How Serious is Shill Bidding in Online Auctions? Evidence from eBay Motors

Kong-Pin Chen* Ting-Peng Liang† Shou-Yung Yin‡

Ted Chang§ Yi-Chun Liu¶ Ya-Ting Yu‖

September 18, 2020

Abstract

Using data from eBay Motors, this paper empirically estimates the prevalence of shill bidding and its effects on the outcomes of online auctions. Since most bidders have partially concealed their IDs, we first develop a procedure to identify the bidders. We then construct a shill-bidding index, wherein we split all auctions into one group which most likely contains shill bids, and another which does not. Our estimates indicate that around 9% of bidders are shill bidders and 22% of all listings contain shill bids. Using the instrumental variable approach for the regression, we show that shill bidding is actually two practices in one. First, it starts the auction with a reserve price as a ratio to the Blue Book price 0.08 lower than the usual auctions, thereby increases trade probability by 0.38%. Second, phantom bids are placed to compete with other bidders, which increase the ratio of the transaction price to Blue Book price by 0.055 on average, but have no effect on transaction probability. The construction of the shill index also enables us to test, and reject, a recent theory that links shill bidding and sniping.

Keywords: Online auctions, shill bidding, machine learning, eBay motors, reserve price

JEL code: D44, C38, C36

*Institute of Economics, Academia Sinica, Taiwan. E-mail: kongpin@gate.sinica.edu.tw.

†Department of Information Management, National Sun Yat-sen University, Taiwan. E-mail: liang@mis.nsysu.edu.tw.

‡Department of Economics, National Taipei University, Taiwan. E-mail: syyin@mail.ntpu.edu.tw.

§Institute of Economics, Academia Sinica, Taiwan. E-mail: f95323027@ntu.edu.tw.

¶Taiwan Semiconductor Manufacturing Company, Taiwan. E-mail: gfm2234@gmail.com.

‖Institute of Economics, Academia Sinica, Taiwan. E-mail: yyating@gmail.com.

1 Introduction

Although the practice of shill bidding is well known in the online auction literature (see, for example, Ockenfels et al. 2006 and Steiglitz 2007), its empirical prevalence and effects on auction outcomes are less understood.[1] There has been a sizable theoretical literature on how sellers can increase their revenue by placing shill bids,[2] but only very limited empirical literature.

In principle, if a bid comes from the same IP address as the seller's, then it strongly indicates shill bidding.[3] In practice, however, researchers have no access to IP addresses of bidders or sellers. Even if the full ID is observable, it might still be difficult to detect shill bidding. This is because shill bidding has become something of an industry, where professionals place phantom bids on behalf of sellers.[4] Without IP addresses, researchers can presumably identify shill bidders by looking for those who habitually enter a seller's listing and exhibit behavior resembling shilling (e.g., they bid often but rarely win). Unfortunately, for the majority of bidders, the researchers cannot observe their full IDs.[5] Furthermore, sellers who place shill bids can create as many IDs as they see useful, making identification of shill bidders more difficult. Finally, unless the bidder himself admits (i.e., there is a proof of "intention"), otherwise it is difficult to say for sure who actually shill bids, even if a bidder's behavior closely resembles one. In other words, a precise definition of shill bidding is hard to obtain. All these factors contribute to the reason why there is so little empirical literature on this topic.

Because of these difficulties, the literature has focused on experimental studies. Tre-

[1] According to eBay, "Shill bidding is when someone bids on an item to artificially increase its price, desirability, or search standing", and is forbidden by eBay. (See https://www.ebay.com/help/policies/selling-policies/selling-practices-policy/shill-bidding-policy?id=4353.) This paper is concerned with the first intention of shill bidding, i.e., to increase the item's price.

[2] See Graham et al. (1990), Bag et al. (2000) and Izmalkov (2004) for the independent valuation model; and Vincent (1995) and Chakraborty and Kosmopoulou (2004) for the common value model.

[3] For studies which trace user IP address to detect shill bidding, see Mamun et al. (2013) and Mamun (2015).

[4] See, for example, the discussion in NamePros: https://www.namepros.com/threads/giant-shill-bidding-operation-at-namejet-exposed.1013479

[5] Often, the bidders choose to partially conceal their IDs, so that only the first and last alphanumerics of the ID are observable.

vathan et al. (2008) used computer software to simulate the behavior of shill bidders in the laboratory, and found that shill bidding increased transaction price by 1%-25%. Nikitkov and Bay (2015) conducted field experiments selling computer memory cards online. In the half of the auctions in which they participated in shill bidding, the average price was 16-44% higher than in the other half, where they did not. Engelberg and Williams (2009) ran a field experiment in which 30 pairs of Chicago Cubs tickets were listed in online auctions. In the treatment group of each pair, they used the "discover and stop" strategy to shill in the last five minutes of the auctions. This strategy increased the average price of sold tickets from $73.79 (control group) to $76.65 (treatment group). Grether et al. (2015) conducted field experiments on an online used automobile auction platform (Copart Inc.) in two locations (New York and Texas). By manipulating the values of the minimum increment of bids, they successfully identified a group of seven bidders in Texas who were extremely likely to be shill bidders.[6] Kosmopoulou and De Silva (2007) and McCannon and Minuci (2020) conducted experiments to show that, when bidders are fully aware that the sellers can shill bid, they become conservative in their bids, so that the seller's revenue is less than when shill bidding is impossible. We are aware of only one empirical paper, Kauffman and Wood (2005), concerned with "reserve price shilling," rather than the type of shill bidding most of the literature addresses.[7]

The extant literature leaves certain questions open. First, how prevalent is shill bidding? Although this is a practice that increases the seller's revenue at the cost of the bidders (and indeed is forbidden by some platforms like eBay), it might be a phenomenon of only slight theoretical interest if it occurs rarely. Second, most of the experiments have been conducted using products with relatively low prices. Is shill bidding more prevalent for more expensive products such as cars? Third, and perhaps more importantly, what

[6] Contrary to common thinking, these bidders usually submitted aggressive bids with large increments over the standing price, in order to make the auction more "emotionally exciting."

[7] The purpose of reserve price shilling is to avoid the fee paid to the auction platform. Since the platform usually charges higher listing fees for items with higher starting prices (which is equivalent to an open reserve price in the auction), in order to reduce the fees, some sellers set a low starting bid, then enter and place shill bids up to the levels that they actually desire. However, for automobiles (our data), eBay charges a fixed listing fee, plus another fee when items are sold (McGrath and McGrath, 2010). Therefore, fees are independent of starting bids, and there is no need for reserve price shilling.

are the empirical implications of shill biddings for the auction outcomes, especially the sale probability and transaction price? This question is not easy to answer, as the literature has pointed out that sellers who intend to shill often deliberately start with a low reserve price (e.g., Steiglitz, 2007; Kauffman and Wood, 2005).[8] Therefore, the design of the field experiments in the literature, in which the sellers only place bids to compete without considering the reserve price, cannot fully capture the impact of shill bidding. Moreover, since the value of the reserve price not only is endogenous to whether the seller intends to shill but also influences the transaction price and sale probability, there will be possible estimation bias if this endogeneity is not controlled for. To correct this bias, we use three instrumental variables to endogenize the value of the seller's reserve price.[9] The extant empirical literature regarding the relationship between reserve price and the auction outcomes is rather mixed.[10] Our approach provides a chance not only to reinvestigate this issue, but also to quantify the extent by which the shill bidders reduce the reserve price to attract bids.[11]

Given that the bidder's ID is often partially concealed and that there is not a precise definition of shill bidding, our empirical study consists of two steps. First, we identify the bidders based on the habit and reputation score in each bidder's bidding history. Second, we construct an objective measure of how likely it is that one is a shill bidder, and whether an auction contains shill bids. The scores are constructed by considering several regularities typical of a shill bidder, and weights are assigned to each of the regularities to construct a summary score for its likelihood. Based on the scores and a maximum

[8] The practice is to set a low starting price to attract early bids. Once bids are placed, the shill bidder enters phantom bids to compete, sometimes even creating a bidding fever.

[9] As far as we know, no literature which investigated how reserve price affects transaction outcomes (see next section) has taken this endogenous nature into consideration. Choi et al. (2016), however, considered endogenous entry.

[10] Ariely and Simonson (2003), Häubl and Popkowski Leszezye (2003), Reiley (2006), Brown and Morgan (2009), and Choi et al. (2016) showed that an increase in reserve price usually decreased the number of bidders but increased the transaction price conditional on sale. Barrymore and Raviv (2009), Ku et al. (2006), Simonsohn and Ariely (2008), and Bajari and Hortaçsu (2003) found a negative effect of reserve price on revenue. Lucking-Reiley et al. (2007) and Einav et al. (2015) found no effect at all.

[11] It should be emphasized that the sellers are not given the option to set an open reserve price in eBay. The sellers are only given the option to set a secret reserve price. Therefore, the de facto reserve price for the seller is the starting price or bid, which the sellers must provide when they list an item (except for fixed-price listings). Hereafter, we will use the two terms "reserve price" and "starting bid" interchangeably.

likelihood procedure that together identify the shill bidders, a dummy is constructed to reflect whether a listing contains shill bids. Finally, a mixed-process model simultaneously estimates the trade probability and transaction price equations. Since reserve price is the seller's endogenous choice, in the estimation we also adopt an instrumental variable approach to endogenize the value of the reserve price. The IV approach shows that shill bidding actually consists of two practices, first setting a low reserve price, then placing phantom bids. The first aspect of shill bidding is one that is ignored in all previous literature. Our estimation shows that a shilled listing starts with a reserve price (as a ratio to Blue Book price) 0.08 lower than the average. Together with the phantom bids, it results in 0.38% higher sale rate, and a 0.055 higher transaction price (again, as a ratio to Blue Book price). These results imply that shill bidding strictly increases the seller's revenue.

Our empirical model also enables us to test a recent theory proposed by Bose and Daripa (2017). They theoretically showed that, if the bidders suspect seller's shilling, then they will strategically snipe (i.e., place bids only in the last minute of the auction).[12] Therefore, the only chance for the sellers to shill is near the end of the auctions. This theory has the strong implication that shill bids occur mainly just before the auction ends. Our construction of a shill bid index can help to test this implication. Specifically, if the theory is correct, the bidders who places bids near the end of the auction should be substantially more likely to be shill bidders than those who place bids earlier. Using several measures for the "last minute" of the auction, we do not find any positive correlation between sniping and the probability of being a shill bidder. In fact, the correlation is significantly negative. Our result thus rejects the theory's prediction.

2 Data

The data were collected from all the listings of Toyota cars in eBay Motors for a nine-month period from June 18, 2008 to March 6, 2009. During this period, there were 37,357 listings and 351,595 bid records. We first deleted 2,808 listings of new cars from the sample

[12] For discussion of sniping, see Steiglitz (2007) and Ockenfels et al. (2006).

for two reasons. First, there is no Blue Book value for new cars. The Blue Book value is important for the empirical studies related to the prices of used cars, as it is a good indicator of the value of used car based on its characteristics, and is widely consulted. In the literature, it serves as a good proxy that summarizes variables concerning a car's characteristics and conditions which affect its value.[13] Second, not only do new cars account for less than 8% of our sample, but also only 147 of the 2,808 new cars were sold (5% of all new cars; while the sale rate of cars in the whole sample is 19%). This implies that eBay Motors is predominantly a used-car platform. We further deleted observations that were posted-price listings and listings with best offer, as they were not auctions.

Our empirical study consists of two related parts, using different samples. In the first part, we dealt with the problem that for the majority of bidders, their IDs were concealed to various degrees. As to the sellers, although their IDs were fully revealed, some were missing during data collection. We deleted them from the sample, as there is no way to know whether two listings are from the same seller without their IDs. In all, there remained 7,653 sellers in the sample. The bidder's ID can be either fully revealed or partially concealed. Moreover, a seller can choose to conceal the IDs of all the bidders in her listing. In that case, all the bidders' IDs in that listing will be completely concealed. For a partially concealed ID, we only observed its first and last alphanumeric digits. Among the 181,819 bids that remained after we deleted new cars and non-auctions, 3,369 IDs were fully revealed, 151,230 were partially revealed, and 27,220 were completely concealed.[14] Similarly, we deleted all bids with completely concealed IDs, together with any listing whose sellers chose to conceal the bidders' IDs in their listings. That gave us 9,473 regular auctions and 8,968 buy-it-now auctions, and we used these 18,441 listings and 154,599 bids to identify the bidders, and then to construct the shill-bid index from the bidders' bidding histories in the listings. The definition of variables and the summary statistics of the sample for the first part are reported in Tables 1 and 2.

