"An Empirical Analysis of Electronic Auction Data: Estimating Marginal Effects and Recovering the Cumulative Joint Distribution Function of Valuations"

Kevin O'Connor

Economics of Auctions Research Paper

Introduction:

If one was to imagine a fundamental theme to economics, it would likely be how to allocate resources in the presence of scarcity. Logically, a line of questions follows this supposition, such as "How do we maximize efficient allocation?", or "are there market mechanisms that improve efficient allocation?" The First Theorem of Welfare Economics (Debreu, 1959), implies that the uninhibited exchange between buyers and sellers in the market will result in efficient allocation, conditioned on the absence of a market failure. This theorem hinges on the idea that there must be a sufficient number of market participants to avoid the exercise of market power, which can reduce efficiency in allocation. A hallmark of today's evolving economic landscape is the consolidation of sellers, and the atomization of specialty digital products and services that have ambiguous valuations.

An important example of this is the sale of ads on search engines, a market in which Google has the leading role. In this market, agents pay to have their website linked to the search results of specific words or phrases. For example, if one was to search "Office Supplies" on Google, the first several results would have links to office supply vendors such as W.B. Mason. With no conceivable limit on the various combinations of words in phrases across different languages one could search, this has led to prioritization of the most common words or phrases that relate to potential interest in a business's products or services. The most valuable phrases are those that are either most common or the most specific. This results in considerable information asymmetry between Google and ad buyers. How possibly can Google efficiently set prices on billions of words and phrases from hundreds of languages across the globe? The traditional competitive market model doesn't seem to apply in such settings.

Enter the auction. It has been demonstrated over 50 years ago that a second-price auction is an efficient mechanism for the sale of a single indivisible item with ambiguous value (Vickery, 1961). This auction model is becoming increasingly important in today's economy for the reasons stated above. Primarily that traditional competitive markets will not create the efficiency needed for markets of multitude atomized digital products and services.

In this paper, we will demonstrate how second price auctions result in an efficient outcome that also lends to revenue for the seller competitive with what the competitive market produces. We will also demonstrate empirically how different factors increase revenue to the seller, such as courting additional bidders. To accomplish these aims, we will need to analyze broad data from a product that is sold both in the open competitive market, and in second price auctions.

Furthermore, auctions result in other benefits to the seller that competitive markets do not always produce. In the process of bidding, bidders reveal quite a bit of information about themselves and how they value an item. We will demonstrate how from a second-price auction, a seller can recover distributions of valuations for the item being sold. This information can be of critical importance to price discovery, reserve price setting, and general valuation exercises.

Using a custom dataset from 97 auctions for the Bowflex Selecttech 552 dumbbell set, we conduct an empirical analysis to determine the marginal effects of courting additional bidders to an auction. We demonstrate that, in line with orthodox auction theory, that the more bidders there are in an auction, the higher the revenue generated for the seller. Further, we demonstrate that this market results in efficient allocation. From the 97 auctions, we can also recover the cumulative distribution function and probability density function of valuations.

A Brief Literature Review:

It is generally accepted that William Vickery initiated the empirical study of auctions and auction mechanisms. In 1961, Vickery published his theory of the independent private values model and Bayesian Nash Equilibrium. Results of his study resulted in key concepts in auction theory such as The Revenue Equivalence Theorum, and the theory that at certain objectives such as efficiency can be obtained by several auction structures. (1) From this point, the scope of auction theory began to expand. Some important studies that followed include Roger Myserson's work on optimal auction design (Myserson, 1981), James Dana and Kathryn Spiers "Designing a Private Industry" (Dana and Spiers, 1994), that began to further study how to design and implement auctions to achieve specific outcomes, such as allocative efficiency. In recent decades, much of the emphasis in auctions research has been placed on recovering information about bidders from auction data, and the applications and design of electronic auctions. The key area of interest in this paper is in regard to online second price auctions. In 2017, it was demonstrated that eBay auctions using proxy bidding are almost numerically equivalent to traditional sealed bid second price auctions by demonstrating the equilibrium bidding strategy is to bid one's own valuation (Hickman, Hubbard and Paarsch, 2017). As the use of SPA online becomes more ubiquitous, there is a fair amount current research being published. For example, Albert Xin Jiang and Kevin Leyton-Brown have laid out game-theoretical, decision-theoretical, simple (such as ignoring hidden bids), and expectation maximization learning approaches to estimating valuation distributions (Xin Jiang and Leyton-Brown, 2021). For this paper, we will lean on the collective wisdom of auction theory generated over the past 50 years, particularly the previous two papers mentioned, for both lending credence to some assumptions made, as well as describing methodology for retrieving valuation distributions in online SPA.

