Selective Entry in Auctions: Estimation and Evidence

Yunmi Kong[†]

February 7, 2018

Abstract

This paper performs a structural analysis of selective entry in an empirical auction market. The analysis allows for bidders' risk aversion and asymmetry. After documenting empirical evidence of the model, I exploit an entry cost shifter and auction format variation (first-price and English) to establish identification and develop a nonparametric estimation procedure recovering model primitives. I use counterfactual simulations to show how bidders' entry-induced uncertainty about the number of other entrants, combined with risk aversion, can substantially soften the revenue impact of low competition in first-price auctions. This constitutes an important reason for sellers to favor first-price auctions over English auctions.

I would like to thank the New Mexico State Land Office for generous assistance in collecting and interpreting the data. Special thanks are due to current and former chief geologists Stephen Wust, Joe Mraz, and Dan Fuqua for answering an endless stream of questions from me. Chris Barnhill, Kevin Hammit, Tracey Noriega, Philip White, and Lindsey Woods also provided valuable insight on the bidders' perspective. I would like to thank Xiaohong Chen, Jeremy Fox, Emmanuel Guerre, Kei Kawai, Vijay Krishna, Laurent Lamy, John Lazarev, Alessandro Lizzeri, Xun Tang, and participants of the 2015 SITE Summer Workshop, 11th World Congress of the Econometric Society, Stern Friday seminar, NYU IO reading group and Econometrics lunch seminar, and Midwest Econometrics Group for helpful comments and suggestions. I am deeply indebted to Isabelle Perrigne and Quang Vuong for their invaluable guidance and support. All remaining errors are mine. Financial support from NYU-CRATE is gratefully acknowledged.

[†]Department of Economics, Rice University. E-mail: yunmi.kong.01@gmail.com

1 Introduction

Entry and selection have long held a central place in the study of markets. Entry recognizes that the set of entrants in a market is determined by agents' endogenous choice. Selection recognizes that, when choices are endogenous, the subset who choose "yes" (or "no") are not representative of the entire population.

This paper performs a structural analysis of selective entry in an empirical auction market. Selective entry in auctions means bidders for each auction item are endogenously determined and nonrepresentative of those who choose not to bid. The analysis employs a general model allowing for bidders' risk aversion and asymmetry in addition to selective entry. Taking advantage of an entry cost shifter and auction format variation in the data, I establish identification and develop a nonparametric estimation procedure that recovers all the model primitives: bidders' value distributions, utility functions, and entry costs. Using the estimated structural model, I quantify the impact of policies aimed at increasing bidder competition.[1] I also show how bidders' uncertainty about the number of other entrants, caused by the entry process, can alleviate the negative effect of low competition on revenue, particularly in conjunction with bidders' risk aversion. These affect bidding strategies in first-price auctions but not in English auctions, constituting a reason for sellers to favor first-price auctions in settings where competition is low.

The data I use in this analysis come from auctions for the New Mexico State Trust Lands. The New Mexico State Land Office (NMSLO) uses both the first-price sealed-bid (S) format and the English or ascending oral (O) auction format to sell its oil and gas leases. Curiously, in the sealed-bid auctions, leases receiving one bid sell for roughly eight times the reserve price.[2] This tells us a number of things about the underlying auction. First, it must be that bidders face uncertainty about the number of other bidders that will bid; if a bidder knew for certain he was the only bidder, he would have bid the reserve price. This uncertainty implies an entry process whose aggregate outcome is unknown to bidders at the time of bidding.

[1] Low competition is one of the greatest concerns in auctions. See Bulow and Klemperer (1996).

[2] 17% of all sealed-bid auctions in the sample receive just one bid.

Second, there is reason to suspect an important role for risk aversion. Auction theory says expected revenue is unaffected by entry-uncertainty if bidders are risk-neutral; it is when bidders are risk-averse that uncertainty about the number of bidders increases revenue in first-price auctions.[3]

In describing the entry process, bidders say they first look through the published list of auction items to select the ones they like enough to bid on, and then analyze only those items in detail to decide how much to bid. The described process is selective - the bidder's preliminary signal about the item must be high enough to warrant bidding - and values are finalized or learned only after incurring some effort (an entry cost).

Motivated by the empirical evidence, I analyze an auction model with selective entry, where bidders' utility functions and value distributions are generalized to allow risk aversion and asymmetry among bidders. I model both the first-price sealed-bid auction and the English auction, under the independent private values (IPV) paradigm.[4] The selective entry process follows the two-stage model laid out in Marmer, Shneyerov, and Xu (2013) and Gentry and Li (2014). In stage 1, each potential bidder observes a private signal that is correlated with his (unknown) value for the item and chooses whether to enter the auction. Entry incurs an entry cost. In stage 2, the bidders who chose to enter learn their valuations but not the number of entrants, and submit bids.

In equilibrium, a bidder will enter the auction if his signal exceeds an entry threshold. Therefore, the bidders that enter are not representative of the entire bidder pool. They are selected based on a pre-entry signal correlated with value, so they have higher values in expectation. This is the sense in which entry is selective. The model spans two extremes: the perfectly selective model of Samuelson (1985), in which signals equal values, and the nonselective model of Levin and Smith (1994), in which bidders have no signal about value prior to entry. In particular, it allows for imperfect selection through a joint distribution of signals and values. I refer to this general model whenever I use the term "selective entry."

In the bidding stage, the first-price sealed-bid auction has a pure-strategy equilibrium in

[3]See McAfee and McMillan (1987), Matthews (1987), Krishna (2010).

[4]The IPV assumption is discussed in Section 3.

which bids are monotonic in values. In the English auction, it is a dominant strategy for each bidder to bid his value for the item.

After specifying the model, I carefully consider identification of the model from data as follows. The observable data include all bids and bidder identities in the first-price sealed-bid (S) auction, only the transaction price and winner's identity in the English auction, and (a measure of) the number of potential bidders. The number of realized entrants in each auction is observed for S but not for O; this is a common limitation of English auction data. The auction model is not identified from observed bids alone, in either auction format, absent some excludable variation in the data that can be exploited. I use variation in the auction format and variation in the entry threshold to establish nonparametric identification of the full model. The details are as follows. First, I show that value distributions conditional on entry - that is, conditional on a bidder's signal exceeding an entry threshold - are identified for asymmetric bidders from the English auction data. The argument is based on a result from reliability theory and extends a theorem in Athey and Haile (2002) to cases where the number of entrants is unobserved. Second, value distributions conditional on point values of the signal are subsequently identified if there is excludable and continuous variation in the entry threshold, as discussed previously in Gentry and Li (2014). The entry cost shifter provides this variation. Third, if the pool of potential bidders is the same for the two auction formats S and O, a bidder's utility function is identified as the function that satisfies his first-order condition for bidding in S, which maps value distributions identified in the first two steps to bid distributions observed in S. This follows the logic of Lu and Perrigne (2008), but the distributions here must be conditioned on entry thresholds. Finally, the entry cost is identified as that which makes the bidder with a marginal signal indifferent between entering and not entering the auction.

Closely based on this identification strategy, I develop a multi-step estimation procedure that recovers the structural parameters of the auction model nonparametrically. The first part is a sieve maximum likelihood estimator that estimates bidders' value distributions from English auction data. I assess the finite sample performance of this estimator in a Monte

Carlo study. The rest of the estimation procedure estimates bidders' utility functions and entry costs by following the constructive identification argument above step by step.

Finally, I apply the procedure to the New Mexico auction data. According to the estimation results, entry is indeed selective; for a representative auction item, the marginal entrant (whose pre-entry signal just meets the entry threshold) has a median value that is 40% lower than that of entrants overall. Bidders are moderately risk-averse, at a level similar to what has been measured in previous studies of risk aversion. In counterfactual simulations using the estimated model, I first ask how the entry process affects current auction revenue. I find that bidders' entry-induced uncertainty about the number of other entrants, combined with risk aversion, boosts revenue in first-price sealed-bid auctions while not affecting English auctions. The boost is largest when the number of realized entrants is low, alleviating the negative effect of low competition. Empirically, the magnitude of this boost is often similar to that of having one additional entrant. As such, this is a first-order consideration for sellers choosing between first-price sealed-bid auctions and English auctions in low competition environments. Second, I compare revenue under selective entry versus nonselective entry when policies lower the entry threshold, quantifying how much less responsive revenue is to the entry rate when entry is selective. Finally, I ask the theoretically ambiguous question of whether expanding the potential bidder pool would increase revenue,[5] and find that the effect is positive but moderate relative to the size of the expansion.

This paper contributes to the literature by estimating a nonparametric model of selective entry on empirical data. As a result of this analysis, it demonstrates an economically important reason to use first-price sealed-bid auctions over English auctions that has not received attention in the literature to date. The closest related literature can be divided into two main categories. The first category involves testing and identification of selective entry in auction models. Under risk neutrality, Marmer, Shneyerov, and Xu (2013) propose a test for different models of entry, while Gentry and Li (2014) prove identification given excludable variation in the entry threshold. Under risk aversion, Li, Lu, and Zhao (2015) develop a test

[5]Li and Zheng (2009) show that when entry is endogenous, revenue may or may not increase with the number of potential bidders.

for the form of risk aversion by comparing observed entry rates in first-price and ascending auctions for symmetric bidders, while Gentry, Li, and Lu (2015) show identification given excludable variation in potential bidders and a parametric signal-value copula. These papers do not perform estimation on empirical data. Incidentally, my paper diverges from Gentry et al. (2015) in its identification strategy; I exploit variation in the auction format and entry threshold in place of the conditions they exploit.[6] Also, I clarify that the starting point for Gentry and Li (2014)'s analysis is that value distributions conditional on entry are already identified. I identify these distributions from English auction data where the number of entrants and losing bids are unrecorded and bidders are asymmetric.