[13] See, for example, Wykoff (1973), Alberini el al. (1995), Raviv (2006) and Esteban and Shum (2007).

[14] If we consider listings, rather than bids, there were 2,026 listings in which all bidder's IDs were completely concealed, 303 listings in which all bidders revealed their full IDs, 16,849 listings in which bidders' IDs were partially revealed, and 1,289 listings in which some bidders fully revealed and some partially revealed their IDs.

In the second part of our empirical study, for the reason explained earlier, we used only the sample in which we could find the Blue Book values of the cars.[15] Also, the buy-it-now option on eBay is temporary, in the sense that if any bidder places a bid rather than exercise buy-it-now, the buy-it-now option disappears, and the listing then reduces to a regular auction. Therefore, the item in a buy-it-now auction can be sold either at the buy-it-now price (when a bidder exercises the option), or at the second highest bid (when a bidder places an eligible bid before buy-it-now is exercised). Since the former is essentially sold with a posted price, we further deleted buy-it-now auctions which were sold at a buy-it-now price from the sample.

After deleting listings whose Blue Book values could not be identified, the buy-it-now listings that were sold at the buy-it-now prices, and listings that had missing values, there remained 10,893 listings, of which 5,268 were regular auctions, and 5,625 were buy-it-now auctions. For these listings, the bidder's IDs were completely revealed in 1,153 bids, and partially revealed in 68,766 bids. There were therefore 69,919 bids, 10,893 listings and, as we will see in the next section, 25,896 distinct bidders in the sample for the second part of our empirical study.[16]

For each listing, the data contain (i) the auction characteristics, including the starting price, the auction duration posted by the seller, whether there is a secret reserve price, whether the item is sold, and the transaction price if it is; (ii) car characteristics, including car age, mileage, vehicle model and body type, fuel type, etc.; and (iii) seller's characteristics, such as whether the seller is a dealer, the seller's experience and their feedback scores. In addition, we also collected the bid history for each listing (for instance, bid amount and time of bid). Among the 10,893 listings, about 17.1% were sold, whose average ratio of transaction price to Blue Book price is 0.70. Tables 3 and 4 report the definitions of variables and summary statistics for the sample of the second part empirical study. Table 5 summarizes the numbers of listings, bidders, and sellers in the two parts of the empirical study, respectively.

[15] There are two main reasons that we could not find the Blue Book value of a used car. First, the seller did not provide the car age. Second, we did not know whether the seller was a dealer.

[16] The sample for the second part of the empirical study is thus a subject of the first part.

3 Empirical Model

In this section, we propose a procedure to identify the shill bidders and the listings which contain shill bids, together with how we handle the problem of partially concealed IDs. We then estimate equations of trade probability and transaction price, taking into consideration the influence of shill bids.

3.1 Identifying Bidders, Shill Bidders and Shilled Listings

In response to the difficulty of precisely defining a shill bidder mentioned earlier, the literature has tried to identify shill biddings through operational definitions, i.e., by the bidder's bidding behavior and the auction process, rather than their source. In that case, since bidders' IDs are usually concealed,[17] the first step towards detecting shill bids is to identify the bidders.

For this purpose, we extend a procedure proposed by Liu (2017). First, a preliminary identification is made through checking the ratings of the bidders. Unlike their IDs, the ratings of the bidders are fully observable. There were 1,691 bidders in the sample who fully revealed their IDs, among whom 697 bid at least twice. These 697 bidders were essentially the only full-ID bidders who we knew for sure appeared at least twice. The average daily change of ratings for these bidders was 0.12. For two bidders with partially concealed IDs but identical first and last alphanumeric characters, if the average difference in their ratings is smaller than 0.12 per day, we tentatively viewed them as the same bidder, otherwise they were viewed as two different bidders. The procedure produced 41,555 distinct bidders. To test how well this identification procedure worked, we applied it to the subsample of bidders whose IDs were completely revealed, and found 1,508 distinct bidders, of which 1,313 were correctly identified when we compared their full IDs (precision rate: 1,313/1,691 = 78%).

Since the criterion that two IDs are classified as identical if the daily change of ratings is less than 0.12 is probably too soft,[18] we applied the Bayesian Information Criterion

[17] For example, a seller whose ID starts with b and ends with k is usually shown as b***k. No matter how long the ID is, there are always 3 asterisks in the middle.

[18] This can be seen from the fact that there were actually 1,691 bidders with full ID, while the criterion

(BIC) to further identify them, based on the informational similarity of their bidding behavior. Specifically, for every bidder in every listing, we gathered information on his reputation score, the number of bids he placed, the length of time between his first and last bids, the lengths of time from his first and last bids to the end of the auction, and whether he won the item. Based on the information, we then compared the similarity of behavior between bidders who had identical concealed IDs to further classify them into smaller groups. With this procedure, we identified 50,994 distinct bidders. Therefore, together with the 1,691 bidders who fully revealed their IDs, we had 52,685 bidders in our sample. Again, we applied this procedure to the subsample of bidders whose IDs were completely revealed, and identified 1,511 distinct bidders. Compared to full IDs, we correctly identified 1,369 bidders, with a precision rate of 1,369/1,691 = 81%. Though not perfect, we believe this is accurate enough to justify our identification procedure as a first step in constructing a bidder's shill-bid index. In total, there were 18,841 listings, 52,685 distinct bidders, and 7,653 sellers in the first part of the empirical study, and 10,893 listings, 25,896 bidders, and 4,433 sellers in the second part of the empirical study (see Table 5).

After every bidder was assigned a distinct ID, we proceeded to investigate the likelihood that a bidder is a shill bidder through his bidding history, then construct a shill dummy for each listing, based on whether it contained a bidder very likely to a shill bidder. Kauffman and Wood (2005) identified shill bidders through their questionable bidding behavior. They reasoned that if a bidder chooses to bid in an auction when he has the chance to place the same or a lower bid in another concurrent auction featuring an identical item, then this bidder is likely to be a shill bidder. Shah et al. (2003) detected shill bidders through estimating how likely a bidder was to participate in and win auctions held by different sellers. Xu, Bates and Shatz (2009) used multiple criteria on the behavior of bidders to check for shill bidding, which included early bidding time and a large number of bids with small bidding increments. Dong, Shatz and Xu (2009, 2012) not only used various questionable bidder behaviors to identify potential shill bidders,

produced only 1,508 distinct bidders.

but also improved and verified the detection model by applying the Dempster-Shafter theory.

In this paper, we adopted a recent identification procedure proposed by Trevathan and Read (2009) and extended by Liu (2017). The underlying assumption of the procedure is that a shill bidder usually exhibits the following characteristics: He (i) usually bids exclusively in the auctions of one particular seller; (ii) tends to have a higher bid frequency;[19] (iii) tends to have very few wins for the auctions participated in; (iv) generally follows a new bid within a very short time; (v) usually out-bids rivals by minimum increments; and (vi) tends to appear early in an auction. As such, this procedure includes the approaches mentioned above as special cases. Following this identification procedure, bidder *i*'s shill-bidding probability in seller *m*'s listing *j*, *m(j)*, is related to seven variables[20]:

$\alpha_{i,m(j)}$ = The percentage of a particular seller's auctions that bidder i has participated in.

$\beta_{i,j}$ = The percentage of bids that bidder i has submitted in listing j.

$\gamma_{i,m(j)}$ = The proportion of wins in a particular seller's auctions that bidder i has participated in.

$\delta_{i,j}$ = The normalized average inter-bid lengths of time for bidder i in listing j.

$\varepsilon_{i,j}$ = The normalized average inter-bid increments for bidder i in listing j.

$\zeta_{i,j}$ = The normalized time between listing j's starting time and bidder i's first bid.

$\eta_{i,j}$ = The normalized time between bidder i's last bid and the auction's expiration time

in listing j.

These seven variables were used to assess the likelihood that bidder i is a shill bidder in listing *j*. Among them, α, β, and η are supposed to be positively, and γ, δ, ε, and ζ negatively, related to shill bidding probability. However, they might have different

[19] eBay shows two formats of bid history; one includes automatic bids (i.e., proxy bids submitted by eBay's automatic system on behalf of bidders), and one does not. Since automatic bids are not shill bids, but only a result of the bidder's submitting a relatively high bid, our data on bid frequency uses the latter count.

[20] For precise derivation of these variables in our data, please see the appendix.

degrees of importance in suggesting the likelihood that one bidder is a shill bidder. We therefore first adopted principal component analysis to generate the principle components. Next, a Gaussian mixture model was adopted to cluster the bidders through the principle components to find the most suspicious group of bidders.[21]

Principle component analysis is a widely used statistical method which generates new variables (named "principle components") through linear combination of the original variables while retaining as much variation as possible from the original data. It is particularly useful in reducing data dimensions and inspecting which variables are more important than the others. For our data, the first principle component (PC1) is a linear combination of the seven original variables above. The seven weights θ_{11}, ..., θ_{17} of PC1, each for one of the variables, are derived through:

$$\max_{\theta_{11},,\theta_{17}} \left[var(PC1)\right] \equiv$$
$$\max_{\theta_{11},,\theta_{17}} \left[var(\theta_{11}\alpha_{i,m(j)} + \theta_{12}\beta_{i,j} + \theta_{13}\gamma_{i,m(j)} + \theta_{14}\delta_{i,j} + \theta_{15}\varepsilon_{i,j} + \theta_{16}\zeta_{i,j} + \theta_{17}\eta_{i,j})\right], \quad (1)$$
$$s.t. \sum_{k=1}^{7} \theta_{1k}^2 = 1.$$

The original variables can turn into a single variable, while retaining as much variation from the original data as possible. The second principle component (PC2) is derived in ways similar to deriving PC1, but bears an additional restriction that $Cov(PC1, PC2)$=0:

$$\max_{\theta_{21},,\theta_{27}} \left[var(PC2)\right] \equiv$$
$$\max_{\theta_{21},,\theta_{27}} \left[var(\theta_{21}\alpha_{i,m(j)} + \theta_{22}\beta_{i,j} + \theta_{23}\gamma_{i,m(j)} + \theta_{24}\delta_{i,j} + \theta_{25}\varepsilon_{i,j} + \theta_{26}\zeta_{i,j} + \theta_{27}\eta_{i,j})\right], \quad (2)$$
$$s.t. \sum_{k=1}^{7} \theta_{2k}^2 = 1, \text{ and } Cov(PC1, PC2) = 0.$$

In other words, PC2 simplifies the original variables while retaining as much variation as possible that is not explained by PC1. Similarly, PC3 simplifies the original variables

[21] Details of how we constructed the seven variables and the Gaussian mixture model are in Appendix A and Appendix B, respectively.

while perserving as much variation as possible which is not explained by PC1 and PC2:

$$\max_{\theta_{31},,\theta_{37}} [var(PC3)] \equiv$$
$$\max_{\theta_{31},,\theta_{37}} \left[var(\theta_{31}\alpha_{i,m(j)} + \theta_{32}\beta_{i,j} + \theta_{33}\gamma_{i,m(j)} + \theta_{34}\delta_{i,j} + \theta_{35}\varepsilon_{i,j} + \theta_{36}\zeta_{i,j} + \theta_{37}\eta_{i,j})\right], \quad (3)$$
$$s.t. \sum_{k=1}^{7} \theta_{3k}^2 = 1,\ Cov(PC1, PC3) = 0,\ \text{and } Cov(PC2, PC3) = 0.$$

In this paper, the first three principal components are used. PC1 explains about 48.8% of the variation of the original variables, and PC2 and PC3 explain another 15.1% and 14.2%, respectively. The coefficients for the principle components are:

$$PC1 = 0.0361\alpha + 0.0908\beta + 0.0467\gamma + 0.4647\delta + 0.2672\varepsilon + 0.5905\zeta - 0.5944\eta,$$
$$PC2 = 0.0278\alpha + 0.0691\beta + 0.1681\gamma + 0.31\delta + 0.8171\varepsilon - 0.3169\zeta + 0.3194\eta, \quad (4)$$
$$PC3 = -0.0316\alpha - 0.0936\beta + 0.9808\gamma - 0.0959\delta - 0.1286\varepsilon + 0.0404\zeta - 0.0308\eta.$$

The value of PC1 in (4) has dominant weights on δ, ζ, and η. As mentioned above, η is positively, and δ and ζ are negatively, related to shill bidding. Furthermore, the coefficient of η is negative in PC1 of (4). Both imply that the value of PC1 should be low for a shill bidder. Similarly, PC2 has a positive and dominant value on ε, and PC3 has a positive and dominant value on γ, in (4). As mentioned above, shill bidding probability should be negatively related to both ε and γ. Therefore, a shill bidder should have low values of PC2 and PC3 too. PC1 captures how long the bidders waited to follow up with another bid and how late they join and leave an auction; PC2 reflects the bid increment they placed and PC3 indicates the winning probabilities of the bidders. Shill bidders are expected to respond quickly to other bids, join and leave auctions in the early stage, place bid increments as small as possible, and win as few as possible. In summary, shill bidders are expected to have low values for PC1, PC2, and PC3.