The Data and Key Assumptions:

For this study, we have chosen to analyze eBay "sold" listings data on the Bowflex Selecttech 552 dumbbell set from two time periods, 2010-2012, and 2021. The data from the earlier period is of 71 auctions, and has been manually entered from "Worthpoint," which displays the sale price, date sold, and condition, along with the original eBay listing. From the listing, it is evident whether shipping costs were included in the price . Shipping prices were not included in all listings, so over an average of 26 listings which included shipping prices, I was able to create a baseline shipping cost across the earlier time period. From this average shipping cost, and whether shipping was included in the price, the total price paid for the winner of each auction was calculated. Three dummy variables for each listing were also created. The first of which was a dummy variable "blkfriday" which has a value of "1" if an auction was completed within 15 days of November 26th. This controlled for any open market sales inducing pressure on the auction market. The second dummy variable "covid" which took a value of "0" for all of the 2010-2012 data, and controlled for the exogenous utility/demand shock effects of the COVID-19 pandemic. During the 2020 U.S. pandemic lockdowns, the prices of this dumbbell set skyrocketed above $1,000 on Amazon marketplace during March 2020 as most U.S. gyms were shut down. Thereby, the consumer's utility from home fitness equipment increased, and demand increased. The third dummy, "new" controlled for the condition of the item, with the value of "1" if the item was new in box, "0" if otherwise.

The 2021 data consists of 26 completed auction listings sourced from eBay, which include highest bid, whether shipping was included, the date sold, number of bids, and the condition of the item as well as the text included with the listing. Average shipping costs were tabulated and the total price calculated for each listing. Ideally, we would have information on how many bidders participated in each auction. However, we can still provide some interpretation. If there was one bid, there was only one bidder in the auction. If there were two bids, there were two bidders in the auction. If there were greater than two bids, we can infer that there were likely greater than two bidders, but cannot say definitively. We can still infer from this however by creating a top-coded censored model using total bid as an instrument for total number of bidders. All of the same dummy variables apply to this data, however, the “covid” dummy variable takes a value of “1” for the 2021 data to control for any residual effects of the shocks on utility and demand that persist from 2020. Lastly, for comparison, eBay data for 30 “Buy it Now” listings for the same product were collected, which are used analogously for competitive market outcomes.

This particular item was evaluated for the following reasons. First, bidders should not care about what other bidders value the item at, so bidders should exhibit private values. Home fitness equipment was chosen because unlike technology items such as computers and video game consoles, the utility afforded by the home fitness equipment is consistent over time, whereas a laptop for sale 5 years ago is worth considerably less today. As opposed to a luxury good which could be speculated upon or exhibit Giffen behavior, the dumbbells are an ordinary and typical good in a consumer’s consumption bucket. This further implies rationality, consistency, and some symmetry in the valuations of the consumer, who are all on aggregate, typical consumers in the United States. From here, we can reasonably assume symmetric independent private values (SIPV). This dumbbell set is one of the most popular weight sets on the internet, which allows for greater ease of data collection. Secondarily, as the set is so ubiquitous, we should expect a low degree of risk aversion. For this study, we will assume risk neutral bidders. Lastly, apart from Black Friday sales, the prices from the manufacturer have remained relatively consistent over the decade. Thereby, this creates some stability in secondary markets which further lend to rationally explainable behavior.

For example, the retail price for this product in 2021 is $400 plus shipping. The Black Friday deal for 2021 was $300 plus free shipping. The savvy online secondary retailer would buy multiple sets at this reduced price, and sell sets later on in the year near the prevailing retail price. The consumer, seeking to pay less than the retail price plus shipping, will participate in auctions that allow them to purchase the set below or up to the prevailing retail price. Thereby, we will expect that with all agents behaving rationally, the average highest bid in auction will approach, but not cross the retail price plus shipping. The average “Buy It Now” price will be very close to, but marginally below, the retail price plus shipping. Lastly, a used dumbbell set should theoretically provide nearly the same utility to the consumer. Rationally, there should exist a small premium for the assurance that comes from a new set being free from damage. Last, we assume perfect information. On aggregate, we should expect rational behavior.