The other category of closely related work consists of empirical applications that estimate models of selective entry in a fully parametric approach. Roberts and Sweeting (2013) and Bhattacharya, Roberts, and Sweeting (2014) examine the merit of different ways to organize bidders' entry - for instance, unrestricted entry, an entry rights auction, or sequential entry - when entry is selective.[7] Roberts and Sweeting (2016) account for selective entry as they study how firm bailouts affect revenue in subsequent auctions. In all of these applications, bidders are modeled as risk-neutral. My paper takes a nonparametric approach with a generalized model where risk aversion plays an important role. Also, the questions addressed in my paper are distinct from those asked in this body of work.

The broader related literature includes empirical studies that model bidders entering nonselectively, as in Levin and Smith (1994). Examples include Bajari and Hortaçsu (2003) on eBay auctions, Athey, Levin, and Seira (2011) on U.S. Forest Service timber auctions, and Krasnokutskaya and Seim (2011) on government bid preference programs, for risk-neutral bidders. My paper also relates to analyses documenting auction behavior consistent with risk aversion. For instance, Ackerberg, Hirano, and Shahriar (2006) use risk aversion to explain bidders' propensity to take the buy-it-now option in eBay.

The paper proceeds as follows. Section 2 introduces the NMSLO auctions and presents

[6] In the NMSLO data, excludable variation in the number of potential bidders is difficult to measure; see Section 6.3.

[7] Sweeting and Bhattacharya (2015) study this question computationally.

empirical evidence. Section 3 lays out an auction model with selective entry for asymmetric, risk-averse bidders. Section 4 discusses identification of the structural model from data, and Section 5 develops an estimation procedure. Section 6 discusses estimation details specific to the NMSLO data along with the estimation results. Section 7 discusses insights gleaned from counterfactual simulations. Section 8 concludes. The appendix collects all proofs not presented in the main text.

2 New Mexico's Oil and Gas Lease Auctions

2.1 Overview

The New Mexico State Land Office (NMSLO) administers oil and gas leases on its trust lands. These leases grant the lessee the exclusive right to drill the leased land for a specified number of years. In return for the lease, the lessee pays the lessor an upfront lump sum "bonus", which can be considered the price of the lease, in addition to an annual rental and royalties on production.[8] These leases are sold via monthly auctions where bidders bid on the amount of the bonus. 80% of the leases auctioned since 2005 are located in the Permian Basin, a long established oil and gas producing area where the geology is well known. To avoid excessive heterogeneity in auctions, I focus my study on leases located within the Permian Basin.

The NMSLO uses both the first-price sealed-bid (S) and English or ascending oral (O) auction in the Permian Basin, generally splitting available land between the two formats.[9] It employs a public reserve price of approximately $15.625 per acre. 320 acres (half a square mile) is by far the most common tract size, and leases have a term of five years. The annual rental is either $0.50 or $1 per acre depending on geographic location. The royalty rate is

[8] The agency's stated goal is to "optimize revenues while protecting the health of the land for future generations." It is responsible for administering 13 million acres of subsurface estate; this is roughly 17% of the total area of New Mexico.

[9] The NMSLO uses symbols S and O for "sealed" and "oral" to represent the two types of auctions, so I use the same. Appendix Table 8 uses production data to check that assignment between the two formats is fairly random.

determined by the SLO's assessment of tract potential; tracts deemed "regular" are assigned a "V0" lease prefix with 16.67% royalty, and tracts deemed "premium" are assigned a "VB" lease prefix with 18.75% royalty.[10]

Conversations with agency staff as well as bidders reveal that valuations of a lease are idiosyncratic by bidder. Firms have different probabilities of drilling the tract within the five-year lease term, which depends on how they see the lease fitting into their overall portfolio and development strategy.[11] They also differ in well and field design, recovery rates, aggressiveness of hedging programs, cash flow, and alternative options for land acquisition to name some examples. All of these things factor into how they value a lease.

The NMSLO records the dollar amount and bidder identity for every bid submitted in the first-price sealed-bid (S) auction. Only the transaction price and winner identity are recorded for the English (O) auction. Hence, the number of entrants for each lease is observed in the data for S but not for O. Figure 1 shows that overall, the number of entrants per auction is not high, with a mean and median of 3 sealed bids. 46% of auctions receive 2 or fewer bids. SLO staff say they usually observe about 15 to 20 bidders in the auction room, and the number of bidder names observed over time is larger, so those who bid on a given lease are a small fraction of the entire bidder pool.

2.2 Entry and uncertainty about the number of entrants

Figure 2 plots a histogram of the price per acre obtained in auction, separately for each auction format. One immediately noticeable difference between the two formats is that O has a large concentration at the bottom end of prices, while S does not. Those first three bars in the O histogram consist almost entirely of reserve price sales, where the tract sold exactly at the reserve price. By nature of the auction format, a reserve-price O sale indicates that only one bidder raised his hand at the English auction. So the mass at the bottom of the O histogram is caused by auctions with one entrant.

[10] Rarely, the NMSLO assigns a "VC" label with 20% royalty. VC leases comprise only 0.2% of the leases auctioned in the Permian Basin since 2005.

[11] Drilling is not obligated; it is an option. It is very common for leases to expire without drilling.

We know from Figure 1 that 17% of S auctions also receive only one bid. So why doesn't the S histogram have a similar mass at the reserve price? It turns out that in the S auctions, one-bidder leases sold for on average 8 times the reserve price. Clearly, the lone bidders in the S auction did not know beforehand that they would be the only bidder. If they had known for certain, they would have bid the minimum acceptable bid, so these leases would have sold at the reserve price, just as in the O auction. This demonstrates that bidders are uncertain about the number of entrants they will compete against. To generate such uncertainty, there must be an entry process whose aggregate outcome is unknown to bidders at the time of bidding.

Conversations with bidders reveal that there is a cost to bidding, part of which is the work required to analyze a tract up for lease.[12] There is also an associated opportunity cost; a firm may have to forgo bidding on other leases in order to bid on that one. In describing the timeline of entry, bidders say they first look through the published list of auction items to select the ones they like enough to bid on. This decision is based on things they can discern from the tract description. Then they do work to analyze only those tracts in detail and decide how much to bid. This suggests a selective entry process - the bidder's preliminary signal about the item must be high enough to warrant bidding - and values are finalized only after incurring an entry cost as described above.

2.3 Risk aversion

Let n denote the number of entrants in an auction. As seen in the example of one-bidder auctions, bidders make a sizable adjustment to their S bids in response to uncertainty about n. There is reason to suspect an important role for risk aversion in these auctions. Theory (see Krishna (2010)) says that if risk-neutral bidders with independent private values face uncertainty about n in the S auction, they will bid a weighted average of the optimal bid under each realization of n. The weightings are such that cases of "overbidding" balance out cases of "underbidding", and expected revenue is unaffected by the uncertainty. On the

[12] Bidders say they analyze S auction and O auction tracts the same way.

other hand, McAfee and McMillan (1987) and Matthews (1987) show that when bidders are risk-averse, uncertainty about n does increase expected revenue in the S auction. This makes intuitive sense; a risk-averse bidder is willing to bid more to reduce the risk of losing when more bidders turn out than expected. The majority of bidders in the NMSLO auctions are local independent operators rather than major integrated companies, so risk aversion with regards to these auctions seems natural.[13][14]

Meanwhile, neither risk aversion nor uncertainty about the number of entrants affects revenue in the O auction, where bidders have a dominant strategy for bidding. Therefore, if bidders with independent private values are uncertain about the number of entrants and risk-averse, revenue equivalence does not hold; revenue in S should exceed revenue in O. According to Table 1, which regresses auction price on auction format and observable characteristics of the lease, the data aligns with this expectation. This observation is also consistent with early intuition in natural resource policy that preceded (and did not have the benefit of) the auction theory work cited here. Surveying federal and state natural resource auctions to compare oral and sealed bidding, Mead (1967) remarks that even when a lack of bidder interest "results in one-bidder sales under sealed bid procedures, such sales may yield a price close to a competitive price," and where "competition is unreliable, sealed bidding is the more appropriate method since it introduces a measure of uncertainty."

[13] I manually collected partial data on the employee count at firms bidding in the Permian Basin in 2005-2014 from firm websites, Businessweek, Cortera, Hoovers, Linkedin, Manta, Wikipedia, Yahoo, and other sources. There were 187 names making sealed bids. Excluding bids from the most frequent bidder which I discuss in the next section, 38% of sealed bids came from bidders with up to 10 employees, 30% came from bidders for which I could not find employee information, 16% came from bidders with 11-49 employees, and the remaining 16% came from bidders with 50+ employees.

[14] It is interesting to note the place of risk aversion in the experimental literature, where bidders are often found to overbid relative to risk-neutral Nash in first-price auction experiments. Risk aversion has long been considered a candidate explanation for the overbidding (for instance, see Cox et al. (1982), Cox et al. (1983)). In a more recent paper, Bajari and Hortaçsu (2005) take data from a first-price auction experiment and compare four alternative structural models in their ability to recover bidders' value distributions from observed bids. They find that a risk-averse Bayes-Nash model performs better than both risk-neutral Bayes-Nash and behavioral models of bidding.

2.4 Asymmetry among bidders

While there may be many types of asymmetry among bidders who participate in the NMSLO auctions, the difference between the most frequent bidder, called Yates Petroleum, and the rest of the bidders is particularly striking in this data. The most frequent bidder won nearly 30% of all auctions, while no other bidder's share exceeded single digit percentages. In terms of entry, the most frequent bidder bid on 80% of all S auctions held since 2005, while the next highest bid rate was only 21%. As Maskin and Riley (2000a) show, asymmetry among bidders has implications for auction revenue, with different implications for different auction formats.

