Comparing Open Outcry and Online Auctions: Evidence from North Dakota Mineral Auctions*

Eric Karsten†

May 2020

Perspectives in Computational Social Science
Dr. Richard Evans

Abstract

I study a switch from in person open outcry English auctions to online eBay-style ascending English auctions for mineral rights in a private values framework with unobserved heterogeneity. Using a natural experiment which alternated auction format for 5 years, I find that leases sold in online auctions yield higher revenues than similar parcels sold in in person auctions. What explains this revenue difference? I estimate a structural model of bidder values and then use it to perform counterfactual policy experiments which show that the revenue premium of online auctions can be explained by higher participation relative to the participation of in person auctions.

*This paper is based on (but not identical to) my MA thesis.

†I am grateful to Thomas Covert for advising the thesis on which this paper builds. I would like to thank Ali Hortaçsu, Victor Lima, Yixin Sun, and the participants of the MAPSS micro seminar for their helpful commentary. Finally, I would like to thank Drew Combes, Susie Entzel, and Kristie McCusker at North Dakota Trust Lands as well as Gus Rivero at EnergyNet for their assistance in gathering and understanding the data used in this paper.

1 Theoretical Model

I will first set out some preliminary assumptions about the setting which hold across all auction formats. I use notation similar to that of Hernández, Quint, and Turansick (2020) due to the similarity of my model to theirs. For simplicity of notation, I will refer to their paper by HQT from here on in.

Assumption 1 (Private Values and Unobserved Heterogeneity)**.** *The value of tract t is different across each bidder i and is multiplicatively decomposed as follows:*

$$\ln(V_{it}) = \psi(Z_t) + \theta_t + \varepsilon_{it}.$$

It is comprised of:

- *A deterministic part, $\psi(Z_t)$, where Z_t is a vector of observable characteristics of the parcel being auctioned and ψ is a function, not necessarily linear in the components of Z_t, which maps observable characteristics into a valuation.*
- *A component of heterogeneity, θ_t, observed by the bidders, but not by the econometrician. This is an iid draw of a random variable with cumulative density function F_θ, and probability density function f_θ.*
- *A private value, ε_{it}, observed by only bidder i. This is an iid draw from a distribution with cumulative density F_ε and probability density f_ε.*

Assumption 2. *The densities f_θ and f_ε are continuous and real analytic.*

This assumption is required for identification using the results from HQT. Their identification proof relies on this assumption because it allows them to prove properties of these densities by proving those properties on their power series expansions. The probability density of any common family of continuous distributions (Normal, Poisson, Weibull, etc.) satisfies this assumption, so it is a less restrictive assumption than the parametric or semi-parametric form that is used to approximate the true distribution. I maintain two further assumptions about bidding behavior from Haile and Tamer (2003).

Assumption 3. *Bidders do not bid more than they are willing to pay.*

Assumption 4. *Bidders do not allow an opponent to win at a price they are willing to beat.*

In both the online and in person ascending English auctions, the private values assumption means that a bidder will not change their valuation after observing other bids in the auction. It is a well-known result that bidders should drop out of the auction at their value $\exp\left(\psi(Z_t) + \theta_t + \varepsilon_{it}\right)$. Thus, the transaction price will be $\exp\left(\psi(Z_t) + \theta_t + \varepsilon_{it}^{(2)}\right)$ where $\varepsilon_i^{(2)}$ is the second order statistic of the draws from the private value distribution. Observing transaction price alone is not sufficient to identify the distributions of the unobserved heterogeneity (f_θ) and the private valuations (f_ε).

The first identification issue is that in online auctions the observed number of bidders is not the same as the number of bidders who "drew" from the value distribution and considered bidding on the parcel. This motivates the need for an entry model. The model essentially has bidders arrive in sequence having drawn a value. They then place a bid informed by their value so long as the current high bid doesn't exceed their value. I describe this model in detail in the structural estimation section, it is the same one used in HQT. In order to identify the distribution of how many bidders will arrive at each auction, I need that distribution to have positive probability on a bounded support, observations of the number of bidders entering each auction, and an entry model which translates actual number of bidders into a distribution of how many might be observed placing bids. HQT Theorem 3 non-parametrically identifies the distribution of bidders considering each auction from this data.