Efficient by design: Proof of efficiency in eBay proxy bid model

eBay utilizes a fixed-time semi-sealed-bid modified second-price format with proxy bidding. For analogy, let's assume bidder A places a bid for an item at $100. The highest bid is displayed as $100. Bidder B, places a bid of $200. The highest bid now is displayed as $\$100 + \varepsilon$, where ε is some marginal increment calculated by eBay that is scaled in proportion to the current highest hidden bid, where $\$100 + \varepsilon \leq \$200$. Bidder A is unaware of Bidder B's maximum bid, and chooses to bid $\$110 + \varepsilon$.

Assuming $\$110 + \varepsilon \leq \$200$, eBay will proxy bid on bidder B's behalf increasing the displayed maximum bid to $\$110 + 2\varepsilon$ assuming that $\$110 + 2\varepsilon \leq \$200$. This continues until the fixed time expires, and the winner pays the second highest price plus some marginal increment that is $\leq$ to the winner's highest bid. In the event of a tie, the agent that placed the bid first wins. The winner of this format will always be the highest bidder, conditional on the fact that they place their bid first in the event of a tie. However, the highest bidder winning is not a sufficient condition for efficiency. The highest bidder must also have the highest valuation. So, we will formally prove that the outcome is efficient, through determination of the Bayesian Nash Equilibrium bidding strategy.

$As\ the\ winning\ conditions\ depends\ upon\ the\ actions\ of\ the\ two\ highest\ bidders, we\ assume\ two$

$representative\ agents\ A\ and\ B\ who\ are\ the\ two\ highest\ bidders\ and\ have\ the\ two\ highest\ valuations.$

$From\ bidder\ A's\ perspective, there\ are\ three\ possible\ outcomes\ when\ bidding\ above\ thier\ valuation:$

$$(i)\ b_B > b_A > v_A$$

$$(ii)\ b_A > b_B > v_A$$

$$(iii)\ b_A > v_A > b_B$$

$In\ outcome\ i\ and\ ii, bidder\ A\ will\ either\ lose, or\ win\ above\ thier\ valuation.$

$Outcome\ iii\ looks\ promising, however, outcomes\ are\ uncertain. Bidder\ A\ would\ be\ better\ off$

$bidding\ v_A\ in\ case\ b_B\ turns\ out\ being\ higher\ than\ v_A. The\ only\ winning\ solution\ where$

$there\ is\ no\ chance\ to\ pay\ over\ value\ is\ to\ bid\ v_A. This\ is\ symmetric, as\ bidder\ B\ faces\ the\ exact$

$same\ circumstances. Similarly, we\ will\ compare\ bidding\ below\ valuation.$

$$(i)\ b_B > v_A > b_A$$

$$(ii)\ v_A > b_A > b_B$$

$$(iii)\ v_A > b_B > b_A$$

$Similarly\ to\ bidding\ above\ valuation, there\ is\ only\ one\ favorable\ outcome\ (ii)\ where\ bidder\ A\ wins.$

$However, the\ odds\ of\ winning\ would\ still\ be\ greater\ if\ A\ bids\ thier\ valuation, while\ maintaining$

$the\ same\ payoff.\ Bidder\ B\ faces\ identical\ circumstances.$

$$\therefore \beta(v_i) = v_i\ is\ the\ symmetric\ Nash\ equilibrium\ for\ this\ game.$$

$$As\ there\ is\ no\ profitable\ deviation\ from\ this\ strategy, by\ the\ definition\ of\ Bayesian\ Nash\ Equilibrium,$$

$$If v_A > v_B, \qquad \beta(v_A) = v_A > \beta(v_B) = v_B\ \forall\ v_i \in\ \mathbb{R}_n^+$$

$$\therefore The\ individual\ with\ the\ highest\ valuation\ will\ always\ win. This\ format\ is\ efficient\ by\ design.$$

Comparison to competitive market revenues:

As we have demonstrated, the outcome from this auction is always efficient, however, efficiency is not the primary motive for many economic agents. In particular, a seller seeks to increase their revenue. If the efficient mechanism does not produce at least equitable revenue to the open market, why would a seller choose it? For description of the open market outcome, we take the mean total price of 30 sold “Buy it now” listings on eBay. The mean total price is $406.33 for new in box listings. When observing the auction data for new in box listings, we see the average selling price is $418.13. To see if these means are significantly different, we will run a T-test on the null hypothesis that the means are equal to one another.