3 Model of auctions with selective entry for asymmetric, risk-averse bidders

Motivated by the empirical evidence, I build a model of auctions with selective entry, where bidders are allowed to be risk-averse and asymmetric. I assume independent private values (IPV). This is not to say common values are not present in the empirical application; as in most real world auctions, private and common value components likely coexist. I provide a few comments relating to this modeling assumption. First, the Permian Basin is an area where knowledge of the geology is more complete due to a long history of development and production dating back to the 1920s. Permits for new seismic surveys are rarely requested in the basin, as these are only done in areas that are not well known and much of the basin has already been drilled in the past. Electric wireline logs from all drilled wells are made public by the New Mexico Oil Conservation Division.[15] Moreover, the data sample covers years 2005-2014, coinciding with the boom in horizontal drilling and hydraulic fracturing, which also

[15] Where "seismic testing and geophysical logs+ have revealed the extent of producing zones with high degrees of certainty, the geologic risk [...] is much lower than that of a Wildcat well." - Extract from New Mexico Oil Conservation Division Hearing, Case No. 14744, Jalapeno Corp. and Harvey E. Yates Company's Proposed Findings of Fact and Conclusions of Law.

reduce production uncertainty.[16] This is important because the key feature of common value auctions is that I as a bidder think other bidders have value-relevant information - e.g. about how much oil is underground - that I do not possess. If all bidders have similar assessments of underlying oil, non-common components are the primary driver of value differences, and that is a private value auction.

Second, under the common values paradigm, it is difficult to rationalize higher revenue in S than in O. As shown in Milgrom and Weber (1982) and discussed in Krishna (2010), the O auction should outperform the S auction in that case, and uncertainty about the number of entrants should make revenue in S fall even more relative to O due to the linkage principle. In order to make feasible progress analyzing model features indicated by the major revenue patterns discussed in Section 2, I stay within the IPV paradigm rather than adopt the fully general paradigm of both private and common value components.

3.1 Setup

To accommodate the asymmetry discussed in section 2.4, there are 2 subgroups of bidders, denoted by $m \in \{1, 2\}$. The number of potential bidders N_m in each subgroup is assumed to be common knowledge. Each subgroup has potentially different distributions of values and signals, $F_m(v, s)$, where v are values and s are signals that I describe below. $F_m(v|s)$ is the distribution of values conditional on a signal. Utility functions and entry costs may also vary by subgroup. To allow for risk aversion, let $U_m(\cdot)$ be the twice continuously differentiable utility function of bidders from subgroup m with $U_m(0) = 0$, $U'_m(\cdot) > 0$ and $U''_m(\cdot) \leq 0$.

The entry process follows the "Selective Entry model" or "Affiliated Signal Model" laid out in Marmer, Shneyerov, and Xu (2013) and Gentry and Li (2014). In this entry model, an indivisible good is auctioned via a two-stage auction game. In stage 1, each potential

[16] "Previously an oil company had to intersect a geologic structure to be successful so the affected areas were very small and the possibility of a dry hole was large. However with these new technologies the oil companies are now able to drill into the oil bearing source rock so there are no dry holes, and the target area is now in the form of a blanket (your oil bearing formation) as opposed to a point (a geologic structure)." - http://www.niobraranews.net.

To quote another person familiar with the industry, "for unconventional, all firms will have very similar resource assessments. The oil is there, there's no uncertainty about that in shale. Firms don't even wildcat in unconventional. The risk is more about, how naturally fractured is the shale, will we hit a fault line, etc."

bidder i observes a private signal s_i of his (unknown) private value v_i. The signal s_i may be informative about v_i, but it need not be perfectly informative. Upon observing this private signal, all potential bidders simultaneously choose whether to bid on the item, i.e. whether to enter the auction. Entry incurs an entry cost $c_m(x)$, where x are any entry cost shifters. In stage 2, the bidders who chose to enter in stage 1 learn their valuations v and submit bids for the object being sold.

This model nests two standard models of entry that lie on opposite ends of a spectrum: the Levin and Smith (1994) model, in which bidders have no informative signal about their value prior to entry, and the Samuelson (1985) model, in which bidders know their exact v prior to entry. Unlike these models, the entry model used here allows for imperfect correlation between a bidder's pre-entry signal s_i and his post-entry value v_i. Without loss of generality, first-stage signals are normalized to have a uniform marginal distribution on [0,1]: $s \sim U[0,1]$.

Assumptions

A1 Independence across bidders: $(v_i, s_i) \perp (v_j, s_j)$ for all $j \neq i$.

A2 $F_m(v, s)$ have square support $[\underline{v}, \bar{v}] \times [0,1]$.

A3 Stochastic ordering: $s' \geq s$ implies $F_m(v|s') \leq F_m(v|s)$.

A4 $F_m(v|s)$ is differentiable in s and continuously differentiable in v.

A5 The reserve price r is the lower bound of v, i.e. $\underline{v} = r$.

Assumptions A1-A4 mostly follow Gentry and Li (2014). Assumption A1 is an implication of the IPV paradigm; A2 means all values in the support of v have positive density conditional on any value of s; A3 roughly says bidders with higher signals are likely to have higher values; and A4 is about differentiability and smoothness of the value distribution. Assumption A5 effectively attributes all non-bidding to the entry cost. While a more general model would encompass non-bidding due to a combination of reserve price and entry cost, it is not clear how to identify $F(r|s)$ separately from the entry cost without some normalization. In the context of the NMSLO auctions, the agency considers the reserve price a "starting point" for serious bidders and tries not to offer tracts for which the reserve price might be binding. Thus r is in the vicinity of $\underline{v}$ for offered tracts.

In the remainder of this section I discuss the pure strategy Bayesian Nash equilibrium for this auction model. For economy of notation, I will omit conditioning on x (entry cost shifters) and $\{N_m\}$ (number of potential bidders) in this section. Conceptually, imagine the case where x and $\{N_m\}$ are being held constant at specific values for clarity.

3.2 First-price sealed-bid auction (S)

Suppose for now that the stage 1 entry decision involves entry thresholds $\bar{s}_m \in [0, 1]$ such that a bidder in subgroup m chooses to enter if and only if $s_i \geq \bar{s}_m$. Each subgroup may have a different $\bar{s}_m$. (I write $\{\bar{s}_m\}$ as shorthand notation for $\{\bar{s}_1, \bar{s}_2\}$.) Since $s \sim U[0, 1]$, the entry threshold $\bar{s}_m$ also represents the probability that a potential bidder from subgroup m will not enter the auction. Then for bidders from subgroup m, the distribution of values conditional on entry is

$$F_m^*(v; \bar{s}_m) \equiv \frac{1}{1 - \bar{s}_m} \int_{\bar{s}_m}^{1} F_m(v|t)dt. \tag{1}$$

By Assumption A2, the support of this distribution remains $[\underline{v}, \bar{v}]$. The related density is denoted $f_m^*(v; \bar{s}_m)$. Then the probability that a bidder does not enter or enters and draws value less than v is $\bar{s}_m + (1 - \bar{s}_m)F_m^*(v; \bar{s}_m) = \bar{s}_m + \int_{\bar{s}_m}^{1} F_m(v|t)dt$. This probability is increasing in the entry threshold $\bar{s}_m$; namely, its derivative with respect to $\bar{s}_m$ is $1 - F_m(v|\bar{s}_m) > 0$.

Equilibrium bidding

First, consider stage 2 of the auction game, in which bidders learn their v_i and decide what to bid. Suppose a monotonic pure-strategy equilibrium is being played, and every competitor from subgroup m employs a monotonic bidding strategy denoted by $b_m(v)$. Define

$$G_m(b|\{\bar{s}_m\}) \equiv \bar{s}_m + (1-\bar{s}_m)F_m^*(b_m^{-1}(b);\bar{s}_m),$$

which is the probability that a potential bidder from subgroup m either does not enter or enters and bids less than b. This object is conditional on the entry thresholds $\{\bar{s}_m\}$, so all subsequent functions of G_m are also conditional on $\{\bar{s}_m\}$. The probability of winning W, for an entrant from subgroup 1 who bids $b \geq r$, is the probability that all competitors either do not bid or bid less than b, i.e.

$$W_1(b|\{\bar{s}_m\}) \equiv G_1(b|\{\bar{s}_m\})^{N_1-1}G_2(b|\{\bar{s}_m\})^{N_2}.$$

Note that uncertainty regarding the number of entrants is built into W, as each potential bidder will not bid with probability $\bar{s}_m$. Now, the expected stage 2 profit of a subgroup 1 entrant who draws value v_i and bids $b \geq r$ is

$$\pi_1(v_i, b|\{\bar{s}_m\}) \equiv U_1(-c_1+v_i-b)W_1(b|\{\bar{s}_m\}) + U_1(-c_1)(1-W_1(b|\{\bar{s}_m\})). \tag{2}$$

To elaborate, in stage 2, the bidder has entered the auction, so he pays the entry cost c whether or not he wins. If he wins, he gets utility $U_1(-c_1+v_i-b)$, and if he loses, he gets $U_1(-c_1)$.

Then, the bidder's maximization problem in choosing bid b given his value v_i is

$$\max_b \pi_1(v_i, b|\{\bar{s}_m\}),$$

which yields a first-order condition for bidding:

$$\frac{U_1'(-c_1+v_i-b)}{U_1(-c_1+v_i-b)-U_1(-c_1)} = \frac{dW_1(b|\{\bar{s}_m\})/db}{W_1(b|\{\bar{s}_m\})}. \tag{3}$$

Intuitively, the bidder chooses a bid that balances its marginal effect on the probability of winning against its marginal effect on utility conditional on winning. In equilibrium, $b = b_m(v)$. Writing out the function W in terms of value distributions and imposing $b = b_1(v)$

results in the following differential equation for the equilibrium bidding strategy of subgroup 1:

$$\begin{aligned}\frac{1}{b_1'(v)} \quad = \quad & \frac{[\bar{s}_1 + (1-\bar{s}_1)F_1^*(v;\bar{s}_1)]}{(N_1-1)(1-\bar{s}_1)f_1^*(v;\bar{s}_1)}\left\{\frac{U_1'(-c_1+v-b_1(v))}{U_1(-c_1+v-b_1(v))-U_1(-c_1)}\right.\\ & \left.-\frac{N_2(1-\bar{s}_2)f_2^*(b_2^{-1}(b_1(v));\bar{s}_2)}{b_2'(b_2^{-1}(b_1(v)))[\bar{s}_2+(1-\bar{s}_2)F_2^*(b_2^{-1}(b_1(v));\bar{s}_2)]}\right\}.\end{aligned} \tag{4}$$

Switching the subgroup 1 and 2 subscripts in (4) gives the equivalent differential equation for subgroup 2, forming a system of two first order differential equations. Lebrun (1999), Maskin and Riley (2000a), and Maskin and Riley (2000b) among others have derived systems of differential equations characterizing equilibrium bidding with asymmetric bidders. In general, closed-form solutions cannot be obtained for the bid functions.