The second identification issue is that I need to separately identify f_θ and f_ε. The HQT identification strategy here relies on the fact that more bidders will shift the order statistic which is drawn from private values f_ε without shifting the draw from the unobserved heterogeneity f_θ. In order to do this, HQT requires a participation shifter. For this, I follow Compiani, Haile, and Sant'Anna (2020) and use number of adjacent leases as a participation shifter. This participation shifter means that different auctions will draw from different distributions of the number of bidders, allowing the separate identification. HQT Theorem 4 provides the necessary non-parametric identification conditional on the entry model, based on observation of the transaction price and participation shifter.

2 Data

Estimation requires three kinds of data: auction results, price data, and market research data suggestive of tract quality. Auction results come from the State of North Dakota's land trust website. [1] The sequence of bids in the online auctions comes from the platform EnergyNet, which is responsible for conducting the online auctions. [2] The State of North Dakota provided auction notices showing which parcels were nominated for each auction and which firm nominated them. Table 1 provides summary statistics of some of the relevant characteristics of the auctions to provide the reader with the necessary context to imagine the magnitudes of the estimates later in the paper. One relevant note regarding this table is that it appears that open outcry auctions bring in more revenue, this isn't a particularly meaningful comparison though because it isn't comparing the differences in revenue between auctions for similar parcels with similar market oil prices.

Variable	Open Outcry	Online
Revenue ($/Acre)	736.89 (2202.32)	190.38 (1208.66)
Log Revenue	3.61 (2.66)	2.76 (1.99)
Number of Bidders		2.98 (1.69)
Tract Size (Acres)	84.74 (47.06)	89.46 (49.05)
Adjacent Leases	1.82 (2.3)	0.86 (1.84)

Table 1: Summary Statistics: Mean (SD) of several variables in the dataset.

Daily spot prices for the Clearbrook pricing center (used to price the Bakken shale formation) come from S&P Platts. These data cover May 2012 through December 2019. As we see from figure **??**, there is extensive variation in the auction prices over the period being studied.

Additional descriptive data covering 1951 to 2020 comes from DrillingInfo, which provides the lease location, the production decisions that the lessee made, and the quantity of oil they extracted on a monthly basis. I use these data to construct a picture of what a particular firm might expect a new tract to produce based on results from similar nearby tracts that were reporting production at the time of bidding. In the appendix, figure **??** shows variation in space and time of mean oil extracted and figure **??** shows the same thing for gas.

3 Structural Estimation

The reduced form section prompts questions about why there is such a large difference in the competitiveness of the online and open outcry auctions. In order to shed light on some

1. `https://www.land.nd.gov/Minerals/AuctionHistory`
2. `https://www.energynet.com/page/Government_Sales_Results_Previous`

of these questions, I will estimate a structural model of the distributions of bidder valuations from the online auctions. This will allow me to run counterfactual policy experiments where I simulate what would have happened in the open outcry auctions had they been held on the online platform under different conditions for entry. The goal of this is to tease out the magnitude of reduction in bidder participation which might replicate the revenue difference between the online auctions and the open outcry ones. A structural model has the advantage that I am comparing apples to apples by simulating what would have happened in an open outcry auction had it been held online instead. The reduced form alternative is a comparison to a similar but not identical parcel from an online auction.

The apples-to-apples nature of this comparison relies on some modeling assumptions. I am assuming that valuations for the online auctions and the open outcry auctions are drawn from the same distribution. Given the temporal overlap of these auctions and the ex-ante possibility that a parcel could be nominated for either auction time, this appears to be a reasonable assumption. I am also assuming that bidders in online auctions do not engage in any collusive or uncompetitive behavior which would mask my ability to correctly recover value distributions. Finally, I am assuming the form of an entry model to back out how number of bidders drawing from the value distribution corresponds to the number of bidders observed bidding in the auction. Recall that some bidders might consider a parcel, but never bid because by the time they arrive at the auction, they see that the current high bid exceeds their valuation. I will use the same entry model as was used to describe eBay auctions in HQT. This model features the property that not all bidders who draw valuations will be observed placing bids. The specifics of this model are described in step 2 of the following estimation process. My entry model is reduced form in the way that parcel quality influences entry probability, but it makes specific parametric assumptions about the distribution of bidders considering the auction (Poisson) and about the process by which bidders increment their bid throughout the auction.