$$H_o: (Mean(TotalPrice) if\ buyitnow) - (Mean(TotalPrice) if\ auction, if\ new = 1) = 0$$

Two-sample t test with equal variances

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
0	69	418.1322	9.279473	77.08109	399.6153	436.6491
1	30	406.3333	11.14433	61.04003	383.5406	429.1261
Combined	99	414.5568	7.285609	72.4909	400.0987	429.0148
diff		11.79884	15.88957		-19.73756	43.33524

```
    diff = mean(0) - mean(1)                                      t =   0.7426
H0: diff = 0                                     Degrees of freedom =       97

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.7702         Pr(|T| > |t|) = 0.4595          Pr(T > t) = 0.2298
```

With a t-value of .7426, we fail to reject the null hypothesis. We can arrive at the conclusion that the auction results in no statistically significant difference in seller revenue than they would on the open competitive market. This is an interesting finding that is consistent with rational agents. In one meta-analysis of cross-item eBay auctions, in 50% of auctions, bidders end up paying more than the “Buy it now” prices (2). Perhaps this is because this item has a consistent and stable price from the manufacturer, and there is little ambiguity around what the item is worth. This result is consistent with the perfect information assumption.

Estimating the Marginal Return of Courting Additional Bidders:

The traditional wisdom of auction theory predicts that all else equal, having more bidders tends to lead to increased seller revenue from increased competition. This, in fact, is a benchmark result of the revenue equivalence theorem that only holds under symmetric independent private values. As we have assumed SIPV, we should expect in our data that there is a definable marginal increase in total price induced by increasing the number of bidders in the auction.

There are a few considerable limitations in the data we have collected. First of which, is that we only have data on the number of bids, and not the number of bidders. To overcome this difficulty, we will have to make some assumptions. If there is only a single bid, then there is only a single bidder.

If there are two bids, there are almost certainly two bidders. Any more than two bids results in a loss of inference. Two bidders could potentially enter into a bidding war, bidding any number of times before the auction ends. What we can assume, is that a higher number of bids is correlated with a higher number of bidders. By utilizing the greater variation in total bids, perhaps we will be able to infer the marginal effects of bidder increase. To do this, we will run two regressions with total bids as an instrument for the number of bidders. We will run a two stage least squares IV regression, and compare it with general method of moments IV regression, while in both models controlling for black Friday sales and item condition.

```
. ivregress 2sls TotalPrice (Bidders=Bids) New BlkFriday

Instrumental variables 2SLS regression          Number of obs   =         25
                                                Wald chi2(3)    =       9.68
                                                Prob > chi2     =     0.0215
                                                R-squared       =     0.2572
                                                Root MSE        =     59.154

------------------------------------------------------------------------------
  TotalPrice | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     Bidders |   -11.2344   27.17971    -0.41   0.679    -64.50565    42.03686
         New |   56.09921   24.29672     2.31   0.021     8.478511    103.7199
   BlkFriday |   69.70071   28.29929     2.46   0.014     14.23511    125.1663
       _cons |   371.6215    67.3538     5.52   0.000     239.6104    503.6325
------------------------------------------------------------------------------
Instrumented: Bidders
 Instruments: New BlkFriday Bids

. ivregress gmm TotalPrice (Bidders=Bids) New BlkFriday

Instrumental variables GMM regression           Number of obs   =         25
                                                Wald chi2(3)    =      11.99
                                                Prob > chi2     =     0.0074
                                                R-squared       =     0.2572
GMM weight matrix: Robust                       Root MSE        =     59.154

------------------------------------------------------------------------------
             |               Robust
  TotalPrice | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     Bidders |   -11.2344    24.5483    -0.46   0.647    -59.34819     36.8794
         New |   56.09921   24.23847     2.31   0.021     8.592691    103.6057
   BlkFriday |   69.70071   21.68396     3.21   0.001     27.20093    112.2005
       _cons |   371.6215   70.20144     5.29   0.000     234.0292    509.2137
------------------------------------------------------------------------------
Instrumented: Bidders
 Instruments: New BlkFriday Bids
```

Noting in the above output tables, even while using an instrumental variable, we do not arrive at a statistically significant marginal effect of the number of bidders on the selling price. Although this result is discouraging, if we loosen our assumptions further, we may be able to arrive at a reliable estimate.