A monotone pure-strategy equilibrium exists for this stage 2 bidding game, as can be shown by relating this game to the framework in Reny and Zamir (2004).[17] As for uniqueness, the equilibrium may not be unique if $U_1(\cdot) \neq U_2(\cdot)$ or $c_1 \neq c_2$. In estimation, I will assume that a single equilibrium is being played in the data. I denote the bidding strategies of the selected equilibrium with an asterisk, $b_1^*(v|\{\bar{s}_m\})$ and $b_2^*(v|\{\bar{s}_m\})$.

Equilibrium Entry

Now going back to stage 1 of the auction game where potential bidders decide whether to bid based on their signal s, the expected profit from entering the auction for a potential bidder i from subgroup 1 is

$$\Pi_1(s_i|\{\bar{s}_m\}) \equiv \int_{v=\underline{v}}^{\bar{v}} \pi_1\left(v, b_1^*(v|\{\bar{s}_m\})\middle|\{\bar{s}_m\}\right) f_1(v|s_i)dv. \tag{5}$$

This is the expectation of $\pi_1(v_i, b^*(v_i))$ with respect to v_i given the signal s_i. Naturally, a potential bidder will choose to enter whenever the expected profit of entering the auction

[17] Reny and Zamir (2004)'s model accommodates first-price auctions with risk-averse, asymmetric bidders and a binding reserve price. The stage 2 bidding game here can be reformulated as a first-price auction with a binding reserve price, where bidders draw $v < r$ with probability $\bar{s}_m$ and draw $v \sim F_m^*(v;\bar{s}_m)$ otherwise, so that each bidder does not bid with probability $\bar{s}_m$.

exceeds the profit of not entering; that is, whenever $\Pi_1(s_i|\{\bar{s}_m\}) \geq U_1(0)$. The utility of not entering the auction is $U_1(0)$ because the entry cost is paid only upon entry.

Taking the derivative of $\pi_1(v, b_1^*(v|\{\bar{s}_m\})|\{\bar{s}_m\})$ with respect to v and using the envelope theorem shows that $\pi_1(v, b_1^*(v|\{\bar{s}_m\})|\{\bar{s}_m\})$ is strictly increasing in v. In addition, $F_m(v|s)$ are stochastically ordered in s by the model's assumptions. Therefore, $\Pi_m(s_i|\{\bar{s}_m\})$ is strictly increasing in s_i, yielding a threshold strategy for entry in which bidders enter the auction if and only if s_i exceeds a threshold. This is true for a bidder regardless of the entry strategies used by other bidders as long as signals are independent across bidders. Hence, all bidders will indeed use threshold strategies when deciding whether or not to bid in an auction. An implication of this is that bidders who enter are not representative of the overall bidder pool; they are the ones with higher signals and therefore stochastically dominant value distributions, according to assumption A3.

In equilibrium, there is a zero profit condition for marginal entrants such that bidders whose signals are equal to the entry threshold just break even from entering. Using asterisks to indicate equilibrium entry thresholds, the following system of equations must be satisfied in equilibrium:

$$\begin{aligned} \Pi_1(\bar{s}_1^*|\{\bar{s}_m^*\}) &= U_1(0) \\ \Pi_2(\bar{s}_2^*|\{\bar{s}_m^*\}) &= U_2(0). \end{aligned} \tag{6}$$

3.3 English (ascending oral) auction (O)

Now I model the English (O) auction. I am more concise here since many concepts are analogous to the S auction and repeated. Suppose the stage 1 entry decision involves entry thresholds $\bar{s}_m^o \in [0, 1]$, where the "o" superscript indicates the O auction. Equation (1) still applies here, with $\bar{s}_m^o$ replacing $\bar{s}_m$. If only one bidder enters the auction, that bidder wins and pays the reserve price r.

Equilibrium bidding

First, consider stage 2 of the auction game. Under the private value paradigm, the English auction is equivalent to the second price auction. Even with selective entry, risk aversion, and asymmetry, it is still a dominant strategy for each bidder to bid his value. The probability of winning W^o, for an entrant from subgroup 1 who draws value v_i, is the probability that all competitors either do not enter or enter but have values less than v_i:

$$W_1^o(v_i|\{\bar{s}_m^o\}) \equiv [\bar{s}_1^o + (1-\bar{s}_1^o)F_1^*(v_i;\bar{s}_1^o)]^{N_1-1}[\bar{s}_2^o + (1-\bar{s}_2^o)F_2^*(v_i;\bar{s}_2^o)]^{N_2}.$$

Note that this expression also represents the distribution of the highest competing bid, since my probability of winning is equal to the probability that $\underset{j\neq i}{max}\{v_j\} < v_i$. (If no one else has bid, the highest competing bid is the reserve price r.) The expected stage 2 profit of an entrant from subgroup 1 who draws value v_i and bids his value is then

$$\begin{aligned}\pi_1^o(v_i|\{\bar{s}_m^o\}) \quad = \quad & \underbrace{U_1(-c_1+v_i-r)(\bar{s}_1^o)^{N_1-1}(\bar{s}_2^o)^{N_2}}_{\text{no other bidders enter}} + \underbrace{U_1(-c_1)(1-W_1^o(v_i|\{\bar{s}_m^o\}))}_{\text{lose}} \\ & + \underbrace{\int\limits_{y=\underline{v}}^{v_i} U_1(-c_1+v_i-y)dW_1^o(y|\{\bar{s}_m^o\}).}_{\text{win against other entrants}}\end{aligned} \tag{7}$$

In the first part of (7), $(\bar{s}_1^o)^{N_1-1}(\bar{s}_2^o)^{N_2}$ is the probability that no other bidders enter the auction. In the last part of (7), the integral is the expectation of $U_1(-c_1+v_i-y)$ with respect to y, the highest competing bid. In an English auction, the highest competing bid is the price paid upon winning. $\pi_2^o(v_i|\{\bar{s}_m^o\})$ for a subgroup 2 entrant can be written by switching the 1 and 2 subscripts in (7). It is easy to see that $\pi_m^o(v_i|\{\bar{s}_m^o\})$ is strictly increasing in v_i.

Equilibrium Entry

Now going back to stage 1 of the auction game, the expected profit from entering the auction for a potential bidder from subgroup 1 with signal s_i is

$$\Pi_1^o(s_i|\{\bar{s}_m^o\}) \equiv \int_{v=\underline{v}}^{\bar{v}} \pi_1^o(v_i|\{\bar{s}_m^o\})f_1(v_i|s_i)dv. \tag{8}$$

This is the expectation of $\pi_1^o(v_i|\{\bar{s}_m^o\})$ with respect to v_i given s_i. The bidder will choose to enter whenever the expected profit of entering exceeds the profit of not entering; that is, whenever $\Pi_1^o(s_i|\{\bar{s}_m^o\}) \geq U_1(0)$. Since $\pi_m^o(v_i|\{\bar{s}_m^o\})$ is strictly increasing in v_i and $F_m(v|s)$ is stochastically ordered in s, $\Pi_m^o(s_i|\{\bar{s}_m^o\})$ is strictly increasing in s_i given independence of signals across bidders. Hence, bidders will indeed use threshold strategies for entry, entering if and only if s_i exceeds a threshold.

In equilibrium, a zero profit condition must be satisfied for marginal entrants:

$$\begin{aligned} \Pi_1^o(\bar{s}_1^{o*}|\{\bar{s}_m^{o*}\}) &= U_1(0) \\ \Pi_2^o(\bar{s}_2^{o*}|\{\bar{s}_m^{o*}\}) &= U_2(0). \end{aligned} \tag{9}$$

4 Identification

I now investigate whether this auction model is identified from observable data. The goal is to see whether model primitives $F_m(\cdot|\cdot)$, $U_m(\cdot)$, and c_m for each subgroup can be recovered uniquely from the data at hand. For economy of notation, I omit conditioning on entry cost shifters x and the number of potential bidders $\{N_m\}$ unless it is necessary.

4.1 Identification of the English (O) auction model

I consider the English auction first. The observables here are the transaction price, identity of the winning bidder, and the number of potential bidders in each subgroup, $\{N_m\}$. I refer to these as the "O data." Bids other than the transaction price are not observed, nor the number of entrants. Since it is a dominant strategy for each bidder to bid his value in the English auction, the observed transaction price is the second highest value among all entrants unless there is only one entrant. If there is only one entrant, the transaction price equals the reserve price. Entrants are those whose pre-entry signals exceeded the entry threshold, i.e. $s_i \geq \bar{s}_m^o$. For the rest of this section, I omit the o superscript that indicates the O auction.

Identification of value distributions

Athey and Haile (2002), in their Theorems 2 and 3, establish that the value distributions of asymmetric IPV bidders are identified from transaction prices and winner identities alone when the number of bidders is known. In doing so they reference a result from Meilijson (1981) in the reliability theory literature. Can this reasoning be extended to my setting where the number of entrants is unobserved?