3.1 Estimation Process

I will estimate the densities f_θ and f_ε from the online auction data using the method from HQT described in the identification section. There are three steps to the estimation procedure. The first step is to run the regression of the second highest bid in each auction (which must be equal to the second highest bidder's valuation) on the observable predictors of auction value

$$\ln\left(B_{it}^{(2)}\right) = \beta Z_t + T_{it}.$$

We define in this equation T_{it} to be the transaction residual whose distribution we wish to decompose into a θ component of unobserved heterogeneity and an ε component of private value.

The second step in the estimation process is to estimate the entry model. We define our entry process as occurring in two parts. First bidders arrive at the auction according to some Poisson process with parameter λ. As they arrive in sequence at the auction, they draw a private value from f_ε. If their private value is higher than the current high bid, they place some bid randomly on the interval between the high bid and their value. If the private value is lower than the current high bid, that bidder doesn't enter the auction. At the end of the auction, bidders bid up to their value. This entry model means that the firms observed placing bids are only a subset of those which drew values from f_ε and considered bidding in the auction. Thus it is necessary to estimate $P\left(N_{obs} = k \mid N = n\right)$. I do this by simulating a million entries processes for each n from 1 to 100.

Having worked out how many observed bidders to expect conditional on the number of bidders who drew a value for the auction, we need to estimate the parameterization of the Poisson distribution. One identifying assumption is that there is some participation shifter, an instrument that makes some auctions have more participants without changing the bid residuals of the auctions. I will follow Compiani, Haile, and Sant'Anna (2020) and use number of firms with preexisting adjacent leases as a participation shifter. I will call this shifter X which is a random variable taking one of the three values in $\{0, 1, 2+\}$. We then estimate the Poisson parameter λ_x for each of these values of the shifter using MLE on the following

$$\arg\max_{\lambda_x} \sum_{m=1}^{M_x} \ln\left(\sum_{n=2}^{\infty} \frac{1}{1 - e^{-\lambda_x} - \lambda_x e^{-\lambda_x}} \frac{\lambda_x^n e^{-\lambda_x}}{n!} P\left(N_{obs} = n_m \mid N = n\right)\right)$$

where we have truncated realizations of 0 or 1 bidders out of the Poisson distribution. Figure 1 has kernel density estimates which visually suggest that the truncated Poisson distribution fits our data quite well.

The third and final step in the estimation process is to recover the distributions of values. Following HQT Theorem 5, we know that the pdf of the transaction price residual conditional a Poisson entry model is

$$f_T(t) = \frac{\lambda^2 e^{-\lambda}}{1 - e^{-\lambda} - \lambda e^{-\lambda}} \int_{-\infty}^{\infty} f_\theta(t-s)(1 - F_\varepsilon(s)) f_\varepsilon(s) e^{\lambda F_\varepsilon(s)} ds$$

the final step is to define a quasi-parametric approximation to each f_θ and f_ε and then we perform MLE on the above likelihood of each observed transaction price using the fitted λ based on the participation shifter. The quasi-parametric approximations come from Gallant and Nychka (1987) and are defined as:

$$f_Z(z) = \left(\sum_{k=0}^{K} \beta_{Z,k} H_k \left(\frac{z - \mu_Z}{\sigma_Z} \right) \right)^2 \varphi \left(\frac{z - \mu_Z}{\sigma_Z} \right)$$

where $H_0(x) = 1$, $H_1(x) = x$, $H_k(x) = \frac{1}{\sqrt{k}} \left(x H_{k-1}(x) - \sqrt{k-1} H_{k-2}(x) \right)$ are the probabalist's Hermite polynomials. For this to be a valid density, we need the condition that $\sum_{k=1}^{K} \beta_{Z,k}^2 = 1$. Additionally, we cannot separately identify two means, so we normalize the mean of the distribution of unobserved heterogeneity to zero. Thus the space over which we are performing MLE is $(\mu_\varepsilon, \sigma_\varepsilon, \sigma_\theta, \vec{\beta}_\varepsilon, \vec{\beta}_\theta)$ to semi-parametrically identify the two distributions of interest.