To overcome the lack of statistical significance, we will make a single assumption. If each bidder is following the Bayesian Nash Equilibrium strategy, they will simply bid their true valuation. What follows from this assumption is that each bidder will only bid once, and any additional bid is from a new bidder. To see if this is a reasonable assumption, we will run a tobit regression of total bids on bidders, with a lower limit of 1, and an upper limit of 3- assuming that any more than 5 bids constitute 3 or more bidders. The results of this tobit regression back up our assumption (Appendix 1). The coefficient on total bids is 1, with a P-value of .000. We will try to run a similar regression with the total bids as proxy for the number of bidders.

. regr TotalPrice Bids BlkFriday New

Source	SS	df	MS			
				Number of obs	=	25
				F(3, 21)	=	2.83
Model	33875.4386	3	11291.8129	Prob > F	=	0.0634
Residual	83891.0214	21	3994.81054	R-squared	=	0.2876
				Adj R-squared	=	0.1859
Total	117766.46	24	4906.93583	Root MSE	=	63.205

TotalPrice	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
Bids	-.4327761	1.118725	-0.39	0.703	-2.759291	1.893739
BlkFriday	64.91414	28.83879	2.25	0.035	4.940588	124.8877
New	56.92848	26.0776	2.18	0.041	2.697139	111.1598
_cons	349.2018	22.73443	15.36	0.000	301.9229	396.4806

Even with these assumptions made, we do not arrive at a statistically significant estimate for the marginal effect. We conduct an F-test to verify this (Appendix 2). This leaves two likely possibilities; either our sample size is too small, or increasing the number of bidders does not have an impact in this case.

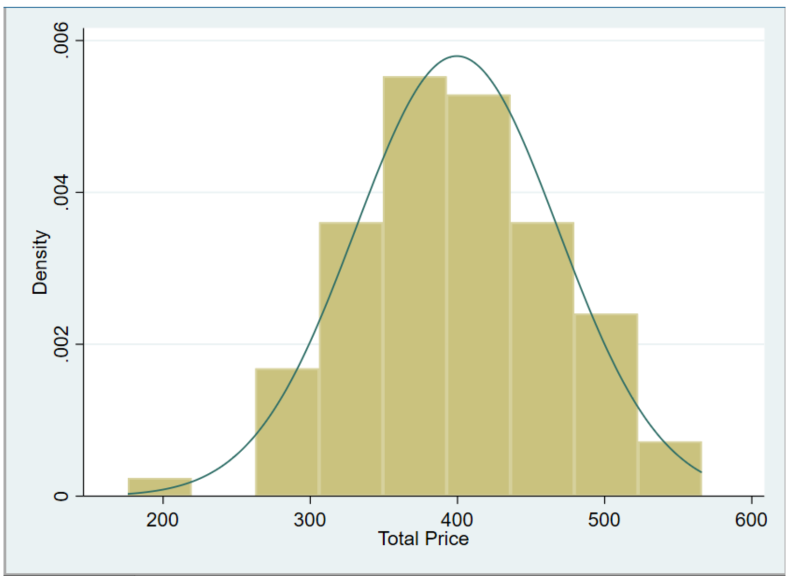
In 2013, economist Brian Larue found evidence of a negative invitation effect in multi-unit sequential auctions of complete information for livestock in Quebec. During periods of generally lower hog prices, in effort to increase competition, the auctioneers invited bidders from Ontario to participate. Using maximum likelihood estimators, Larue determined that the presence of Ontario bidders depressed the prices in the early sequences of the auction. (3) Among other reasons, such as the tendency to over-estimate how common symmetry among bidders exist, Larue points out in the event of sequential auctions, bidding strategies may shift. For example, if a consumer has the ability to participate in a less competitive auction for the same item, it would make sense for them to wait for later in the sequence. In fact, as there are perpetually more auctions for this same item on eBay it is almost as if there is an infinite sequence of auctions for the dumbbell set in question. Perhaps there is a change in the bidding strategy that results. Perhaps the data set has too few observations. Perhaps our assumptions of SIPV are incorrect. Interestingly, we have a statistically significant positive relationship between sales made within 15 days of Black Friday in every model ran. From this glaringly counterfactual result, it can be concluded that it is most likely that the sample size is too small to properly infer marginal effects.

Estimating the Distributions of Valuations:

It can be very beneficial to sellers and bidders to understand the distribution of valuations. Whether it is used to refine setting the optimal reserve price, make forecasts of future auction revenues, or set a bid strategically, understanding valuations is of importance.