To think about this, I reframe the problem. If I temporarily assign a placeholder transaction price of r to auctions with 0 entrants, transaction prices in this auction are observationally equivalent to the second highest value out of $N_1 + N_2$ bidders who draw their values from distribution $H_m(v) \equiv [\bar{s}_m + (1 - \bar{s}_m)F^*_m(v; \bar{s}_m)]$. The distribution $H_m(\cdot)$ has an atom at its infimum representing the probability that a potential bidder does not enter the auction. Meilijson (1981)'s result need not apply here, because it assumes distributions are non-atomic. In Proposition 1, I show that a later paper in the reliability theory literature, Nowik (1990), allows me to extend the reasoning of Athey and Haile (2002) to auctions where each potential bidder bids with some probability and the realized number of entrants is unobserved.[18]

Proposition 1. $H_m(v) \equiv [\bar{s}_m + (1 - \bar{s}_m)F^*_m(v; \bar{s}_m)]$ *is identified, and consequently* $\bar{s}_m$ *and* $F^*_m(v; \bar{s}_m)$ *are identified.*

Simple intuition for Nowik (1990)'s whole proof is difficult to provide, but intuition for identifying $\bar{s}_m$ is as follows. Though the number of entrants is unobserved, we can deduce when there are no entrants - the item does not sell - and when there is one entrant - the transaction price equals the reserve price. The empirical probabilities of these events provide a system of equations that can be solved to identify $\{\bar{s}_m\}$. Specifically, the proof of Proposition 1 shows that

$$\bar{s}_m = \frac{\text{P(no one enters)}}{\text{P(winner's subgroup}= m\,\&\text{ winning bid}= r)/N_m + \text{P(no one enters)}}.$$

Now having identified $\bar{s}_m$ and $F^*_m(v; \bar{s}_m)$, the value distribution conditional on entry, it

[18] Independently, Komarova (2013) illustrates how an alternative approach to identifying asymmetric second-price auction models could extend to a stochastic number of bidders.

remains to identify $F_m(v|s)$, the value distribution conditional on a specific value of the signal. Since $F_m^*(v;\bar{s}_m) \equiv \frac{1}{1-\bar{s}_m}\int_{\bar{s}_m}^1 F_m(v|t)dt$, we have $F_m(v|\bar{s}_m) = -\frac{\partial}{\partial \bar{s}_m}[(1-\bar{s}_m)F_m^*(v;\bar{s}_m)]$. However, taking this derivative is feasible only if there is continuous variation in $\bar{s}_m$ and if the variable causing $\bar{s}_m$ to vary leaves $F_m(v|s)$ unaffected. Intuitively, if the entry threshold increases exogenously from $\bar{s}$ to $\bar{s}+\epsilon$, then the resultant change in value distribution conditional on entry can be attributed entirely to the dropping out of bidders whose pre-entry signals lie between $\bar{s}$ and $\bar{s}+\epsilon$. As $\epsilon \to 0$, this allows for identification of the distribution of values conditional on a specific signal. On the other hand, if the threshold increase is endogenous, the resultant change cannot be attributed entirely to those bidders dropping out, since the value distribution of remaining bidders changes simultaneously. Formally, the following assumptions allow recovery of $F_m(v|s)$ from $F_m^*(v;\bar{s}_m)$.

Assumptions:

A6 Exclusion restriction: $F_m(v|s,x) = F_m(v|s)$ while $\bar{s}_m(x)$ depends on x through $c_m(x)$

A7 Continuous variation in $\bar{s}$: $\forall s \in [0,1]$, $\exists x$ such that $\bar{s}_m(x) = s$

The exclusion restriction A6 enables $\bar{s}_m(x)$ to vary via changes in an entry-cost shifter x while leaving $F_m(v|s)$ unaffected.[19] Assumption A7 gives continuous variation of $\bar{s}_m(x)$ on the support of s. These correspond to the identifying assumptions in Gentry and Li (2014). Other forms of exclusion could work as well, the key being that there be some variation in $\bar{s}$ that the target $F(\cdot)$ is invariant to. Rewriting the relationship between $F_m^*(v;\bar{s}_m)$ and $F_m(v|s)$ with explicit conditioning on x gives

$$F_m^*(v;\bar{s}_m(x)) \equiv \frac{1}{1-\bar{s}_m(x)}\int_{\bar{s}_m(x)}^1 F_m(v|t)dt.$$

Rearranging and taking the derivative of both sides with respect to $\bar{s}_m$:

[19] An entry-cost shifter for the NMSLO auctions is discussed in section 6.2.

$$F_m(v|\bar{s}_m(x)) = -\frac{\partial}{\partial \bar{s}}[(1-\bar{s})F_m^*(v;\bar{s})]\bigg|_{\bar{s}=\bar{s}_m(x)}.$$

Given the assumptions, the right-hand side is calculable for any $\bar{s}_m$. Hence $F_m(v|s)$ is identified. If $\bar{s}_m$ varies continuously over a subset of [0,1], then $F_m(v|s)$ is identified for s in that subset.

Nonidentification of utility function $U_m(\cdot)$ and entry cost c_m

The discussion so far establishes that bidders' value distributions are identified from O data. However, Proposition 2 states that this is not the case for bidders' utility functions and entry costs.

Proposition 2. *$U_m(\cdot)$ and c_m are not identified from the O data.*

While Appendix B provides formal proof, the nonidentification result is intuitive; since bidders simply bid their value in the O auction, bids do not contain any information on the form of utility. The equilibrium entry conditions do, but they only provide a single restriction with which to identify both $U_m(\cdot)$ and c_m. To summarize, bidders' value distributions are identified from the English auction data, but utility functions and entry costs are not.

4.2 Identification of the sealed-bid (S) auction model

Next, I consider identification of the first-price sealed-bid auction model. The goal is to identify $F_m(\cdot|\cdot)$, $U_m(\cdot)$, and c_m from observable data. For the S auction, these include $\mathbb{G}(b_1...b_{N_1}, b_{N_1+1}...b_{N_1+N_0})$ (the joint distribution of bids, using a placeholder for non-bids), the number of potential bidders by subgroup, $\{N_m\}$, and the number of entrants, which I will denote $\{n_m\}$. I refer to these as the "S data."

It helps to rewrite the first-order condition for bidding in (3) as an inverse bid function in the style of Guerre et al. (2000). To do this, I follow the notation of Li et al. (2015) and define $\tilde{U}(x) \equiv U(x-c) - U(-c)$. Also, define $\lambda(x) \equiv \frac{\tilde{U}(x)}{\tilde{U}'(x)}$. Note that $\tilde{U}(0) = 0$ and therefore $\lambda(0) = 0$. Then the first-order condition for bidding can be rewritten as

$$\frac{\tilde{U}_m(v_m - b)}{\tilde{U}'_m(v_m - b)} = \lambda_m(v_m - b) = \frac{W_m(b|\{\bar{s}_m\})}{dW_m(b|\{\bar{s}_m\})/db}. \tag{10}$$

It is useful to know that $\lambda'(x) = \frac{[\tilde{U}'(x)]^2 - \tilde{U}''(x)\tilde{U}(x)}{[\tilde{U}'(x)]^2} = 1 - \frac{\tilde{U}''(x)\tilde{U}(x)}{[\tilde{U}'(x)]^2}$, where $\tilde{U}''(x) = U''(x-c) \leq 0$ and $\tilde{U}(x) > 0$ for $x > 0$. This means $\lambda(x)$ is strictly increasing for $x > 0$ and therefore invertible if x is restricted to $[0, \infty)$. Since $v - b \geq 0$, I can use the inverse of $\lambda(\cdot)$ to define an inverse bid function. Namely, each bidder's private value can be expressed in terms of the corresponding bid b, an observed distribution W, and the function $\lambda^{-1}(\cdot)$:

$$v_m = b + \lambda_m^{-1}\left(\frac{W_m(b|\{\bar{s}_m\})}{dW_m(b|\{\bar{s}_m\})/db}\right) \equiv \xi_m(b). \tag{11}$$

If $\lambda(\cdot)$ were known - as would be the case if bidders were risk neutral - we could compute $\xi_m(b)$ for any observed bid b, and this would at least identify the distribution of v_m conditional on entry, $F^*_m(v; \bar{s}_m)$.[20] However, $\lambda(\cdot)$ is unknown here.

Proposition 3. $[F_m(\cdot|\cdot), U_m(\cdot), c_m]$ *are not identified from the S data without further assumptions.*

While Appendix B provides formal proof, intuition for non-identification is as follows. There are two functions - the utility function and the value distribution - to be identified, but there is only one observable function - the bid distribution - available for use when trying to recover these objects. In general, there would be multiple pairs of utility functions and value distributions, along with an entry cost, that can rationalize the data. This relates to the nonidentification result in Guerre et al. (2009), which states that the first-price auction model with risk-averse bidders is not identified from the observed distribution of bids.

Remark 1. In the symmetric bidders case, the structure is nonparametrically identified if the same entry threshold $\bar{s}$ is observed for two different values of exogenous N (number of potential bidders) in the data. Alternatively, Gentry et al. (2015) show that the model can be identified with a parametric restriction on the copula describing the joint distribution of s and v, if there is exogenous variation in the number of potential bidders.

[20] See Guerre et al. (2000).

Despite non-identification of the model, the entry threshold $\bar{s}_m$ is directly identified. The number of entrants n_m has a binomial distribution with probability of entry $(1 - \bar{s}_m)$, so $E[n_m] = N_m(1 - \bar{s}_m)$, and hence $\bar{s}_m = 1 - E[n_m]/N_m$.