Our next step is to divide the bidders into different groups through the Gaussian mixture model. The basic assumption behind the Gaussian mixture model is that the sample data may be drawn from more than one population with unknown parameters. For example, our research data consist of shill bidders and non-shill bidders, and these two types of bidders may come from different populations and display different bidding behaviors. However, we don't know either the proportion of shill bidders or how differently they behave as compared to the non-shill bidders. For a given number of populations, the Gaussian mixture model uses maximum likelihood estimation on the sample data to estimate the population mean and variation of each underlying population. Moreover, for every data point, it estimates the probability it belongs to each of the estimated populations. Thus, we can group the bidders based on how likely it is that they belong to the estimated underlying populations.

By using the Gaussian mixture model on each principle component, bidders were divided into different normally distributed groups. In determining the optimal numbers of groups for PC1-PC3, we referred to the Bayesian Information Criterion (BIC) and Integrated Completed Likelihood criterion (ICL). The results indicated that for both PC1 and PC3, the values of BIC and ICL became stable beyond three to five clusters, suggesting that clustering PC1 and PC3 into four groups is optimal. However, for PC2, BIC suggested there should be eight clusters or more, while ICL showed the optimal number of clusters was 1. Moreover, the values fluctuated greatly between different simulations. Thus, we opted for a number of four clusters, the same as those for PC1 and PC3. This seems to be a reasonable number that is located between the numbers indicated by BIC and ICL, and goes along with the numbers of clusters of PC1 and PC3. Figures 2 to 4 show the density distributions of PC1, PC2, and PC3 and how the bidders are grouped. We then took a conservative approach, and assumed that a shill bidder must be in the lowest-scored group in at least two PCs.[22] Finally, the percentage of shill bidders was 8.85%.

A shilled listing was then defined as a listing participated in by at least one shill bidder. In our sample, 22.2% of the listings were deemed to be shilled listings. This number is similar to the numbers from some other literature. In Kauffman and Wood

[22] The reason we do not require a shill bidder to belong to the lowest group in all three PCs is that the lowest PC1 and PC2 groups have an empty intersection. This could indicate that trying to shill through placing low bids right after other bids in the early stage of an auction is not a common strategy adopted by the shill bidders. Bidders may shill bid through placing low bids, or through placing bids right after other bids in the early stage of an auction, but they do not use the two strategies together.

(2005), the "premium bids" occurred 23% of the time in rare coin auctions. Nikitkov and Bay (2010) identify 37 out of 186 (19.9%) auctions of cars, laptops, and perfumes as involving their definition of shill bidding.[23] The ratio of reserve price to Blue Book price of the shilled listings is on average 0.134, and that of the non-shilled listings is 0.466. These are important facts which confirm the intuition in the literature (e.g., Steiglitz, 2007; Kauffman and Wood, 2005) that the sellers who practice shill bidding often start with a low reserve price.

Figure 1 plots the bidding process of an auction in our data in which six bidders have participated. Bidder 1 first placed a bid of $750. Then, in a series of almost continuous bids with minimum increments of $5, bidder 2 raised the price to $520. After bids from bidder 3 and 4, and then a high bid from bidder 5, bidder 1 entered again with a series of 5 bids. Bidder 1 eventually lost to bidder 6, who won with $1055. Note that there were two bidders whose behavior resembled a shill bidder, 1 and 2. However, only bidder 2 was deemed a shill bidder by our procedure. There are two reasons for this. First, bidder 1 entered the auction first and with a high bid, which a shill bidder rarely does. Second, although he entered a series of bids on October 28, not only was the increment large ($50) but also the bids were relatively late. On the other hand, bidder 2 not only bid early, with minimum increment, but also avoided late bids (so he would not win by accident).

3.2 The Effects of Shill Bids

In this section, we estimate the trade probability and transaction price, while taking the influence of shill bids into consideration. Since there is a transaction price only when there is a trade, we adopt the standard Heckman two-stage procedure for the estimation. In the first stage, a probit model estimates the factors that affect the transaction probability of a car. In the second stage, an OLS model estimates the transaction prices of the items

[23] Using various criteria and machine-learning techniques, there is a large variation in the proportion of shill bidders in the informational science literature. For example, Alzahrani and Sadaoui (2018) recorded 26%, Alzahrani and Sadaoui (2020) recorded 10.7%, Ganguly and Sadaoui (2018) recorded 5%, and Anowar and Sadaoui (2020) 9%.

that are sold. The first-stage estimation is as follows:

$$
\begin{aligned}
Sold_j = {} & \alpha_0 + \alpha_1 \times StBidR_j + \alpha_2 \times Shill\ Dummy_j + \alpha_3 \times StBidR_j \times Shill\ Dummy_j \\
& + \alpha_4 \times SRP_j + \alpha_5 \times Competitor_j \\
& + \alpha_6 \times ln(Seller\ Score_{m(j)}) + \alpha_7 \times Warranty_j \\
& + \alpha_8 \times AMileage_j + \alpha_9 \times Seller\ is\ Dealer_{m(j)} \\
& + \alpha_{10,d} \times Posted\ Duration_{d(j)} + \alpha_{11,f} \times Fuel\ Type_{f(j)} + \alpha_{12,l} \times Model_{l(j)} \\
& + \alpha_{13,c} \times Car\ Body\ Type_{c(j)} + \alpha_{14,v} \times Vehicle\ Condition_{v(j)} + \varepsilon_{1,j}
\end{aligned} \tag{5}
$$

In the equation, $Sold_j$ equals 1 if the item is sold, and 0 if otherwise. $m(j)$ is seller m in listing j, $d(j) = 1$, 2, 3 is a dummy representing the auction duration of 3, 5, and 7 days, respectively; $f = 1$, ..., 6 is a dummy variable which denotes the vehicle's fuel type; and $l = 1$, ..., 19 is a dummy variable which denotes the 19 models of vehicles. Finally, c= 1, ..., 9, denotes the car body types, and $v = 1$, 2 denotes whether the vehicle title bears the designation "clear" or "salvage". Table 4 reports the summary statistics of car characteristics.

Understandably, there is a large price variation among the cars, depending on their models and characteristics. In order to control for this variation, all prices are normalized to be their values relative to the Blue Book prices. The variable $StBidR_j$ is the ratio of the starting bid to the car's Blue Book value. Since eBay does not allow the seller to post an open reserve price, the de facto reserve price is the starting bid, which is a requirement when a seller lists an item in the regular auction and the buy-it-now auction. In our model, we view the starting bid as the reserve price chosen by the seller. $AMileage_j$ is the car's mileage divided by its age. Although a car's Blue Book value is supposed to be a summary statistic of a car's characteristics that affect its price, it is not directly related to trade probability. A higher Blue Book price does not imply a lower trade probability: it only reflects better characteristics of the car. Rather, characteristics such as fuel type, model, body type, and vehicle condition can better control for the bidder's preference that might affect the car's trade probability. We therefore used these variables to control for the car's trade probability, rather than its Blue Book value.[24]

The seller's reputation score and whether she is a dealer were both expected to influence trade probability, and were used as control variables. The variable *Seller Score*, used as a proxy for reputation, is the total number of positive minus negative feedback for transactions. The number of competitors, defined as the number of similar cars (with the same model and age) on eBay Motors during the time an item was listed, is expected to negatively affect trade probability, while warranty is expected to have a positive effect. Setting a secret reserve price is widely known to have a negative effect on trade probability (Katkar and Reiley, 2006 and Bajari and Hortaçsu, 2003), and we used a dummy (since we did not know its value), SRP, to control for it. This variable is important, as in our sample more sellers used it than not (see Table 6). We followed Lucking-Reiley et al. (2007) to include dummy variables (the $d(j)$'s) for various lengths of posted duration. The posted duration of 10 days was used as the basis of comparison.

The second-stage estimation for the transaction price is an OLS estimation:

$$\begin{aligned} WinBidR_j = {} & \beta_0 + \beta_1 \times StBidR_j + \beta_2 \times Shill\ Dummy_j + \beta_3 \times StBidR_j \times Shill\ Dummy_j \\ & + \beta_4 \times SRP_j + \beta_5 \times Competitor_j + \beta_6 \times ln(Seller\ Score_{m(j)}) \\ & + \beta_7 \times Warranty_j + \beta_8 \times AMileage_j \\ & + \beta_{9,v} \times Vehicle\ Condition_{v(j)} + \beta_{10,d} \times Posted\ Duration_{d(j)} + \varepsilon_{2,j}. \end{aligned} \tag{6}$$

The dependent variable, $WinBidR$, is the winning bid divided by the car's Blue Book value. The literature has adopted two measures of how the characteristics of used cars affect their price premiums. The first was the difference between the car's Blue Book value and its price. In this vein, Andrews and Benzing (2007) used two different measures for the premium or discount for each vehicle: The difference between the highest bid of each listing and the Blue Book value of an automobile, and the difference between the winning price and the Blue Book value. Both measures were based on price difference. The second approach was to use the ratio of the winning price to the Blue Book value instead

[24] We have run a regression adding Blue Book price as a control variable and, as expected, the coefficient is not statistically significant.

of their differences, which was also adopted by Bajari and Hortaçsu (2003). This was the measure that we adopted, as we have already measured the reserve price by ratio. The seller's reputation score is expected to have a positive effect on the transaction price. The influence of warranty on car price is obvious.

As indicated in Newberry (2015), the winning price of a car can be influenced by how many similar cars are in competition with each other, and by the mileage and age of the car. Therefore, we included $Competitor$ and $AMileage$ to control for the effects of competition and a vehicle's average mileage.[25]

3.3 The Endogenization of Reserve Price

As mentioned earlier, the sellers who practice shill bidding usually start with a low reserve price. They use a low reserve price first to attract early bids, then enter to compete, even hoping to create an atmosphere of a bidding fever, so that the bidders are thrilled by the fervor of competing against each other, and on the way increase their willingness-to-pay beyond the level of what it normally would be.[26] Note that even for a the seller who does not intend to shill bid, she will also set the value of the reserve price based on her own consideration. This implies that the reserve price is an endogenous variable, which also happens to correlate with trade probability and transaction price.

To correct for possible bias arising from this endogenous choice, we used an instrumental variable approach. Three exclusive variables, the seller's experience, the BIN option, and the average starting price of other listings by the same seller, served as the instrumental variables to estimate the reserve price. The reason for choosing seller's experience as an instrumental variable is that experience obviously influences the ability of the seller to set the reserve price. While we do not impose any prior restriction between experience and the value of the reserve price, there is reason to believe that more experienced sellers are more likely to set a lower starting bid, while less experienced ones might be more concerned with the risk that the car is sold at a low price, and are more prone to

[25] Average mileage is not in the Blue Book reference.

