors. *The American Economic Review*, 101(4):1535–1546, 2011. ISSN 0002-8282. URL https://www.jstor.org/stable/23045908. Publisher: American Economic Association.

MengâFeng Yen, YuâPei Huang, LiangâChih Yu, and YuehâLing Chen. A Two-Dimensional Sentiment Analysis of Online Public Opinion and Future Financial Performance of Publicly Listed Companies. *Computational Economics*, 59(4):1677–1698, April 2022. ISSN 0927-7099, 1572-9974. doi: 10.1007/s10614-021-10111-y. URL https://link.springer.com/10.1007/s10614-021-10111-y.