4.3 Identification using both the O and S auction

To summarize, the IPV auction model with selective entry, risk aversion, and asymmetric bidders is not identified from either the English (O) or first-price sealed-bid (S) auction alone absent additional conditions beyond A1-A7. However, if the O auction and S auction share the same pool of potential bidders, and thus the same $F_m(v|s)$ and $U_m(\cdot)$, we may be able to gain identifying power by using data from both auction formats simultaneously.[21] Intuitively, value distributions would be identified from O data as explained in section 4.1; then taking these as given, utility functions (or to be precise, $\lambda(\cdot)$) would be identified from S auction bids as the function that satisfies the first-order condition for bidding in (11); finally, entry costs for each auction format would be identified from the respective equilibrium entry conditions (6) and (9). The next proposition confirms this idea.

Proposition 4. *If the O auction and S auction share the same* $F_m(\cdot|\cdot)$ and $U_m(\cdot)$, *then the model primitives for both of these auctions are identified nonparametrically from observable data given A1-A7. The entry cost* c_m *need not be the same between O and S.*

This proposition provides a basis for estimating the full auction model, which is the topic of the next section.[22]

[21] I discuss whether this condition applies to the NMSLO data at the beginning of Section 6.

[22] I leave for future research the question of how to incorporate unobserved heterogeneity when value distributions are identified from English auctions where only transaction prices are observed. Krasnokutskaya (2011) and Hu et al. (2013) are well known papers that study identification of auction models with unobserved heterogeneity. However, methods like these require at least two bids per auction item to deconvolute unobserved heterogeneity while estimating the value distribution.

5 Estimation

I develop an estimation procedure that recovers $F_m(\cdot|\cdot)$, $U_m(\cdot)$, and c_m from the O and S data. The multi-step procedure closely follows the nonparametric identification argument presented in the proof of Proposition 4. It can be summarized in the following four steps. First, I estimate the entry threshold as the observed probability that a potential bidder does not enter the auction. Second, I estimate conditional value distributions from O data using a sieve maximum likelihood estimator. This estimator maximizes the likelihood of observed auction prices and winners, approximating value distributions with Bernstein polynomials. Third, using value distributions as estimated in the second step and observed distributions of sealed bids, I estimate the nonparametric utility function that satisfies the first-order condition for bidding in S. Finally, I estimate the entry cost that makes the bidder with a marginal signal indifferent between entering and not entering the auction. As I describe the estimation procedure, I introduce notation for auction covariates z, which are characteristics that describe the auction item. Signals are normalized so that $s|z \sim U[0,1]$.

5.1 Estimation of entry thresholds $\{\bar{s}_m\}$

As an intermediate step to recovering model primitives $F_m(\cdot|\cdot), U_m(\cdot)$, and c_m, I need to know the selective entry threshold $\bar{s}_m$. Given $\{N_m\}$, $\bar{s}_m$ is a function of entry cost c_m and auction covariates z. Since c_m itself is a function of z and entry-cost shifter x, $\bar{s}_m(z, \{c_m\})$ can be expressed as $\bar{s}_m(z,x)$. Since $s \sim U[0,1]$, the entry threshold is equal to the probability that a potential bidder does not enter the auction.

In the first-price sealed-bid (S) auctions, the number of entrants n_m is observed. So

$$\bar{s}_m(z,x) = 1 - \frac{\mathbb{E}[n_m|z,x]}{N_m},$$

where $\mathbb{E}[n_m|z,x]$ can be estimated nonparametrically. If (z,x) are large vectors such that nonparametric estimation is impractical, one could use parametric methods as well. For instance, one could estimate a binomial logistic regression with $n_m \sim \mathcal{B}(N_m, \frac{1}{1+e^{-\beta X}})$, where

$\mathcal{B}(\cdot,\cdot)$ is the binomial distribution and X is a vector of z and x. Then $\mathbb{E}[n_m|z,x]$ is the value of n_m predicted by that estimation given z, x.

In the English (O) auctions, the number of entrants n_m is not observed as only winning bids are known. However, $\bar{s}_m(z,x)$ can still be computed for each z, x using

$$\bar{s}_m = \frac{\text{P(no one enters)}}{\text{P(winner's subgroup}= m\ \&\ \text{winning bid}= r)/N_m + \text{P(no one enters)}}$$

as shown in the proof of Proposition 1. Nonparametric or parametric methods could be used to estimate these probabilities as a function of z and x. Let $\hat{\bar{s}}_m$ denote the estimated entry threshold.

5.2 Estimation of value distributions $F_m(v|s,z)$ from O data

The task of estimating $F_m(v|s,z)$ is demanding on the data. Recall that not all bids, but only transaction prices, are observed for the O auction. Furthermore, each subgroup has its own distribution, and the distributions are conditional on both signals s and covariates z. Finally, we do not observe the private signals s; all we know is that $s_i \geq \bar{s}_m$ for each entrant. I propose a sieve maximum likelihood estimator using Bernstein polynomial bases to perform this task. Properties of sieve estimators, including sieve maximum likelihood, are discussed in Chen (2007).[23]

In the O data, we see for each item $k = 1, ..., K$ the transaction price (the second highest v among entrants) and the identity of the winning bidder. So the log likelihood of the observed data is the log likelihood of observed transaction prices p and winner's subgroups m conditional on entry thresholds $\{\bar{s}_m\}$ and auction covariates z.[24] Namely,

$$\sum_{k=1}^{K} \log(\text{P(2nd highest } v \text{ among entrants} = p_k\ \&\ \text{winner's subgroup} = m_k|\{\bar{s}_m\}_k, z_k)). \quad (12)$$

[23]Komarova (2017) is an example of using Bernstein polynomials for sieve estimation of distribution functions in an ascending auction framework.

[24]The transaction price is a continuous variable, so one should think of P(2nd highest v among entrants $= p$) as a density.

The idea is to estimate $F_m(v|s,z)$ by finding the conditional distribution of v that maximizes this log likelihood. This is conceptually straight forward, but we do not observe s and only know that $s_i \geq \bar{s}_m$ for each bid. The log likelihood is a more complex function of $F_m(v|s,z)$ than it would be if s were known.

Before writing the mathematical expression for (12), some notation is useful. Let $H_m(p;\bar{s}_m,z)$ be the probability that a bidder either does not bid ($s_i \in [0,\bar{s}_m)$) or does bid ($s_i \in [\bar{s}_m,1]$) with $v \leq p$, when his entry threshold is $\bar{s}_m$ and auction covariates are z. That is, $H_m(p;\bar{s}_m,z) \equiv \bar{s}_m + \int_{\bar{s}_m}^{1} F_m(p|s,z)ds$. Also define the derivative $h_m(p;\bar{s}_m,z) \equiv \partial H_m(p;\bar{s}_m,z)/\partial p$. Now if we let H_m and h_m be shorthand for $H_m(p;\bar{s}_m,z)$ and $h_m(p;\bar{s}_m,z)$, respectively, the likelihood of observing transaction price $p \in (r,\bar{v}]$ and a winning bidder from subgroup 1 is

$$N_1(1-H_1)[(N_1-1)h_1 H_1^{N_1-2} H_2^{N_2} + N_2 h_2 H_1^{N_1-1} H_2^{N_2-1}]. \tag{13}$$

This is the probability that the second highest value among entrants is p and the winner is from subgroup 1 when there are $\{N_m\}$ potential bidders who each "draw" their value from $\{H_m(\cdot;\bar{s}_m,z)\}$, which has built in the fact that each potential bidder from subgroup m has probability $\bar{s}_m$ of not bidding.[25]

Now, to use sieve estimation, $F_m(v|s,z)$ is approximated using a multivariate Bernstein polynomial after normalizing v,s,z to have support in $[0,1]$, the domain of Bernstein polynomials.[26] For example, if z is a scalar,

[25] For completeness, the probability of an observation where no one bids is $\bar{s}_1^{N_1}\bar{s}_2^{N_2}$, and the probability that the lease sells at the reserve price to a bidder from group m is $N_m(1-\bar{s}_m)\bar{s}_m^{N_m-1}\bar{s}_{-m}^{N_{-m}}$. But since both of these probabilities only depend on $\{\bar{s}_m\}$, which are already estimated and do not depend on the object of estimation $F_m(\cdot|s,z)$, including or excluding these observations from the maximum likelihood estimator makes no difference in the result.

[26] These normalized values are only used inside the Bernstein polynomial; when performing all other calculations, true values are used.

$$F(v|s,z) = B_{m,n,l}(v,s,z)$$

$$\equiv \sum_{p=0}^{m}\sum_{q=0}^{n}\sum_{r=0}^{l}\alpha_{p,q,r}\begin{pmatrix} m \\ p \end{pmatrix} v^p(1-v)^{m-p}\begin{pmatrix} n \\ q \end{pmatrix} s^q(1-s)^{n-q}\begin{pmatrix} l \\ r \end{pmatrix} z^r(1-z)^{l-r}.$$

This approximation imposes a restriction that $F(v|s,z)$ be continuous not just in v but also in s and z. The polynomial degrees m, n, l, which are analogous to the bandwidths in kernel estimation, can be chosen to minimize a relevant criterion.[27]

Now the likelihood in (13) can be expressed in terms of polynomial $B(\cdot,\cdot,\cdot)$ by replacing H_m with $\hat{\bar{s}}_m + \int_{\hat{\bar{s}}_m}^{1} B(p,s,z)ds$ and h_m with $\int_{\hat{\bar{s}}_m}^{1} \frac{\partial B(p,s,z)}{\partial p}ds$. Finally, $F_m(v|s,z)$ can be estimated by finding the parameters $\alpha_{p,q,r}$ that maximize the log likelihood in (12). A benefit of using Bernstein polynomials is that they are easy to restrict to satisfy required properties of $F_m(v|s,z)$. Specifically, since $F_m(v|s,z)$ is a cdf, $B(v,s,z)$ should be weakly increasing in v, which means $\alpha_{p,q,r} \leq \alpha_{p',q,r}$ if $p < p'$. Test simulations of the estimator are discussed in Appendix B.