[26] For relevant literature, see Heyman et al. (2004), Ku et al. (2005), Jones (2011), Adam et al. (2011), and Adam et al. (2015).

set a higher starting bid. The average starting price of the seller's other listings during the whole study period might capture the seller's habit of setting the starting price at a certain percentage of the Blue Book value. Finally, according to the theory in Chen et al. (2017) that the optimal reserve price is higher in the auctions with a buy-it-now option than without, we expect a seller to set a higher reserve price when she lists an item with the buy-it-now option. The model for estimating the reserve price is:

$$\begin{aligned} StBidR_j = {} & \gamma_0 + \gamma_1 \times ln(Seller's\ Experience_{m(j)}) + \gamma_2 \times \overline{StbidR}_{m(-j)} \\ & + \gamma_3 \times BIN_j + \gamma_4 \times Shill\ Dummy_j \\ & + \gamma_5 \times SRP_j + \gamma_6 \times Competitor_j + \gamma_7 \times ln(Seller\ Score_{m(j)}) \\ & + \gamma_8 \times Warranty_j + \gamma_9 \times AMileage_j \\ & + \gamma_{10,v} \times Vehicle\ Condition_{v(j)} + \gamma_{11,d} \times Posted\ Duration_{d(j)} + \varepsilon_{3,j}. \end{aligned} \tag{7}$$

For the seller's experience, some studies used her feedback or reputation as a proxy (Kauffman and Wood, 2006; Hu and Wang, 2010; Newberry, 2015), while others used the number of days since the seller joined the auction platform (Chen et al., 2013; Scott, Gregg, and Choi, 2015). In our paper, we used how many transactions the seller had made within a year as a proxy for experience, because we believed that it was more relevant to the current transaction.

Wooldridge's (1995) score test statistic was used to test the endogeneity of reserve price, and we also conducted tests of weak instruments and over-identification restrictions for the starting bid to make sure our selection of the instrumental variables was appropriate and valid.[27] In particular, the results showed that a seller who was more experienced, or who set lower reserve prices in her other listings, was more likely to set a lower reserve price. Also, consistent with Chen et al. (2017), a seller who listed an item with a buy-it-now option set a higher reserve price. The test and estimated results are summarized in Table 9 , where † indicates the 1% level of significance for the score test.

[27] We rejected the nulls of weak instruments and overidentifying restrictions ($\chi^2 = 54.34, p - value = 0.000$; $\chi^2 = 87.66, p - value = 0.001$), which indicated that the variables ($Seller's\ Experience_{m(j)}, \overline{StbidR}_{m(-j)}, BIN_j$) were sufficiently correlated with reserve price.

4 Results and Discussion

Tables 7 and 8 present the estimation results, without controlling for the endogeneity of the reserve price, and Table 9 for when it is considered. Also, to see the importance of shill bids in affecting auction outcomes, in the second and third columns of Table 7 we report the regression results and marginal effects without controlling for the shill bid (i.e., without the shill bidding dummy), while in Table 8 are the results and effects when we control for it. The reserve price's effect on transaction probability is negative and highly significant regardless of whether shill dummy is controlled for (the coefficients are -1.696 and -1.783; the marginal effects are -0.280 and -0.289, respectively), a result consistent with almost all literature.[28] The effect of reserve price on the transaction price is positive and highly significant, again regardless of whether the shill dummy is controlled for (0.52 and 0.566, respectively). This is consistent with the literature which shows a positive effect of reserve price on price,[29] but inconsistent with some others.[30] Note that the impacts of the reserve price have been consistently smaller in Table 7 than in Table 8. The reason for this is clear: these effects are partially ameliorated by the seller's shill bidding so that, when it is controlled for (in Table 8), the true and large effects emerge.

In the price equation, the coefficient for shill dummy in Table 8 is 0.0766 and is significant at the 1% level, implying that shill bidding increases the ratio of the transaction price to Blue Book price by about 0.0766. However, since a shill bidder also sets a lower reserve price, this has to be tempered by the interaction term between shill dummy and reserve price, which is negative (-0.334) and significant at the 1% level. Table 8 also shows that shill bid itself does not increase sale rate (the coefficient of shill bid dummy is not significant.).[31] What increases sale rate is the fact that shill bidder sets a lower reserve

[28] Ariely and Simonson (2003), Bajari and Hortaçsu (2003), Häubl and Popkowski Leszezye (2003), Reiley (2006), Ku et al. (2006), Simonsohn and Ariely (2008), Brown and Morgan (2009), Barrymore and Raviv (2009), Einav et al. (2015), and Choi et al (2016).

[29] Ariely and Simonson (2003), Häubl and Popkowski Leszezye (2003), Reiley (2006), Brown and Morgan (2009), Barrymore and Raviv (2009), Einav et al. (2015), and Choi et al (2016).

[30] Bajari and Hortaçsu (2003), Kamins et al. (2004), Ku et al. (2006), and Simonsohn and Ariely (2008) find a negative effect on transaction price. Lucking-Reiley et al. (2007) and Einav et al. (2015) find no effect of reserve price.

[31] This can be easily understood by the fact that shill bids are rarely placed before there has been any bid.

price, as the coefficient for the cross term with reserve price is negative and significant. This is consistent with the literature (Steiglitz, 2007; Kauffman and Wood, 2003) which postulates that the sellers who practice shill bids generally set a lower reserve price, then drive up the transaction price by placing phantom bids.

Note that although the reserve price's effect on the transaction price has a flatter slope for listings with $ShillDummy = 1$, they also have a greater intercept. This is exactly because the sellers who shill usually start with a low price. What is more important, Table 8 shows that, for a listing containing shill bids, if there is a, for example, 0.1 decrease in reserve price, its sale rate increases by $(0.289 - 0.238) \times 0.1 = 0.51\%$,while its transaction price, conditional on sale, also increases by $0.0766 - (0.566 - 0.334) \times 0.1 \cong 0.0534$ as a ratio to Blue Book value. In other words, reducing the reserve price increases both the sale probability and the transaction price as long as the reduction is less than 0.33, meaning that if the shill bidding increases both sale rate and transaction price only if he does not engage it in a way but setting a very low reserve price. In this sense, the literature that shows auctions with a low reserve price yield higher revenue might be partially an illusion caused by shill bidding.

Most of the other variables also influence the trade probability and transaction price in a way consistent with the intuition and the literature. For example, longer listing duration tends to result in higher transaction price. Higher mileage reduces transaction price, and having a warranty increases transaction price. Warranty also has a negative effect on transaction probability, perhaps exactly because cars with a warranty are more expensive. Dealers have a harder sale, a result that is also found in several other studies (e.g. Andrews and Benzing, 2007; Lewis, 2011). Setting a secret reserve price reduces trade probability and increases transaction price, which is consistent with the literature.[32] The number of simultaneous competing listings reduces transaction probability, but increases transaction price. The former result is intuitive, but we cannot explain the latter.

Table 9 shows the regression results when the endogeneity of reserve price is considered. Experience is negative and significant at the 1% level, while $\overline{StbidR}_{m(-j)}$ and BIN

[32] Katkar and Reiley (2006), and Bajari and Hortaçsu (2003).

are positive and significant at the 1% level. The influences of the three instrumental variables on reserve price are thus consistent with the literature and our expectation that a more-experienced seller sets a lower reserve price, that a seller tends to set a higher reserve price if she sets higher ones in other listings, and that the listing with BIN option has a higher reserve price than that without the BIN option. The estimated coefficients and significant values of most variables are the same between the Heckman model and the endogeneity model. The IV approach also shows that a shill bidder starts an auction with a ratio of reserve price/Blue Book price 0.08 lower on average. This increases trade probability by $(0.248 - 0.2) \times 0.08 = 0.38\%$ and transaction price by $0.0726 - (0.554 - 0.334) \times 0.08 = 0.055$.

It is interesting to see that the coefficient for the shill dummy is not statistically significant in the trade probability equation, but is highly significant in the price equation. This echoes the results in the Heckman model above, in that what increases the trade probability is not the shill bid per se, but simply the fact that shill bidders set a lower reserve price. Shill bidding increases the ratio of transaction price to Blue Book value by 0.0726, which can be a substantial amount considering that the price of a used car is generally in the thousands. The effect of reserve price on the sale rate and the transaction price are both lower than in the Heckman model. The differences are not substantial, meaning that the consideration of the endogenous reserve price changes the estimation results quantitatively, but not qualitatively. This said, the IV approach proves two aspects of shill bidding that cannot be shown by the Heckman model. First, it shows that shill bidding is two practices in one, one is to set a low reserve price to allure entry; the other to play phantom bids to increase the price. The literature has invariably focused on the second aspect, and therefore cannot appropriately assess the full impact of shill bidding. Second, we also quantify the degree that the seller lowers the reserve price, so that we can calculate the average increase in both transaction rate and transaction price.

5 Last-Minute Shill Bids

In a recent paper, Bose and Daripa (2017) propose a new theory of sniping, based on the bidder's strategic reaction to shill bids. They reason that, when the bidders are aware of the possibility of shill bidding, a strategic response to avoid competing with the shill bidders is to delay placing bids and, in particular, to place bids right before the last minute of the auction. In that case, the only possible time for the sellers to place phantom bids is during the last minute, when there is a positive probability that their bids cannot go through.

If this theory is true, then shill bids will occur primarily just before the end of the auctions. The empirical implication for this is that there will be a larger share of shill bids during the last minutes of an auction than before the last minutes. It should, however, be emphasized that this theoretical prediction actually conflicts with the common thinking in the literature. As can be seen in Section 3.1, most of the literature takes early bids, rather than last-minute bids, as one of the signs of shill bids.[33] In fact, in defining shill indexes, this criterion is explicitly taken into consideration through ζ and η. To do justice to this theory, we again clustered the 52,685 bidders from the first-part data through principle component analysis and a Gaussian mixture model, but this time without ζ and η. As a result, the first two principle components are used, and the results are:

$$\begin{aligned} PC1 &= 0.0172 \times \alpha + 0.1315 \times \beta + 0.0896 \times \gamma + 0.7733 \times \delta + 0.6127 \times \varepsilon, \\ PC2 &= -0.0284 \times \alpha - 0.0842 \times \beta + 0.9927 \times \gamma - 0.0764 \times \delta - 0.0291 \times \varepsilon, \end{aligned} \tag{8}$$

The two principal components explain 63.9% of the variations of the original indexes, excluding ζ and η. Although the 3rd principal component can explain an additional 19.3% of the variations, the distribution of PC3 doesn't clearly show a group susceptible of being shill bidders, so we elect to cluster the bidders with only the first two principle components.

Similar to what constitutes shill bidding, there is not a precise mathematical definition

[33] Kauffman and Wood (2003); Xu et al. (2009); Dong et al. (2009); Trevathan and Read (2009), and Liu (2017).

of sniping, and especially of what constitutes the "last minute" of an auction. Researchers usually use the last 5 or 10 minutes of an auction as a threshold.[34] We propose three definitions on what constitutes the "last minute" of an auction. Specifically, we use the last 30 minutes, the last 5 minutes, and the last 5% of the auction duration as three possible thresholds for the last minute. Our aim is to test whether the bidders who are most likely to be shill bidders predominantly bid in the last minute. If the prediction in Bose and Daripa (2017) is correct, then the proportion of shill bidders should be substantially and significantly higher in the last minute. The results are reported in Table 10. The proportion of shill bidders who appear during the last minute is actually smaller, and is significant at the 1% level through t-tests, for all three thresholds. Our empirical results therefore do not support the theoretical prediction in Bose and Daripa (2017). Since their model assumes that bidders are aware of shill biddings and strategically react to it, our result suggests that most bidders are probably not aware of shill bids or, even if they are, do not believe it important enough for them to react to it.