3.2 Estimation Results

The first step is to control for observable tract heterogeneity. I run two versions of two different specifications. The first includes the controls from reduced form model 1 (year * township * range fixed effects). The second includes the controls from reduced form model 5 (log price, quantity, parcel characteristics and provisions, royalty rate). Additionally, for each of these models, I do a version where I control for the number of bidders and a version where I do not in the first stage estimation.

In the second step, I calibrate the entry model for each value of our participation shifter. I begin by testing whether our instrument for shifting participation is relevant and meets the exclusion restriction. From table 2, it is clear in the first regression that our participation shifter does indeed shift participation. This is even more evident and quite compellingly shown in figure 1. In table 2, we also test whether adding controls for number of bidders eliminates the reduced form effect that our participation shifter has on the log bonus residual. It appears that the effect is greatly diminished in magnitude, but it doesn't go away entirely. This could be attributed to the fact that 2+ adjacent firms is collapsing a few categories into

	Dependent variable:				
	Number of Bidders	Log Bonus			
	(1)	(2)	(3)	(4)	(5)
1 Adj. Firm	0.245*	0.097	0.045	0.403***	0.149
	(0.147)	(0.071)	(0.055)	(0.148)	(0.098)
2+ Adj. Firms	1.535***	0.153*	0.064	1.231***	0.344***
	(0.139)	(0.083)	(0.065)	(0.148)	(0.103)
Year * Location FE		Yes	Yes		
Spline Price + Quantity				Yes	Yes
Number of Bidders FE			Yes		Yes
Observations	815	815	815	794	794
R^2	0.132	0.907	0.946	0.270	0.692
Adjusted R^2	0.130	0.884	0.932	0.256	0.683
Note:			*p<0.1; **p<0.05; ***p<0.01		

Table 2: Regressions testing whether number of firms with adjacent leases affects log bonus other than by increasing the number of bidders.