5.3 Estimation of utility functions and entry costs

For ease of notation in this section, consider a fixed z and x. I now construct the utility function that satisfies the first-order condition for bidding, nonparametrically. Let $J_m(b|\{\bar{s}_m\})$ be the distribution of submitted sealed bids conditional on entry thresholds $\{\bar{s}_m\}$, and let α indicate quantiles. Since bid functions are monotonic, a bidder whose value equals the α-quantile of values conditional on entry ($F_m^{*-1}(\alpha;\bar{s}_m)$) will bid an amount equal to the α-quantile of bids conditional on entry ($J_m^{-1}(\alpha|\{\bar{s}_m\})$). Then the first-order condition as stated in (10) can be rewritten in the following form that is useful for estimation of $\lambda(\cdot)$:

$$\lambda_m(F_m^{*-1}(\alpha;\bar{s}_m) - J_m^{-1}(\alpha|\{\bar{s}_m\})) = \frac{W_m(J_m^{-1}(\alpha|\{\bar{s}_m\})|\{\bar{s}_m\})}{dW_m(J_m^{-1}(\alpha|\{\bar{s}_m\})|\{\bar{s}_m\})/db}.$$

[27] The criterion used when analyzing the NMSLO data is discussed in Appendix B.

The estimator in section 5.2 yields $\hat{F}_m^*(\cdot;\bar{s}_m)$, the estimated value distribution conditional on entry. Then in order to construct $\lambda_m(\cdot)$, it remains to estimate the equation's other components, $J_m(\cdot|\{\bar{s}_m\})$ and $W_m(\cdot|\{\bar{s}_m\})$.

Submitted sealed bids are directly observed in the data. Meanwhile, $W_m(b|\{\bar{s}_m\})$ is the probability of winning when an entrant from subgroup m bids b in the S auction, conditional on $\{\bar{s}_m\}$. This is equivalent to the distribution of the highest competing bid facing a subgroup-m bidder, which is also observed in the data. So $J_m(b|\{\bar{s}_m\})$ and $W_m(b|\{\bar{s}_m\})$ are conditional distributions of variables that are directly observed in the data. Methods for estimating such objects are well known. I take a nonparametric approach, using bivariate Bernstein polynomials to approximate these conditional distributions.

Having thus estimated $\hat{F}_m^*(\cdot;\cdot)$, $\hat{J}_m(\cdot|\cdot)$, and $\hat{W}_m(\cdot|\cdot)$, I use the first-order condition above to construct $\hat{\lambda}_m(\cdot)$ as the function that maps $\hat{F}_m^{*-1}(\alpha;\bar{s}_m) - \hat{J}_m^{-1}(\alpha|\{\bar{s}_m\})$ to the right-hand side expression, for every quantile α.

Then, since $\lambda(x) \equiv \frac{\tilde{U}(x)}{\tilde{U}'(x)}$, I can compute $\tilde{U}(y) = exp \int_1^y 1/\hat{\lambda}(t)dt$; this computes $\tilde{U}(\cdot)$ to scale, with the scale normalization $\tilde{U}(1) = 1$. Other normalizations can be chosen as well; the scale is easily adjustable. Now, the expected stage 2 profit in (2) can be rewritten in terms of $\tilde{U}(\cdot)$ as $\pi_m(v_i, b) \equiv \tilde{U}_m(v_i - b)W_m(b) + U_m(-c_m)$. Then the expected profit from entering the auction in (5) can be rewritten as

$$\Pi_m(s_i) \equiv \int_{v=\underline{v}}^{\bar{v}} \tilde{U}_m(v - b^*(v))W_m(b^*(v))f_m(v|s_i)dv + U_m(-c_m).$$

Then from equilibrium entry condition (6) and the location normalization $U(0) = 0$,

$$U_m(-c_m) = -\int_{v=\underline{v}}^{\bar{v}} \tilde{U}_m(v - b^*(v))W_m(b^*(v))f_m(v|\bar{s}_m)dv.$$

All the objects needed to compute the right hand side have been estimated, so this provides a value for $U_m(-c_m)$. Now, by definition, $\tilde{U}_m(c_m) = U_m(c_m - c_m) - U_m(-c_m) = -U_m(-c_m)$. Hence, $c_m = \tilde{U}_m^{-1}(-U_m(-c_m))$. The right-hand side of this is known, providing an estimate of c_m, the entry cost in the S auction.[28] Finally, by definition of $\tilde{U}(\cdot)$, $U_m(w) = \tilde{U}_m(w +$

[28] $c_m(z, x)$ can be computed for any z, x by performing this computation for those values of z, x.

$c_m) + U_m(-c_m)$.

The entry cost in the O auction can be estimated by similarly solving for c_m^o using equilibrium entry condition (9), plugging in the estimated $\hat{F}_m(\cdot|\cdot)$, $\hat{U}_m(\cdot)$, and $\hat{\bar{s}}_m^o$. This completes the estimation of $F_m(\cdot|\cdot), U_m(\cdot)$, and c_m.

6 Estimation using NM auction data

In this section, I discuss estimation details and results that are specific to the NMSLO auction data. First, I check whether the pool of potential bidders is the same in the first-price sealed-bid (S) and English (O) auctions by comparing all bidder names observed in S to all winner names observed in O.[29] In auctions of Permian Basin leases in 2005-2014, 97% of S bids came from bidders that had also won in the O auctions, and 98% of bidders that won in the O auctions had also bid in the S auctions. The top 5 bidder names by number of wins are the same for both auction formats and in the same order. I conclude that the bidder pools underlying the two auction formats are mostly overlapping if not the same.

Next, I specify the estimation sample. The most common tract size is 320 acres, and acreage determines the reserve price. Since restricting my estimation to 320-acre tracts allows me to homogenize size and reserve price while still keeping the majority of data, I do so. I also drop outlier dates in which the supply of Permian Basin acreage is abnormally high or low, meaning greater than 15,000 acres or less than 5,000 acres where the SLO's target is 10,000 acres. This leads to dropping 8 out of 120 auction dates. As a result, the sample I use for estimation are all auctions of 320-acre leases in the Permian Basin in 2005-2014 on dates when Permian acreage stays between the bounds mentioned. There are 1059 S auctions and 935 O auctions in this sample.

I also define the subgroups of bidders. Following the discussion in section 2.4, I categorize the most frequent bidder (Yates) as the sole member of subgroup 2 and all other bidders as subgroup 1.

[29] Since only winners are observed in O.

6.1 Covariates z

With many covariates or lease characteristics z, nonparametric estimators can suffer from the curse of dimensionality, meaning there is not enough data to condition estimates on every combination of covariate values. In order to overcome this problem, I take a single index approach, in which the value distribution and entry cost depend on the vector z only through a scalar index $z'\beta$, i.e. $F_m(v|s,z) = F_m(v|s,z'\beta)$ and $c_m(z,x) = c_m(z'\beta,x)$. Since the entry threshold $\bar{s}_m(z,x)$ is determined by $F_m(v|s,z'\beta)$ and the entry cost $c_m(z'\beta,x)$ through (6), it follows that $\bar{s}_m(z,x) = \bar{s}_m(z'\beta,x)$. Furthermore, since sealed bids are determined by $F_m(v|s,z'\beta)$, $c_m(z'\beta,x)$, and $\bar{s}_m(z'\beta,x)$ through the first-order condition in (3), it follows that sealed bids are also dependent on z only through $z'\beta$. I estimate β by regressing the log of submitted sealed bids on observable lease characteristics z. Table 2 lists the lease characteristics considered. The ones marked with an asterisk are ultimately used to form the index $z'\beta$. Appendix B provides the rationale for this choice, a detailed explanation of the covariates, and results of regressing the log of submitted sealed bids on these covariates.

6.2 Choice of entry cost shifter x

Notation x refers to a variable that shifts the entry cost $c_m(z,x)$, and therefore the entry threshold, without affecting $F_m(v|s,z)$. In the NMSLO auction data, the amount of land offered for auction outside the Permian Basin is a candidate for x. Recall that 80% of leases auctioned by the NMSLO since 2005 are located in the Permian Basin. Of the remaining 20%, three fourths are in exploratory or "frontier" areas that have historically seen little to no drilling, and one fourth are in other established basins such as the San Juan basin.[30] Figure 3 shows the variation in acreage offered over time, excluding dates that have been dropped as described earlier. Unlike Permian Basin acreage, which the agency explicitly aims to keep at around 10,000 acres each month, acreage offered from frontier areas varies considerably from month to month, ranging from 0 to 32000 acres, as the agency does not

[30] The San Juan basin is located in the northwest corner of the state, as opposed to the Permian Basin in the southeast corner.

set a target. Frontier acreage on a given date is determined by the agency's capacity to review and approve land nominations prior to auction.

Bidders say that when there is a lot of acreage being auctioned on a single date, they have to "filter down to the ones they can spend more time analyzing." This suggests that the number of tracts a firm can analyze each month is finite, and bidders raise the bar on which tracts merit bidding as acreage offered increases. So variation in non-Permian acreage may be an entry cost shifter in this data. To explore this idea, regressions in Table 3 measure the relationship between acreage outside the Permian Basin and the number and size of submitted sealed bids inside the Permian Basin, controlling for covariates including oil and gas prices and year fixed effects.

In columns (1) and (2), the effect of non-Permian acreage on the number of bids submitted on each lease inside the Permian Basin is significant and negative, supporting the anecdotal evidence that it causes entry costs to increase and entry to decrease. Meanwhile, we may be concerned that increases in non-Permian supply also negatively affect bidders' values. While the exogeneity of this instrument cannot be tested, it is encouraging that columns (3) and (4) do not detect a negative effect of non-Permian supply on the level of bids (in log deflated dollars), controlling for the number of bids submitted. It seems plausible that non-Permian supply affects the entry threshold for but not the value of a Permian tract.