6 Conclusion

In this paper, we empirically reinvestigated the extent to which shill biddings affect auction outcomes. To overcome data limitations, we first used a procedure to identify bidders who partially concealed their IDs. Based on behavioral assumptions on shill bidders, we then constructed a shill index for the bidders, and identified listings which contained shill bids. It is shown that shill bidding consists of two practices. First, it starts with the auction with a lower reserve price. In our data, the reserve price (as a ratio to the Blue Book price) of a listing that contains shill bids is reserve price 0.08 lower than average. Second, after there is a bid, the shill bidder enters phantom bids to compete in the auction to increase the transaction price. We want to emphasize that since the first aspect of shill bidding is one that is ignored in all the previous literature, which thus fails to address the full impact of shill bidding. In all, shill bidding increases the ratio of transaction price to Blue Book price by 0.055, and transaction probability

[34] See the survey in Ockenfels et al. (2006).

by 0.38%. However, the increase in transaction rate does not comes from shill bidding per se, but simply because shill bidders set a lower reserve price. Since it increases both transaction rate and price, shill bidding strictly increases the seller's expected revenue.

Finally, we showed that bidders who snipe were less likely to be shill bidders. These results contradicted the theoretical prediction of Bose and Daripa (2017). Possible explanations might be that the bidders are not aware of shill bids when they participate in the auctions, or do not feel they warrant attention, or that the sellers might think that placing shill bids in the last minute is too risky.

References

[1] Adam, M.T.P, Krämer, J., Jähnig, C., Seifert, S., and Weinhardt, C., (2011), "Understanding Auction Fever: A Framework for Emotional Bidding," *Electronic Markets*, 21, 197-207.

[2] Adam, M.T.P., Krämer, J., and Müller, M.B. (2015), "Auction Fever! How Time Pressure and Social Competition Affect Bidders' Arousal and Bids in Retail Auctions," *Journal of Retailing*, 91(3), 468-485.

[3] Alberini, A., Harrington, W., and McConnell, V. (1995), "Determinants of Participation in Accelerated Vehicle-Retirement Programs" *The RAND Journal of Economics*, 93-112.

[4] Alzahrani, A., and Sadaoui, S. (2018), "Scraping and Preprocessing Commercial Auction Data for Fraud Classification," *ArXiv*, abs/1806.00656.

[5] Alzahrani, A., and Sadaoui, S. (2020), "Clustering and Labeling Auction Fraud Data," In *Data Management, Analytics and Innovation*, 269-283, Springer, Singapore.

[6] Andrews, T., and Benzing, C. (2007), "The Determinants of Price in Internet Auctions of Used Cars," *Atlantic Economic Journal*, 35(1), 43-57.

[7] Anowar, F., and Sadaoui, S. (2020), "Detection of Auction Fraud in Commercial Sites," *Journal of Theoretical and Applied Electronic Commerce Research*, 15(1).

[8] Ariely, D., and Simonson, I. (2003), "Buying, Bidding, Playing, or Competing? Value Assessment and Decision Dynamics in Online Auctions," *Journal of Consumer Psychology*, 13(1&2), 113-123.

[9] Bag, P.K., Dinlersoz, E.M., Wang, R. (2000), "More on Phantom Bidding," *Economic Theory*, 15(3), 701-707.

[10] Bajari, P., and Hortaçsu, A. (2003), "The Winner's Curse, Reserve Prices, and Endogenous Entry: Empirical Insights from eBay Auctions," *The RAND Journal of Economics*, 34(2), 329-355.

[11] Barrymore, N., and Raviv, Y. (2009), "The Effect of Different Reserve Prices on Auction Outcomes," Robert Day School of Economics and Finance Research Working Paper No. 2009-13.

[12] Bose, S., and Daripa, A. (2017), "Shills and Snipes," *Games and Economic Behavior*, 104, 507-516.

[13] Brown, J., and Morgan, J. (2009), "How Much Is a Dollar Worth? Tipping Versus Equilibrium Coexistence on Competing Online Auction Sites," *Journal of Political Economy*, 117(4), 668-700.

[14] C. Biernacki, G. Celeux and G. Govaert, "Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 7, pp. 719-725, July 2000, doi: 10.1109/34.865189.

[15] Chakraborty, I. and Kosmopoulou, G. (2004). "Auctions with Shill Bidding," *Economic Theory*, 24(2), 271-282.

[16] Chen, J.R., Chen, K.P., Chou, C.F. and Huang, C.I. (2013), "A Dynamic Model of Auctions with Buy-it-Now: Theory and Evidence," *Journal of Industrial Economics*, 61(2), 393-429.

[17] Chen, K.P., Ho, S.H., Liu, C.H. and Wang, C.M. (2017), "Optimal Listing Strategy in Online Auctions," *International Economic Review*, 58(2), 421-437.

[18] Choi, S., Nesheim, L., and Rasul, I. (2016), "Reserve Price Effects in Auctions: Estimates from Multiple Regression-Discontinuity Designs," *Economic Inquiry*, 54(1), 294-314.

[19] Dong, F., Shatz, S.M., and Xu, H. (2009), "Inference of Online Auction Shills using Dempster-Shafer Theory," *2009 Sixth International Conference on Information Technology: New Generations* (pp. 908-914). Washington DC: IEEE Computer Society.

[20] Dong, F., Shatz, S.M., Xu, H., and Majumdar, D. (2012), "Price Comparison: A Reliable Approach to Identifying Shill Bidding in Online Auctions?" *Electronic Commerce Research and Applications*, 11(2), 171-179.

[21] Engelberg, J. and Williams, J. (2009). "eBay's Proxy Bidding: A License to Shill," *Journal of Economic Behavior and Organization*, 72(1), 509-526.

[22] Einav, L., Kuchler, T., Levin, J., and Sundaresan, N. (2015), "Assessing Sale Strategies in Online Markets Using Matched Listings," *American Economic Journal: Microeconomics*, 7(2), 215-247.

[23] Esteban, S., and Shum, M. (2007), "Durable-Goods Oligopoly with Secondary Markets: The Case of Automobiles", *The RAND Journal of Economics*, 38(2), 332-354.

[24] Ganguly, S., and Sadaoui, S. (2018, June), "Online Detection of Shill Bidding Fraud Based on Machine Learning Techniques", In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, 303-314. Springer, Cham.

[25] Graham, D.A., Marshall, R.C., Richard, J.-F. (1990), "Phantom Bidding against Heterogenous Bidders," *Economics Letters*, 32, 13-17.

[26] Grether, D., Porter, D., and Shum, M. (2015), "Cyber-shilling in Automobile Auctions: Evidence from a Field Experiment," *American Economic Journal: Microeconomics*, 7(3), 85-103.

[27] Häubl, G., and Popkowski Leszczyc, P. (2003), "Minimum Prices and Product Valuations in Auctions," *Marketing Science Institute Reports*, 03-117.

[28] Heyman, J.E., Orhun, Y., and Ariely, D. (2004), "Auction Fever: The Effect of Opponents and Quasi-Endowment on Product Valuations," *Journal of Interactive Marketing*, 18, 7-21.

[29] Hu, Y., and Wang, X. (2010), "Country-of-Origin Premiums for Retailers in International Trades: Evidence from eBay's International Markets," *Journal of Retailing*, 86(2), 200-207.

[30] Izmalkov, S. (2004), "Shill Bidding and Optimal Auctions," MIT Working Paper.

[31] Jones, M.T., (2011), "Bidding Fever in eBay Auctions of Amazon.com Gift Certificates," *Economics Letters*, 113, 5-7.

[32] Kamins, M.A. , Dreze, X. , Folkes, V.S. (2004), "Effects of Seller-supplied Prices on Buyers? Product Evaluations: Reference Prices in an Internet Auction Context," *Journal of Consumer Research* 30, 622-628 .

[33] Katkar, R. , Reiley, D.H. (2006), "Public versus Secret Reserve Prices in eBay Auctions: Results from a Pokémon Field Experiment," *Advances in Economic Analysis and Policy* 6 (2), 7.

[34] Kauffman, R. J., and Wood, C. A. (2005), "The Effects of Shilling on Final Bid Prices in Online Auctions," *Electronic Commerce Research and Applications*, 4(1), 21-34.

[35] Kosmopoulou, G. and De Silva, D. G. (2007). "The Effect of Shill Bidding upon Prices: Experimental Evidence," *International Journal of Industrial Organization*, 25(2), 291-313.

[36] Ku, G., Galinsky, A.D., and Murnighan, J.K. (2006), "Starting Low but Ending High: A Reversal of the Anchoring Effect in Auctions," *Journal of Personality and Social Psychology*, 90(6), 975-986.

[37] Ku, G., Malhotra, D., and Murnighan, J.K. (2005), "Towards a Competitive Arousal Model of Decision-Making: A Study of Auction Fever in Live and Internet Auction," *Organizational Behavior and Human Decision Processes*, 96, 89-103.

[38] Lewis, G. (2011), "Asymmetric Information, Adverse Selection and Online Disclosure: The Case of eBay Motors," *American Economic Review*, 101, 1535-1546.

[39] Liu, Y.C. (2017), "Applying Clustering to Analyze Bidding Behaviors and Shill Bidding in Online Auction," Master's Dissertation. National Sun Yat-Sen University, Taiwan.

[40] Lucking-Reiley, D., Bryan, D., Prasad, N. and Reeves, D. (2007), "Pennies from eBay: The Determinants of Price in Online Auctions," *The Journal of Industrial Economics*, 55(2), 223-233.

[41] Mamun, K. (2015), "Combating Shill Bidding in Real Time: Prevention, Detection and Response," *Computer and Information Science*, 8(2), 24-36.

[42] Mamun, K., and Sadaoui, S. (2013), "Combating Shill Bidding in Online Auctions," *International Conference on Information Society, i-Society.*, 170-176.

[43] McCannon, B.C., and Minuci, E. (2020), "Shill bidding and trust," *Journal of Behavioral and Experimental Finance*, 26, 100279.

[44] McLachlan G.J. and Rathnayake S.I. (2014), "On the Number of Components in a Gaussian Mixture Model," *WIREs Data Min. Knowl. Discov.*, 4:341–55.

[45] McGrath, S. and McGrath, L. (2010), "The Complete Idiot's Guide to Ebay," 2nd Edition. Indianapolis, Indiana, USA: Alpha Books.

[46] Newberry, P.W. (2015), "The Effect of Competition on eBay," *International Journal of Industrial Organization*, 40, 107-118.

[47] Nikitkov, A.N. and Bay, D. (2010), "Online Auction Fraud: An Empirical Analysis of Shill-bidding Practice," *Journal of Forensic and Investigative Accounting*, 2(3), 191-228.

[48] Nikitkov, A., and Bay, D. (2015), "Shill Bidding: Empirical Evidence of Its Rffectiveness and Likelihood of Detection in Online Auction Systems," *International Journal of Accounting Information Systems*, 16, 42-54.

[49] Ockenfels, Reiley, D., and Sadrieh, A. (2006), "Online Auctions," In T.J. Hendershott (Ed.), Handbooks in Information Systems I, Handbook on Economics and Information Systems, pp.571-628, Amsterdam: Elsevier Science.

[50] Raviv, Y. (2006), "New Evidence on Price Anomalies in Sequential Auctions: Used Cars in New Jersey," *Journal of Business & Economic Statistics*, 24(3), 301-312.

[51] Reiley, D. (2006),"Field Experiments on the Effects of Reserve Prices in Auctions: More Magic on the Internet," RAND *Journal of Economics* 37(1), 195-211 .

[52] Scott, J.E., Gregg, D.G. and Choi, J.H. (2015), "Lemon Complaints: When Online Auctions Go Sour," *Information Systems Frontier*, 17, 177-191.

[53] Shah, H.S., Joshi, N.R., Sureka, A., and Wurman P.R. (2003), "Mining eBay: Bidding Strategies and Shill Detection," In: Zaïane O.R., Srivastava, J., Spiliopoulou, M., Masand, B. (eds). WEBKDD 2002 – Mining Web Data for Discovering Usage Patterns and Profiles. *WebKDD 2002. Lecture Notes in Computer Science, 2703.* Berlin, Heidelberg, Germany: Springer.

[54] Simonsohn, U., and D. Ariely. (2008), "When Rational Sellers Face Non-Rational Consumers: Evidence from Herding on eBay," *Management Science*, 54(9), 1624-37.