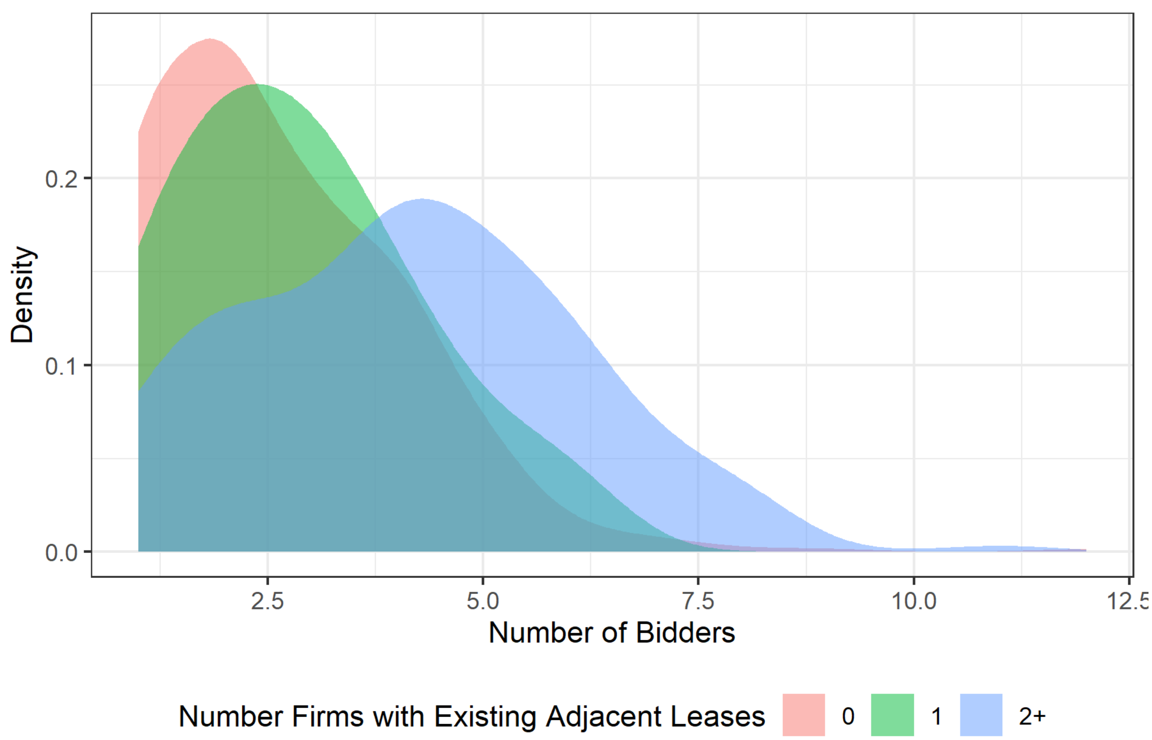

Figure 1: Distributions of bidder entry by number of firms with adjacent preexisting leases to the lease up for auction.

one and more granular categories would make this significance disappear. Due to limited numbers of observations for the higher categories, I am unable to use a more granular instrument. Thus adjacent lessees appears to be a strong, but not entirely exogenous instrument for bidder entry.

	0 Adj. Leases	1 Adj. Leases	2+ Adj. Leases
λ	4.66 (0.2)	5.01 (0.35)	9.23 (0.52)
Observations	362	122	159
-2 Log Likelihood	1136.08	381.05	619.66

Table 3: Entry Model Estimates

Having provided some justification for the instrument, I use MLE to estimate λ for each value of the participation shifter. We see in table 3 that each estimate of λ for each value of the instrument is distinct with fairly tight standard errors.

	Mod1 w/ N	Mod5 w/ N	Mod1 no N	Mod5 no N
μ_ε	-1.75 (0.13)	-3.36 (0.27)	-3.51 (1.35)	-3.57 (0.55)
σ_ε	2.84 (0.18)	5.48 (0.35)	6.01 (2.28)	6.04 (0.86)
σ_θ	1.71 (0.11)	3.29 (0.22)	3.97 (1.51)	3.86 (0.56)
Observations	643	623	643	623
-2 Log Likelihood	1086.45	1064.26	1385.32	1297.98

Table 4: Value Model Estimates

Now we take the estimates from the entry model and fit a normal model of each f_θ and f_ε. While the identification result allows for a more general semi-parametric form, a normal-normal model captures the salient residual decomposition of interest. The results from this estimation are presented in table 4. Impressively across the three different models we see consistency in the relative magnitudes of the estimates. Not a lot should be read into the estimates of the means, the fact that they are negative makes sense because the winning bids generally come from the right tail of the distribution. From the standard errors we see that consistently the variance on ε, the individual values, is generally larger than the variance on the unobserved heterogeneity θ.

3.3 Counterfactual Comparisons

Recall that the structural model allows me to test how reductions in bidder participation could affect the auction results in the open outcry auctions. Quantifying this participation reduction could suggest competing stories about what accounts for the revenue difference.

The first competing story is that it is costly to participate in an in person auction because it involves hiring a local representative or transporting an employee to the site of the auction. Online auctions do not have these entry costs. The second story is that in person auctions might have colluding parties. A collusive cartel would only have to outbid those not in the cartel and thus auction results would have the appearance of fewer auction participants. While I do not have the data to distinguish between these two explanations, I am able to provide bounds on these explanations. I can show the reduction in average number of bidders which explains the difference in auction revenue. This can be interpreted as either the number of participants who didn't attend because of higher costs, or it can be interpreted as the size of the members of the collusive bidding ring considering the typical auction.

The policy experiment I perform is scaling the entry parameters λ by some scale factor. This simulates reduced participation across all auction groups by participation shifter. It retains the feature that "better" parcels according to the exogenous participation shifter will still attract more bids. I then draw a number of bidders N for each open outcry auction and simulate draws from f_θ and f_ε. This allows me to compute the winning bid, number of bidders considering the parcel, number of bidders bidding on the parcel, surplus of winning bidder, and proportion of auctions with only one bidder observed. As λ shrinks, fewer bidders bid in the typical auction and therefore surplus is higher and the average winning bid is lower.