In the case of a second price auction, when we store the selling price, we are actually storing the value of the second highest valuation, as the Bayesian Nash Equilibrium is to bid one's valuation. When estimating distributions, it can be helpful to make parametric assumptions. For example, we know that in large sample of data, many randomly distributed values take on a normal distribution.

In fact, if we plot all the total sale prices and impose the normal distribution, it seems like the distribution of second highest values fits well.



To test what seems intuitively possible, we conduct a Shapiro-Wilk normality test (Shapiro and Wilk, 1965) For both the "new in box" listings, and the aggregated listings, we fail to reject the null hypothesis that the sale prices are not distributed normally. From here, we have a closed form solution for the joint probability distribution function of second-highest valuations for all 97 auctions.

$$f^{2:n}(v) = \frac{1}{(\sigma = 9.279)\sqrt{2\pi}} \exp\left\{-\frac{(v - 418.13)^2}{172.2}\right\}$$

$$F^{2:n}(v) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{v - 418.13}{9.279\sqrt{2}}\right)\right] = F(v)^n + nF(v)^{n-1}(1 - F(v))$$

$$Define\ \phi(H; n)\ as\ the\ solution\ to\ H = \phi^n + n\phi^{n-1}(1 - \phi)$$

$$\phi = \left(\frac{H}{(1 + (n - 1))}\right)^{\frac{1}{n}}$$

$$F(v) = \left(\frac{\frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{v - 418.13}{9.279\sqrt{2}}\right)\right]}{\big(1 + (n-1)\big)} \right)^{\frac{1}{n}}$$

In conclusion:

In summary, from a custom manually created dataset of 97 eBay auctions for the Bowflex Selecttech dumbbell set, we have successfully demonstrated that the auction market efficiently allocates resources in a manner that keeps seller profit equitable with what can be expected from the open market. We were able to recover the joint cumulative distribution function of valuations from the sale total price information. Unfortunately, we were unable to demonstrate the marginal benefit of courting new bidders in our data, due to the limitations of the data set. However, it may also be possible that the sequential nature of eBay auctions has some warping effect of bidding strategies. Future studies would require a larger data set by utilizing data from a longer time period in order to have a more complete and robust analysis of marginal effects. It may also be interesting to include the 2020 COVID-19 data, and conduct analysis.

Bibliography:

(1) Milgrom, P. (1996, December 9). *Prize Lecture on William Vickery*. Nobelprize.Org. Retrieved November 12, 2021, from https://www.nobelprize.org/uploads/2018/06/vickrey_lecture_milgrom.pdf

(2) Han Lee, Y., & Malmendier, U. The Bidder's Curse. *National Bureau of Economic Research*.

(3) Larue, B., Jeddy, M., & Pouliot, S. (in press). On the Number of Bidders and Auction Performance: When More Means Less. *Reseau SPAA Network*.

Helpful Papers:

(1) Hickman, B, & Hubbard, T, Identification and estimation of a bidding model in electronic auctions. *ECONSTOR*

(2) Xin Jiang, A, & Leyton-Brown, K, Estimating Bidder's Valuation Distributions in Online Auctions. *University of British Columbia*

Appendix:

(Appendix 1)

```
Tobit regression                                Number of obs     =         25
                                                       Uncensored =          8
Limits: Lower = 0                                   Left-censored =          0
        Upper = 3                                  Right-censored =         17

                                                LR chi2(-2)       =    -555.18
                                                Prob > chi2       =          .
Log likelihood = 251.40237                      Pseudo R2         =    10.5994
```

Bidders	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
Bids	1	1.32e-15	7.6e+14	0.000	1	1
_cons	1.60e-14	.	.	.	.	.
var(e.Bidders)	2.72e-29	1.32e-29			1.01e-29	7.35e-29

(Appendix 2)

```
. test _b[Bids]=0

 ( 1)  Bids = 0

       F(  1,    21) =    0.15
            Prob > F =    0.7028
```

(Appendix 3)

```
. swilk(TotalPrice) if New==1

                   Shapiro-Wilk W test for normal data
```

Variable	Obs	W	V	z	Prob>z
TotalPrice	68	0.97069	1.762	1.230	0.10937

```
. swilk(TotalPrice)

                   Shapiro-Wilk W test for normal data
```

Variable	Obs	W	V	z	Prob>z
TotalPrice	96	0.98813	0.947	-0.119	0.54754