[55] Steiglitz, K. (2007), *Snipers, Shills, and Sharks: eBay and Human Behavior*, Princeton University Press, Princeton.

[56] Trevathan, J., and Read, W. (2009), "Detecting Shill Bidding in Online English Auctions," *Handbook of Research on Social and Organizational Liabilities in Information Security*, 46, 446-472.

[57] Trevathan, J., Read, W., and Goel, R. (2008), "Online Auction Deception Using a Shill Bidding Agent," *Journal of Scientific and Practical Computing*, 2(2), 23-38.

[58] Vincent, Daniel, (1995), "Bidding Off the Wall: Why Reserve Prices May Be Kept Secret?" *Journal of Economic Theory*, 65, 575-558.

[59] Wooldridge, J.M. (1995), "Score Diagnostics for Linear Models Estimated by Two Stage Least Squares," *Advances in Econometrics and Quantitative Economics: Essays in Honor of Professor CR Rao*, 66-87.

[60] Wykoff, F.C. (1973), "A User Cost Approach to New Automobile Purchases" *The Review of Economic Studies*, 40(3), 377-390.

[61] Xu, H, Bates, C.K., and Shatz, S.M. (2009), "Real-Time Model Checking for Shill Detection in Live Online Auctions," *Software Engineering Research and Practice*, 134-140.

Table 1: **Definition and Description of Variables for Identifying IDs and Clustering the Bidders**

Variables	Description
Buyer Characteristic	
$NoBids_k$	The number of bids for bidder i participating in listing j.
$RepB_k$	The bidder's reputation for the kth bid in listing j.
$BidAmout_k$	The amount of the bidder i's kth bid in listing j.
$BidIncrement_k$	The amount difference between bids placed by bidder i and the latest bid before that in listing j.
$InterBidTime_k$	The time difference between bids placed by bidder i and the latest bid before that in listing j.
$DiffFirstBid_k$	The difference between the expiration time of listing j and the time of bidder i's first bid.
$DiffLastBid_k$	The difference between the expiration time of listing j and the time of bidder i's last bid.
Seller Characteristic	
n	The number of listings held by seller $m(j)$.
Shill Indices	
$\alpha_{i,m(j)}$	The percentage of the seller's auctions bidder i has participated in given a particular seller.
$\beta_{i,j}$	The percentage of bids that bidder i has submitted in listing j.
$\gamma_{i,m(j)}$	The proportion of wins that bidder i has participated in given a particular seller.
$\delta_{i,j}$	The normalized average inter-bid times for bidder i participating in listing j.
$\varepsilon_{i,j}$	The normalized average inter-bid increments for bidder i participating in listing j.
$\zeta_{i,j}$	The normalized time differences between the starting time and the time of bidder i's first bid in listing j.
$\eta_{i,j}$	The normalized time differences between the expiration time and the time of bidder i's last bid in listing j.

Table 2: **Summary Statistics for Identifying IDs and Clustering the Bidders**

Panel A: Basic Information		**All**	**RA**	**BINA**
The number of listings		18,441	9,473	8,968
The number of bids		154,599	94,082	60,517
for bidder's ID partially concealed		151,230	91,845	59,385
for bidder's ID fully revealed		3,369	2,237	1,132
The number of sellers		7,653	-	-
Panel B: Buyer and Seller Characteristics				
Bid level (154,599 obs.)	**Mean**	**S.D.**	**Min**	**Max**
$NoBids$	4.44	5.21	1.00	67.00
$RepB$	89.34	480.72	0.00	70014.00
$BidAmount$ (USD)	5842.59	6050.12	0.01	67775.00
$BidIncrement$ (USD)	369.91	744.82	0.00	39300.00
$InterBidTime$ $(second)$	28298.67	52532.39	0.00	851640.00
$DiffFirstBid$ (day)	3.45	2.66	0.00	11.00
$DiffLastBid$ (day)	2.92	2.64	0.00	11.00
Seller level (7,653 obs.)	**Mean**	**S.D.**	**Min**	**Max**
n	2.41	6.39	1.00	281.00
Panel C: For Clustering the Bidders				
Bidder-Seller-Listing level (72,597 obs.)	**Mean**	**S.D.**	**Min**	**Max**
α	0.10	0.14	0.00	1.00
β	0.29	0.27	0.00	1.00
γ	0.74	0.36	0.02	1.00
δ	0.40	0.42	0.00	1.00
ε	0.39	0.38	0.00	1.00
ζ	0.51	0.42	0.00	1.00
η	0.47	0.42	0.00	1.00

Note: RA represents regular auction, and BINA represents buy-it-now auction.

Table 3: **Definition and Description of Variables for Regular Auctions**

Variables	Description
Transaction information	
$Sold_j$	A dummy variable indicating whether vehicle in listing j is sold or not.
$WinBidR_j$	The winning price of listing j divided by listing j's Kelley Blue Book value.
Auction characteristic	
$StBidR_j$	The starting price of listing j divided by listing j's Kelley Blue Book value.
BIN_j	A dummy variable indicating whether the listing j is listed under buy-it-now auction.
SRP_j	A dummy variable indicating whether the listing j has a secret reserve price.
$Posted\ Duration_{d(j)}$	Dummy variables indicating whether the duration of the listing j is 3 days, 5 days, 7 days, or 10 days.
$Shill\ Dummy_j$	A dummy variable indicating whether the listing j is a shilled listing.
$Competitor_j$	The number of vehicles with the same model and age as listing j listed auction within the posted duration of j.
Seller characteristic	
$ln(Seller\ Score_{m(j)})$	The natural log of seller $m(j)$'s total number of positive minus negative feedback ratings for transactions.
$Seller\ is\ Dealer_{m(j)}$	A dummy variable indicating whether seller $m(j)$ is a car dealer.
$ln(Seller's\ Experience_{m(j)})$	The natural log of how many transactions seller $m(j)$ has made within a year.
$\overline{StbidR}_{m(-j)}$	The average starting prices of other auctions that seller $m(j)$ has listed during our sample period.
Car characteristic	
$Warranty_j$	A dummy variable indicating whether the vehicle in listing j has a warranty or not.
$AMileage_j$	The mileage of the vehicle in listing j divided by its age. (In 1,000 mile/year)
$Vehicle\ Condition_j$	Whether the vehicle condition is clear, salvage, or other in listing j.
$Car\ Model_j$	The car model of the vehicle in listing j. There are 20 car models in our sample.
$Car\ Body\ Type_j$	The car model of the vehicle in listing j. There are 10 body types in our sample.
$Fuel\ Type_j$	The fuel type of the vehicle in listing j. There are 7 fuel types in our sample.

Note: We add one to the number of transactions and seller's score before taking the natural log to avoid zero value.

Table 4: Summary Statistics of Auction Variables

Variable	Obs.	Mean	S.D.	Min	Max
Trade Information					
$Sold_j$	10,893	0.171	0.376	0	1
$WinBidR_j$	1,858	0.696	0.281	0.046	2.188
Auction characteristic					
$StBidR_j$	10,893	0.392	0.393	0.00000028	2.778
BIN_j	10,893	0.516	0.500	0	1
SRP_j	10,893	0.765	0.424	0	1
$Posted\ Duration_{d(j)} = 3Days$	10,893	0.041	0.198	0	1
$Posted\ Duration_{d(j)} = 5Days$	10,893	0.136	0.343	0	1
$Posted\ Duration_{d(j)} = 7Days$	10,893	0.668	0.471	0	1
$Posted\ Duration_{d(j)} = 10Days$	10,893	0.155	0.362	0	1
$Shill\ Dummy_j$	10,893	0.222	0.416	0	1
$Competitor_j$	10,893	17.046	15.846	0	109
Seller characteristic					
$ln(Seller\ Score_{m(j)})$	10,893	4.181	1.780	0	9.571
$ln(Seller's\ Experience_{m(j)})$	10,893	3.001	1.457	0	9.033
$\overline{StbidR}_{m(-j)}$	8,101	0.394	0.362	0.00000033	2.778
$Seller\ is\ Dealer_{m(j)}$	10,893	0.171	0.450	0	1
Car characteristic					
$Warranty_j$	10,893	0.362	0.481	0	1
$AMileage_j$	10,893	15.067	15.844	0	999.999

Table 4: **Summary Statistics of Auction Variables (Continued)**

Variable	Obs.	Mean	S.D.	Min	Max
Car characteristic					
Car Model					
4*Runner*	10,893	0.101	0.302	0	1
Avalon	10,893	0.035	0.183	0	1
Camry	10,893	0.150	0.357	0	1
Celica	10,893	0.029	0.168	0	1
Corolla	10,893	0.079	0.270	0	1
FJ Cruiser	10,893	0.026	0.159	0	1
Highlander	10,893	0.052	0.221	0	1
Land Cruiser	10,893	0.029	0.167	0	1
*MR*2	10,893	0.009	0.095	0	1
Matrix	10,893	0.019	0.136	0	1
Prius	10,893	0.068	0.253	0	1
*RAV*4	10,893	0.037	0.188	0	1
Sequoia	10,893	0.047	0.211	0	1
Sienna	10,893	0.058	0.235	0	1
Solara	10,893	0.043	0.204	0	1
Supra	10,893	0.012	0.107	0	1
Tacoma	10,893	0.123	0.329	0	1
Tercel	10,893	0.003	0.056	0	1
Tundra	10,893	0.067	0.250	0	1
Yaris	10,893	0.012	0.111	0	1

Table 4: Summary Statistics of Auction Variables (Continued)

Variable	Obs.	Mean	S.D.	Min	Max
Car characteristic					
Vehicle Condition					
Clear	10,893	0.933	0.250	0	1
Salvage	10,893	0.051	0.221	0	1
Other	10,893	0.015	0.123	0	1
Car Body Type					
Convertible	10,893	0.032	0.176	0	1
Coupe	10,893	0.046	0.210	0	1
Hatchback	10,893	0.068	0.251	0	1
Minivan/Van	10,893	0.055	0.228	0	1
Pickup truck	10,893	0.185	0.388	0	1
SUV	10,893	0.282	0.450	0	1
Sedan	10,893	0.290	0.454	0	1
Wagon	10,893	0.011	0.106	0	1
Other	10,893	0.004	0.064	0	1
Unspecified	10,893	0.026	0.160	0	1
Fuel Type					
CNG	10,893	0.000	0.010	0	1
Diesel	10,893	0.000	0.014	0	1
Electric	10,893	0.000	0.019	0	1
Gasoline	10,893	0.941	0.235	0	1
Hybrid – electric	10,893	0.051	0.221	0	1
Other	10,893	0.006	0.079	0	1
Unspecified	10,893	0.000	0.019	0	1

Table 5: Number of Listings, Bids, Bidders, and Sellers

Panel A: The First Part Sample			**All**
The number of listings			18,841
The number of bids			154,599
The number of bidders			52,685
The number of sellers			7,653
Panel B: The Second Part Sample	**RA**	**BINA**	**All**
The number of listings	5,268	5,625	10,893
The number of bids			69,919
The number of bidders			25,896
The number of sellers			4,433

Table 6: Secret Reserve Price and the Number of Bids

	Listing with No Bids	Listing with at Least One Bid
Listing w/o SRP	1,294	1,268
Listing with SRP	1,463	6,868
Bid greater than SRP	-	676
Bid smaller than SRP	-	6,192
The number of listings	2,757	8,136

Table 7: Regression Result of Heckman Model w/o Shill Dummy Control

	Sold		*WinBidR*
Variables	Coefficient	Marginal Effect	Coefficient
StBidR	-1.696***	-0.280***	0.520***
	(0.0667)	(0.0089)	(0.0216)
Shill Dummy			
Shill Dummy × *StBidR*			
SRP	-1.753***	-0.289***	0.383***
	(0.0466)	(0.0076)	(0.0214)
Competitor	-0.0104***	-0.0017***	0.0040***
	(0.00156)	(0.0002)	(0.0005)
ln(*Seller Score*)	0.00480	0.0008	0.0050
	(0.00969)	(0.0016)	(0.0038)
Seller is Dealer	-0.561***	-0.0925***	
	(0.0357)	(0.0063)	
Warranty	-0.145***	-0.0240***	0.154***
	(0.0421)	(0.0069)	(0.0162)
AMileage	0.00102	0.0002	-0.0012***
	(0.000909)	(0.0002)	(0.0003)
Duration = 3 *Days*	0.0546	0.0090	-0.0905***
	(0.0854)	(0.0141)	(0.0292)
Duration = 5 *Days*	0.172***	0.0284***	-0.0719***
	(0.0648)	(0.0107)	(0.0227)
Duration = 7 *Days*	-0.0192	-0.0032	-0.0175
	(0.0517)	(0.0085)	(0.0193)
VCondition = *Clear*	-0.167**	-0.0275**	0.0864***
	(0.0695)	(0.0115)	(0.0242)
VCondition = *Other*	-0.114	-0.0188	0.0285
	(0.130)	(0.0215)	(0.0400)
Constant	0.726**		0.521***
	(0.367)		(0.0368)
ρ			-0.733***
			(0.030)
σ			0.299***
			(0.009)
Inverse Mills Ratio			-0.219***
			(0.014)
Observations			10,893

Notes: (a) The numbers in parentheses are robust standard errors. *, **, and *** indicate the 10, 5, and 1% levels of significance, respectively. (b) σ is the standard error of the residual in the price equation, and ρ denotes the correlation coefficient between errors in the price and trade equations. (c) We also control the fuel type, car model, and car body type in the trade equation.