For counterfactual policy experiments, I use Model 5 without controls for N. The four models are quite consistent (as is seen in table 4), and model 1 is based on section * township * year fixed effects which haven't been estimated for all of the tracts in the open outcry auctions. Model 5 without N doesn't introduce bias into the value calculations that would come from the fixed effects for N.

Results from these simulations are reported in table 5. The actual row on the table represents the actual open outcry auctions. Each of the simulation rows should be interpreted as the counterfactual outcome should the parcel in that auction have been auctioned in an online format instead. These simulations do not include any parcels that were auctioned in online auctions, but they do use the value distributions that were recovered using the online auction data.

Notice in the policy simulation that the experiment which replicates the actual average log bid (scaling entry parameters by between 0.3 and 0.4) also dramatically increases the proportion of auctions with only a single bidder participating. This parameter's estimate

Experiment	Avg Log Price	Avg N	Avg N Obs	Avg Surplus	Prop w/ 1 Bid
Actual	2.06 (0.1)				0.43 (0.01)
Sim. 0.25	1.36 (0.07)	2.11 (0.04)	1.72 (0.03)	3.28 (0.1)	0.5 (0.01)
Sim. 0.3	1.63 (0.06)	2.38 (0.03)	1.88 (0.02)	3.46 (0.08)	0.43 (0.01)
Sim. 0.4	2.19 (0.08)	2.97 (0.04)	2.18 (0.02)	3.6 (0.11)	0.32 (0.01)
Sim. 0.5	2.71 (0.08)	3.58 (0.05)	2.46 (0.03)	3.66 (0.1)	0.24 (0.01)
Sim. 0.75	3.84 (0.11)	5.21 (0.07)	3.1 (0.04)	3.49 (0.08)	0.13 (0.01)
Sim. 1	4.76 (0.11)	6.85 (0.09)	3.63 (0.04)	3.25 (0.06)	0.07 (0.01)

Table 5: Policy experiments where we scale entry parameters of poisson process (by the factor listed in the table) to simulate higher entry costs in the open outcry auctions during the overlap period (2016 and later). The standard errors are boostrapped from multiple iterations of the simulation based on different random samples with replacement of the open outcry leases going into the simulation.

is within a margin of error of the same parameter estimated from our data and not used as a moment in calibrating the simulation. This is verification that the structural model is a good description of what is going on in the auction. in particular, this outside moment lends a lot of credence to the reduced bidder interest story due to higher costs of entry in the open outcry auctions. While we cannot observe the full distribution of number of bidders in the open outcry auctions, the fact that we are able to match one of the moments of that distribution is highly suggestive that reductions in bidder participation account for the revenue difference.

One interesting outcome of this policy simulation is how large of a decrease in number of bidders is necessary to replicate the actual revenue residuals from the open outcry auctions. We see that the decrease necessary is by about 4 participants considering bidding in the typical auction. It suggests that in open outcry auctions only two to three bidders are interested in most tracts. This could be a result of the process of bidding in open outcry auctions. Large firms hire land-men to represent them and to bid on their behalf. These representatives have to be sent out with instructions for which parcels to bid on meaning that most parcels will not garner a lot of interest even if they are valuable. Alternatively, the difference could be the result of some collusion. A collection of firms colluding would look similar to there being many fewer bidders because essentially the firms in the bidding cartel who have agreed to take the parcel only need to outbid firms not in the cartel. While there is no direct evidence of collusion and I do not intend to make that accusation in this paper, the collusion story is not inconsistent with the data we observe.

Another parameter of interest that our experiment uncovers is the increase in average surplus associated with fewer bidders participating in each auction. The average surplus increases as I reduce bidder interest and then decreases when bidder interest gets very small. This is because as interest gets quite small, there begins to be a higher number of auctions where nobody bids. On the other hand there is the competing effect that as there are fewer bidders, but still multiple bidders with some interest in the parcel, the typical difference between the highest bid and the second highest bid will be larger. We see that the experiment which does a good job of replicating the actual average log bid has a surplus about .2 log points higher than the simulation as if the auctions had been held online. This means that the typical winning bidder in an open outcry auction had about a 20% higher surplus than a bidder in an online auction. From table 5, I compute the change in average surplus as $\exp(3.6) - \exp(3.25) \approx \$10$ per acre. The typical lease has 80 acres and the typical participant wins about 3 parcels. Thus we back out an entry cost barrier difference of \$2,400. This is a plausible estimate of the cost of entering an in-person auction relative to an online one, providing yet another simulated moment which suggests the strength of my structural model.