Table 8: Regression Result of Heckman Model with Shill Dummy Control

	Sold		*WinBidR*
Variables	Coefficient	Marginal Effect	Coefficient
StBidR	-1.783***	-0.289***	0.566***
	(0.0752)	(0.0099)	(0.0228)
Shill Dummy	-0.0041	-0.0007	0.0766***
	(0.0438)	(0.0071)	(0.0168)
Shill Dummy × *StBidR*	1.466***	0.238***	-0.334***
	(0.169)	(0.0263)	(0.0613)
SRP	-1.809***	-0.293***	0.381***
	(0.489)	(0.0078)	(0.0196)
Competitor	-0.0100***	-0.0016***	0.0038***
	(0.0016)	(0.0003)	(0.0005)
ln(*Seller Score*)	0.0054	0.0009	0.0037
	(0.0097)	(0.0016)	(0.0038)
Seller is Dealer	-0.559***	-0.0905***	
	(0.0362)	(0.0064)	
Warranty	-0.154***	-0.0249***	0.148***
	(0.0418)	(0.0067)	(0.0161)
AMileage	0.0011	0.0002	-0.0012***
	(0.0009)	(0.0001)	(0.0003)
Duration = 3 *Days*	0.0614	0.0100	-0.0905***
	(0.0862)	(0.0295)	(0.0290)
Duration = 5 *Days*	0.182***	0.0295***	-0.0682***
	(0.0656)	(0.0106)	(0.0224)
Duration = 7 *Days*	-0.0240	-0.0039	-0.0198
	(0.0521)	(0.0085)	(0.0190)
VCondition = *Clear*	-0.153**	-0.0247**	0.0785***
	(0.0704)	(0.0114)	(0.0240)
VCondition = *Other*	-0.102	-0.0165	0.0243
	(0.131)	(0.0212)	(0.0397)
Constant	0.859**		0.505***
	(0.373)		(0.0369)
ρ			-0.719***
			(0.032)
σ			0.293***
			(0.008)
Inverse Mills Ratio			-0.211***
			(0.014)
Observations			10,893

Notes: (a) The numbers in parentheses are robust standard errors. *, **, and *** indicate the 10, 5, and 1% levels of significance, respectively. (b) σ is the standard error of the residual in the price equation, and ρ denotes the correlation coefficient between errors in the price and trade equations. (c) We also control the fuel type, car model, and car body type in the trade equation.

Table 9: Regression Result of Mixed-Process Model

	StBidR	*Sold*		*WinBidR*
Variables	Coefficient	Coefficient	Marginal Effect	Coefficient
$StBidR$		-1.829***	-0.248***	0.554***
		(0.0818)†	(0.0100)	(0.0262)†
$Shill\ Dummy$	-0.0823***	-0.0197	-0.0027	0.0726***
	(0.0052)	(0.0451)	(0.0061)	(0.0171)
$Shill\ Dummy \times StBidR$		1.472***	0.200***	-0.334***
		(0.167)	(0.0219)	(0.0616)
SRP	-0.0951***	-1.822***	-0.247***	0.378***
	(0.0078)	(0.0495)	(0.0076)	(0.0201)
$Competitor$	0.0004***	-0.0099***	-0.0013***	0.0039***
	(0.0001)	(0.0016)	(0.0002)	(0.0005)
$ln(Seller\ Score)$	0.0028	0.0040	0.0005	0.0031
	(0.0022)	(0.0097)	(0.0013)	(0.0038)
$Seller\ is\ Dealer$		-0.556***	-0.0754***	
		(0.0363)	(0.0052)	
$Warranty$	0.0470***	-0.144***	-0.0195***	0.150***
	(0.0052)	(0.0426)	(0.0057)	(0.0162)
$AMileage$	-0.0005*	0.0010	0.0001	-0.0012***
	(0.0003)	(0.0009)	(0.0001)	(0.0003)
$Duration = 3\,Days$	-0.0048	0.0631	0.0086	-0.0903***
	(0.0129)	(0.0860)	(0.0116)	(0.0291)
$Duration = 5\,Days$	0.0041	0.185***	0.0251***	-0.0662***
	(0.0080)	(0.0656)	(0.0089)	(0.0224)
$Duration = 7\,Days$	0.0073	-0.0213	-0.0029	-0.0197
	(0.0055)	(0.0522)	(0.0071)	(0.0190)
$VCondition = Clear$	0.0028	-0.151**	-0.0204**	0.0797***
	(0.0120)	(0.0703)	(0.0096)	(0.0240)
$VCondition = Other$	0.0225	0.101**	-0.0136	0.0241
	(0.0247)	(0.131)	(0.0177)	(0.0395)
$Constant$	0.141***	0.900**		0.513***
	(0.0182)	(0.372)		(0.0376)

Notes: (a) The numbers in parentheses are robust standard errors. *, **, and *** indicate 10%, 5%, and 1% levels of significance, respectively. (b) † denotes 1% level of significance for the test of exogeneity (F=205.33, $p-value$=0.000 for the trade equation; F=11.38, $p-value$=0.000 for the transaction price equation. (c) σ's are the sample standard deviations of the errors in the mixed-process model; and ρ's denote the correlation coefficients between errors in the structure estimation. (d) The subscripts "1", "2", and "3", correspond to the price equation, trade probability, and starting price, respectively. (e) We also control the fuel type, car model, and car body type in the trade equation.

Table 9: **Regression Result of Mixed-Process Model (Continued)**

	$StBidR$	$Sold$		$WinBidR$
Variables	Coefficient	Coefficient	Marginal Effect	Coefficient
$ln(Seller's\ Experience)$	-0.0118***			
	(0.0030)			
$\overline{StbidR}_{m(-j)}$	0.836***			
	(0.0119)			
BIN	0.0283***			
	(0.0051)			
ρ_{12}				-0.719***
				(0.032)
ρ_{13}				0.050
				(0.042)
ρ_{23}				0.044
				(0.028)
σ_1				0.293***
				(0.008)
σ_3				0.194***
				(0.004)
Observations				10,892

Notes: (a) The numbers in parentheses are robust standard errors. *, **, and *** indicate 10%, 5%, and 1% levels of significance, respectively. (b) † denotes 1% level of significance for the test of exogeneity (F=205.33, $p - value$=0.000 for the trade equation; F=11.38, $p - value$=0.000 for the transaction price equation. (c) σ's are the sample standard deviations of the errors in the mixed-process model; and ρ's denote the correlation coefficients between errors in the structure estimation. (d) The subscripts "1", "2", and "3", correspond to the price equation, trade probability, and starting price, respectively. (e) We also control the fuel type, car model, and car body type in the trade equation.

Table 10: Proportions of Shill Bidders before and after Threshold

Shill Bidder %	Criteria: 5%	Criteria: 30 minutes	Criteria: 5 minutes
Before threshold	8.80%	8.42%	8.37%
	(5,176)	(5,593)	(5,741)
After threshold	5.29%	5.33%	5.01%
	(872)	(429)	(297)

Note: The number in parentheses is the number of bidders.

Figure 1: Example of a Shill Bidder

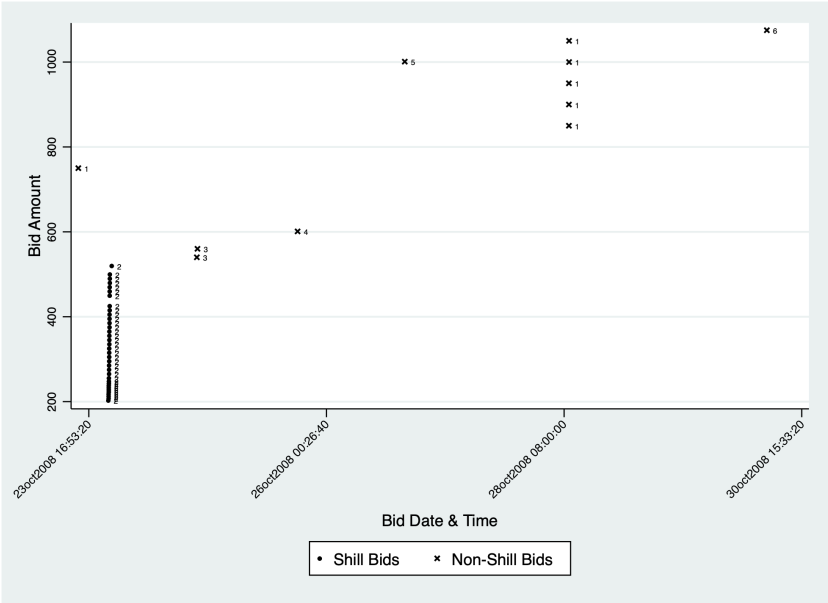


Figure 2: Density Distribution of PC1

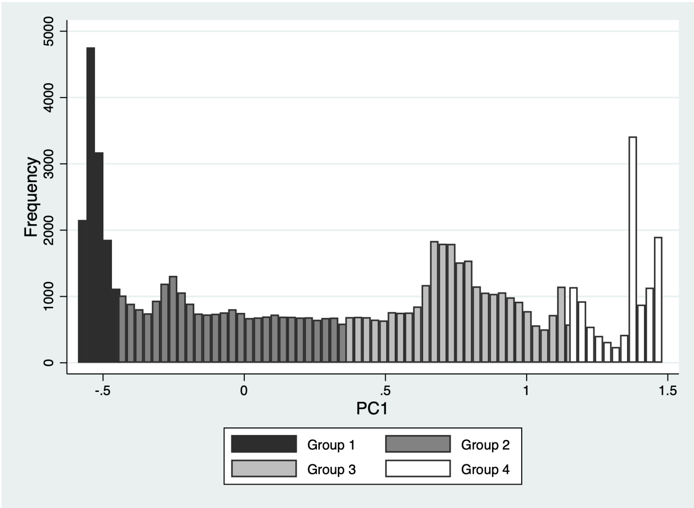


Figure 3: Density Distribution of PC2

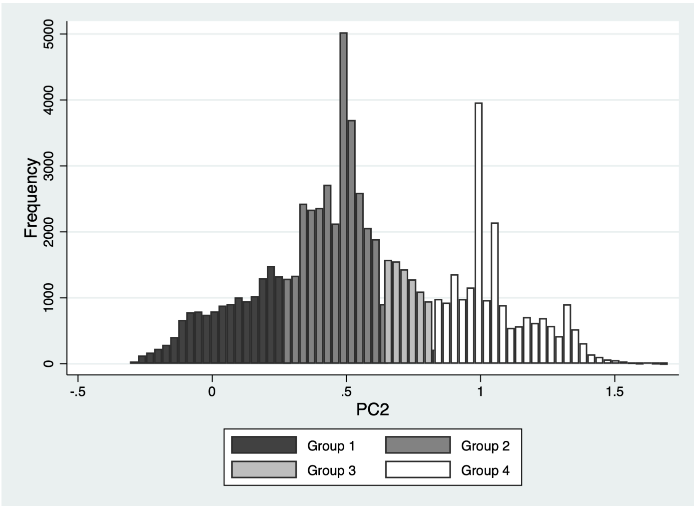
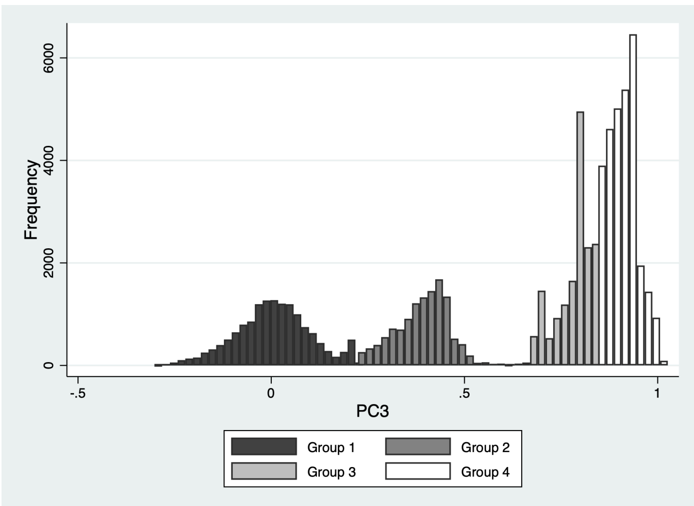