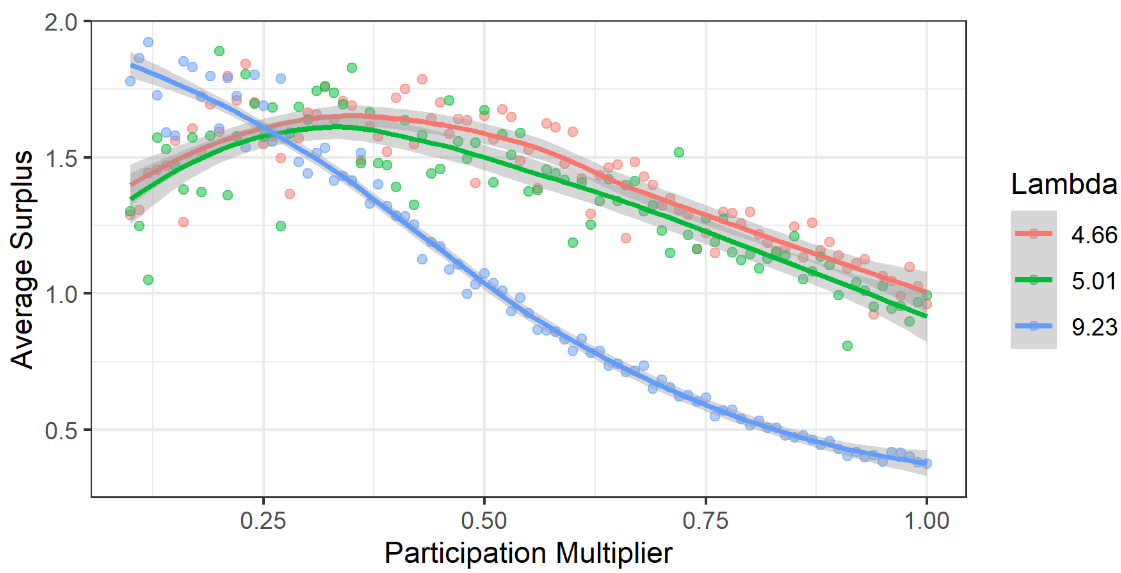

Figure 2: Average Bidder surplus for each lease type as a function of changing auction interest through costs associated with entry.

An additional policy experiment we might wish to perform is to look at the average bidder surplus as a function of bidder participation to see what level of our policy experiment participation multiplier might maximize bidder surplus. This experiment is summarized in figure 2. We see unsurprisingly that for the high quality parcels for which we calibrated a Poisson entry parameter of 9.23, average surplus is maximized when participation is very low, this is because for these parcels if there is only one bidder interested, they will almost certainly have a high enough value to bid on it, so surplus only decreases as more bidders consider the parcel and begin bidding up its value. On the other hand, for the more "typical" parcels, surplus is optimized at a shifter of about .4. This actually corresponds to what our prior experiment recovered as being equivalent to the in person auction, suggesting that in person auctions were essentially optimal in that they optimized surplus for participating bidders but were naturally bad for the state of North Dakota as has been show with prior revenue results in this paper.

References

Compiani, Giovanni, Philip Haile, and Marcelo Sant'Anna. 2020. "Common Values, Unobserved Heterogeneity, and Endogenous Entry in US Offshore Oil Lease Auctions." *Journal of Political Economy* Forthcoming.

Gallant, A Ronald, and Douglas W Nychka. 1987. "Semi-nonparametric Maximum Likelihood Estimation." *Econometrica* 55 (2): 363–390.

Haile, Philip, and Elie Tamer. 2003. "Inference with an Incomplete Model of English Auctions." *Journal of Political Economy* 111 (1): 1–51.

Hernández, Cristián, Daniel Quint, and Christopher Turansick. 2020. "Estimation in English Auctions with Unobserved Heterogeneity." *RAND Journal of Economics* Forthcoming.