Figure 4: Density Distribution of PC3



Appendix: For reference only

Appendix A: Details of Shill Indexes

In this paper, we adopted seven characteristics of shill bidding behavior to determine the likelihood of a bidder i being a shill of listing j. The shill indexes are proposed by Trevathan and Read (2009) and extended by Liu (2017). The definitions and calculations of these shill characteristics are described as follows.

A.1 α rating

$\alpha_{i,m(j)}$ is the percentage of listings held by a particular seller m and participated in by bidder i.

$$\alpha_{i,m(j)} = \frac{n^i_{m(j)} - w^i_{m(j)}}{n_{m(j)}}$$

where $0 \leq \alpha_{i,m(j)} \leq 1$. $n_{m(j)}$ denotes the total number of listings held by seller m; $n^i_{m(j)}$ denotes the number of listings held by seller m and participated in by bidder i; $w^i_{m(j)}$ denotes the number of listings held by seller m and won by bidder i. A shill is likely to participate in auctions held by the same seller. Therefore, a large $\alpha_{i,m(j)}$ means that a lot of listings held by seller m are participated in by bidder i, and bidder i is likely to be a shill of listing j.

A.2 β rating

$\beta_{i,j}$ is the relative number of bidder i's bids among all bids in listing j.

$$\beta_{i,j} = \frac{b^i_j}{b_j/2}$$

where $0 \leq \beta_{i,j} \leq 1$. b_j denotes the total number of bids in listing j, b^i_j denotes the total number of bids placed by bidder i in listing j, and $b_j/2$ denotes the maximum possible number of bids placed by any bidder in a given auction that the bidder should not win. If a bidder places more bids than this number, he will definitely win the auction. A shill wants to lose, so he should not bid more than this number. Therefore, a large $\beta_{i,j}$ means that bidder i bids aggressively in listing j, and he is likely to be a shill of listing j.

A.3 γ rating

$\gamma_{i,m(j)}$ is the proportion of bidder i winning listings held by seller m.

$$\gamma_{i,m(j)} = \frac{5 \times (w^i_m + 0.2)}{n^i_m}$$

where $0 \le \gamma_{i,m(j)} \le 1$. The purpose of shill bidding is to push up the transaction price, not to win auctions, so shills should be less likely to win too many auction. Of course, it is possible to calculate the raw winning probabilities $\frac{w_m^i}{n_m^i - w_m^i}$. However, this would be unfair to the vast number of bidders who bid once or twice in their lives. Such bidders tend to place bids in only one listing and do not win it. This will result in a winning probability of 0%; however, this doesn't mean that these bidders are shills, but rather that they seldom buy things through auctions. In order to fix this problem, Trevathan and Read (2009) set $\gamma_{i,m}$ as above to permit a 20% win-to-lose ratio without penalty. A small $\gamma_{i,m}$ means that the bidder i does not want to win a lot of listings, and then he is likely to be a shill of listing j. If the resulting $\gamma > 1$, we adjust the value to 1.

A.4 δ rating

$\delta_{i,j}$ is the standardized average response time difference between bids placed by bidder i and the latest bid before that in listing j. Shills want to give more time to the honest bidders to consider, so they should respond to a new bid as soon as possible. In order to derive $\delta_{i,j}$, first, the response time of each bid placed by bidder i in listing j is calculated by:

$$\Delta t_{j,k}^i = \begin{cases} 0, & k = 1 \\ t_{j,k}^i - t_{j,k-1}^{-i}, & k > 1, k \in K_j \end{cases}$$

where k is the kth bid in listing j, and K_j is the set of bids in listing j. $t_{j,k}^i$ is the time when the kth bid is placed in listing j (which is placed by bidder i), and $t_{j,k-1}^{-i}$ is the time when the $(k-1)$th bid is placed in listing j (which is placed by someone other than i). The average response time of bidder i in listing j is derived through:

$$\bar{\Delta} t_j^i = \frac{1}{b_j^i} \sum_{k \in K_j^i} \Delta t_{j,k}^i$$

where K_j^i is the set of bids placed by bidder i in listing j. Next, the average response time for each bidder i is rescaled into the range of 0 to 1 for each listing j. It will be compared to the maximum and minimum average response time across all the bidders for each listing:

$$\bar{\Delta} t_j^{max} = \max_{i \in B_j} \bar{\Delta} t_j^i$$

$$\bar{\Delta} t_j^{min} = \min_{i \in B_j} \bar{\Delta} t_j^i$$

$$\delta_{i,j} = \frac{\bar{\Delta} t_j^i - \bar{\Delta} t_j^{min}}{\bar{\Delta} t_j^{max} - \bar{\Delta} t_j^{min}}$$

where $0 \le \delta_{i,j} \le 1$, and B_j is the set of bidders participating in listing j. The smaller $\delta_{i,j}$, the more quickly bidder i responds to new bids in listing j, and he is more likely to be a shill of listing j.

A.5 ε rating

$\varepsilon_{i,j}$ is the standardized average bid increment difference between bids placed by bidder i and the latest bid before that in listing j. Placing a large bid in an auction increases the chance to win. Shills do not want to win, so they tend to place bids as small as possible. The calculation of $\varepsilon_{i,j}$ is similar to that of $\delta_{i,j}$. First, the difference between each bid amount and its last bid amount is calculated through:

$$\Delta p^i_{j,k} = \begin{cases} 0, & k = 1 \\ p^i_{j,k} - p^{-i}_{j,k-1}, & k > 1, k \in K_j \end{cases}$$

where $p^i_{j,k}$ is the bid amount of the the kth bid in listing j (which is placed by bidder i), and $p^{-i}_{j,k-1}$ is the bid amount of the $(k-1)$th bid in listing j (which is placed by someone other than bidder i). Then, for each bidder i, the average difference is calculated within each listing j.

$$\bar{\Delta} p^i_j = \frac{1}{b^i_j} \sum_{k \in K^i_j} \Delta p^i_{j,k}$$

The average differences are rescaled into the range of 0 and 1, compared to the maximum and minimum of all the bidders in each listing:

$$\bar{\Delta} p^{max}_j = \max_{i \in B_j} \bar{\Delta} p^i_j$$
$$\bar{\Delta} p^{min}_j = \min_{i \in B_j} \bar{\Delta} p^i_j$$
$$\varepsilon_{i,j} = \frac{\bar{\Delta} p^i_j - \bar{\Delta} p^{min}_j}{\bar{\Delta} p^{max}_j - \bar{\Delta} p^{min}_j}$$

where $0 \le \varepsilon_{i,j} \le 1$. A small $\varepsilon_{i,j}$ means bidder i tends to place small bids in listing j, and he is more likely to be a shill of listing j.

A.6 ζ rating

$\zeta_{i,j}$ is the average time difference between the first bid placed by bidder i and the starting time in listing j. Since a shill does not want to win, he will not want to place bids too late in an auction. $\zeta_{i,j}$ measures how early the bidder joins in an auction, while $\eta_{i,j}$ measures how early the bidder leaves an auction. Therefore, the difference between the starting time of listing j and the time of bidder i's first bid must be calculated:

$$\Delta t^i_{first} = t^i_{j,first} - t_{j,first}$$

where $t_{j,start}$ is the starting time of the listing j, and $t^i_{j,first}$ is the first bid time of the bidder i in listing j. The difference for each bidder i is then standardized into the range of 0 to 1 according to the maximum and minimum differences of all bidders for each listing j:

$$\Delta t_{j,first}^{max} = \max_{i\in B_j} \Delta t_{j,first}^{i}$$
$$\Delta t_{j,first}^{min} = \min_{i\in B_j} \Delta t_{j,first}^{i}$$
$$\zeta_{i,j} = \frac{\Delta t_{j,frist}^{i} - \Delta t_{j,frist}^{min}}{\Delta t_{j,first}^{max} - \Delta t_{j,first}^{min}}$$

where $0 \leq \zeta_{i,j} \leq 1$. The smaller $\zeta_{i,j}$ is, the earlier bidder i joins in listing j, and the more likely that bidder i is acting as a shill.

A.7 η rating

$\eta_{i,j}$ is the average time difference between the last bid placed by bidder i and the ending time in listing j.

$$\Delta t_{last}^{i} = t_{j,end} - t_{j,last}^{i}$$

where $t_{j,last}^{i}$ is the last bid time of the bidder i in listing j. After deriving the difference between the expiration time and the time of bidder i's last bid in each listing, the differences are then rescaled into the range of 0 to 1 based on the maximum and minimum of the differences of all bidders in each listing j:

$$\Delta t_{j,last}^{max} = \max_{i\in B_j} \Delta t_{j,last}^{i}$$
$$\Delta t_{j,last}^{min} = \min_{i\in B_j} \Delta t_{j,last}^{i}$$
$$\eta_{i,j} = \frac{\Delta t_{j,last}^{i} - \Delta t_{j,last}^{min}}{\Delta t_{j,last}^{max} - \Delta t_{j,last}^{min}}$$

where $0 \leq \eta_{i,j} \leq 1$. The larger $\eta_{i,j}$, the earlier bidder i leaves from listing j, and the more likely that bidder i is acting as a shill.

Appendix B: Note on Gaussian Mixture Model

The following procedure produces replications of clustering with a different number of clusters to avoid the uncertainty from the random states. Specifically, we consider two criteria including the Bayesian Information Criterion (BIC) and Integrated Classification Likelihood (ICL). The former is commonly used in assessing the number of mixture components. Since one non-Gaussian cluster can be represented as two or more Gaussian components, BIC tends to choose more Gaussian components. The later criterion proposed by Biernacki, Celeux, and Govaert (2000) instead considers the Integrated Completed Likelihood criterion equivalent to BIC penalized by the entropy of the corresponding clustering. This criterion leads to the choice of a parsimonious model. The following formula can estimate the Integrated Completed Likelihood:

$$ICL = BIC + EN(\hat{\tau}) \tag{1}$$

where $EN(\hat{\tau})$ denotes the entropy:

$$EN(\hat{\tau}) = -\sum_{i=1}^{N}\sum_{j=1}^{g} \hat{\tau_{ij}} log \hat{\tau_{ij}} \tag{2}$$

We refer readers to McLachlan and Rathnayake (2014) for a comprehensive review of how to select the number of components in a Gaussian mixture model. Note that since the clustering result is sensitive to the initial values of parameters, for a specific number of components, we implement a Gaussian mixture with different random states (RS) 100 times and plot the box plot of BIC/ICL and their difference when allowing for one more component. We also compare the difference of BICs/ICLs to monitor the change of BICs/ICLs, and these results provide another view on where we should stop adding more components.

Figure: Box Plots of BIC and ICL