6.3 Data limitations

Number of potential bidders

While I observe the identities of all bidders in the S auction and all winners in the O auction for every auction, the number of potential bidders is not directly observed. The list of leases to be auctioned and their descriptions are published online for free, and there is no process for registering interest for a particular lease or auction date prior to bidding. One intuitive if imperfect measure of potential bidders for a Permian Basin lease is the number of unique bidder names $\tilde{N}$ observed in Permian Basin auctions that month. This is analogous to the approach used to measure potential bidders in Athey et al. (2011) and Li and Zheng (2012) among others. As the number of leases auctioned each month grows large, $\tilde{N}$ would converge to N. However, in this data the number of leases auctioned each month is not large enough to avoid an undesirable feature of $\tilde{N}$: even if the true N is fixed, variation in the quantity of auction items generates substantial variation in observed $\tilde{N}$. When each potential bidder enters with some probability, by construction the number of unique bidder names observed in the data increases with the number of leases auctioned.

With this in mind, I choose to model N_1 as constant in my sample and measure it as the maximum $\tilde{N}_1$ over all auction dates, which is 23. This is in line with staff remarks that there are usually 15-20 bidders present in the room on auction day, of which 2-3 typically bid on a given lease. Estimation results are quite robust to alternate choices of the constant value of N_1 since the entry threshold $\bar{s}_1$ scales accordingly such that the expected number of entrants, $N_1(1-\bar{s}_1)$, matches the data. For instance, the binomial distribution with $(N,p) = (10, 0.2)$ is in practice not very different from the binomial distribution with $(N,p) = (20, 0.1)$.[31] While it may be inaccurate to assume that N_1 is constant, letting N_1 vary as $\tilde{N}$ is likely to introduce more serious bias given how $\tilde{N}$ is measured. Meanwhile, since there is only one firm in subgroup 2, $N_2 = 1$.

Entry thresholds for the English (O) auction

As shown in Proposition 1, $\bar{s}_m(z,x)$ for the O auctions satisfies

$$\bar{s}_m(z,x) = \frac{\text{P(no one enters}|z,x)}{\text{P(winner's subgroup}=m \text{ \& winning bid}=r|z,x)/N_m + \text{P(no one enters}|z,x)}.$$

Computing this expression requires that the probability of no entrants and probability of reserve price sales to each subgroup be estimated conditional on z and x. In this data, auctions with zero entrants and reserve price sales certainly occur, but they do not occur frequently enough to allow the estimation of their probability conditional on z and x. 103 out of 120 auction dates have gone by without a single lease being unsold in the O auction

[31] Plotting the two distributions confirms this visually.

sample.

In Table 4, I use matched leases to compare entry in the S and O auctions. Each S lease was matched to an O lease located in the same section[32] and auctioned on the same day; unmatched leases are not included in the Table. The number of entrants is not recorded in the O auction, but I know when there are no entrants (the lease goes unsold) and when there is one entrant (the lease sells at the reserve price). I count the occurrence of each of these events in the O auction and compare to the counts in S. A significant difference in entry would manifest itself in these numbers; if entry rates are higher, zero-bidder and one-bidder events will be less frequent.

Overall entry rates appear similar in S and O auctions of comparable tracts. Due to the data limitations described above, I use the $\bar{s}_m(z, x)$ computed from S data as the entry threshold for both auction types. Post-estimation, I can subject the assumption of similar thresholds to a test by computing the expected profit from entering the O auction for a marginal entrant using the entry threshold estimated from S data. According to the equilibrium entry conditions in (9), the expected profit should be zero for a marginal entrant at the true equilibrium threshold. If S entry thresholds well approximate O entry thresholds, the test computation described should yield a value close to zero.[33]

6.4 Other modelling choices and estimation details

Since Yates, the bidder in subgroup 2, is not very selective in bidding (it bid in 84% of all S auctions of 320-acre Permian Basin leases in 2005-2014), I model subgroup 2 as entering nonselectively with probability 0.84 to reduce the burden on the estimator. This way, the value distribution to be estimated for subgroup 2 is not conditional on the signal s. The entry model for subgroup 1 is still fully endogenous and selective. In other words, the value distributions to be estimated are $F_1(v|s, z'\beta)$ and $F_2(v|z'\beta)$.

[32] A one square-mile area.

[33] I compute this post estimation and find that for a typical tract (i.e. a tract with modal $z'\beta$ and x), the expected profit for a marginal entrant in O is roughly \$160. Locally, $\Pi_1^o(\bar{s}_1^{o*}|\{\bar{s}_m^{o*}\})$ (see (9)) changes by about that much for every 0.001 change in $\bar{s}_1^{o*}$. This means the $\bar{s}_1$ computed from S data is very close to (within about 0.001 of) the $\bar{s}_1^{o*}$ that satisfies $\Pi_1^o(\bar{s}_1^{o*}|\{\bar{s}_m^{o*}\}) = 0$.

Following the discussion in section 5.1, the S entry threshold for subgroup 1 is $\bar{s}_1(z'\beta, x) = 1 - \frac{\mathbb{E}[n_1|z'\beta,x]}{N_1}$. I estimate $\mathbb{E}[n_1|z'\beta, x]$, the expected number of entrants conditional on $z'\beta$ and x, parametrically using a binomial logistic regression where

$$n_1 \sim \mathcal{B}\left(N_1, \frac{1}{1 + e^{-(\gamma_0+\gamma_1 x+\gamma_2 z'\beta+\gamma_3(z'\beta)^2+\gamma_4(z'\beta)^3)}}\right). \tag{14}$$

$\mathcal{B}(\cdot,\cdot)$ stands for the binomial distribution. Squared and cubed terms of $z'\beta$ are included to allow for flexible forms of the relationship between $z'\beta$ and $\bar{s}_1$. The results of that estimation are displayed in Appendix B Table 12.

When it comes to estimating the utility function $U_m(\cdot)$, any observed values of $z'\beta$ and x can be used to perform the estimation procedure described in section 5.3 since $U_m(\cdot)$ is independent of z and x. Since estimates of $F^*_m(v; \bar{s}_1|z'\beta)$ and $J_m(b|\bar{s}_1, z'\beta)$ are likely to be most precise where the data is dense, I use $\hat{F}^*_m(v; \bar{s}_1|z'\beta)$ and $\hat{J}_m(b|\bar{s}_1, z'\beta)$ at modal $z'\beta$ and x to estimate $U_m(\cdot)$.

6.5 Estimation results

The estimated value distributions are depicted in Figure 4. Recall from section 4.1 that $F_1(v|s, z'\beta)$ is identified for the signals s that fall in the range of entry thresholds $\bar{s}_1(z'\beta, x)$ available in the data. At modal $z'\beta$, this range is roughly [0.922, 0.944] given (14) and the range of variation in the entry cost shifter x. It spans a 40% change in the probability of entry for each potential bidder from 0.056 to 0.078. In Figure 4, the left panel depicts $\hat{F}_1(v|s, z'\beta)$ for signals s at the bottom and top of this identified range, along with $\hat{F}_2(v|z'\beta)$, at modal $z'\beta$. Comparing subgroup 1 and subgroup 2, there does not seem to be a clear dominance relation of one over the other, but subgroup 2 has a higher median value. Meanwhile, the right panel zooms in to show that higher signals are associated with stochastically dominant value distributions. This dominance was not imposed in estimation and is consistent with selective entry. When the entry threshold is 0.922, the marginal entrant has a median value that is 40% lower than that of entrants overall. When the entry threshold rises, these marginal entrants will be the first to exit, and the value distribution of remaining entrants will stochastically dominate those of the exiters/non-entrants. Going forward, I use estimates imposing stochastic dominance to avoid any crossing in the tails where data is sparse.

Figure 5 depicts the estimated nonparametric utility functions of the two subgroups. Both functions seem to display some risk aversion, corroborating the discussion in section 2.3. Subgroup 1 bidders appear more risk-averse than subgroup 2. To learn the form of risk aversion that these utility functions entail, I fit the Hyperbolic Absolute Risk Aversion (HARA) utility form to $\hat{U}_m(\cdot)$. HARA utility, under the condition $U(0) = 0$, is

$$U(x) = \kappa \frac{1-\alpha}{\alpha} \left[\left(\frac{x}{1-\alpha} + \beta \right)^{\alpha} - \beta^{\alpha} \right]$$

and includes risk neutrality, constant absolute risk aversion (CARA), constant relative risk aversion (CRRA), DARA, DRRA, IARA, and IRRA as special cases. Table 5 displays the fitted HARA parameters.

The coefficient $\hat{\alpha} < 1$ indicates DARA for both utility functions. This is reassuring since DARA is considered more empirically plausible than IARA. The coefficient $\hat{\beta} > 0$ indicates IRRA for $\hat{U}_1(\cdot)$, while $\hat{\beta} = 0$ indicates CRRA for $\hat{U}_2(\cdot)$. κ is just a constant multiple that scales the HARA fit to match the estimated utility functions. Meanwhile, I also fit the CRRA utility function to $\hat{U}_1(\cdot)$ and $\hat{U}_2(\cdot)$ to compare the extent of risk aversion here with estimates in the literature. This yields CRRA parameters of 0.48 and 0.23, respectively. As a comparison, Holt and Laury (2002) measure CRRA parameters centered around the 0.3-0.5 range in laboratory experiments, and Lu and Perrigne (2008) measure roughly 0.59 for the USFS timber auctions.[34]

As for the estimated entry cost, $\hat{c}_1(z'\beta, x)$ is roughly \$11,000 at modal $z'\beta$ with no non-Permian acreage ($x = 0$). This is the minimum expected profit below which a firm will not bother analyzing a tract, as the opportunity cost of doing so is larger than the expected profit. For sake of comparison, the median S auction bid for a 320-acre lease is roughly \$39,500 in this data, and the mean bid is \$101,100. When the entry-cost shifter x increases to 30,000 acres (roughly the maximum observed in the data), the estimated entry cost increases to

[34] A CRRA parameter of 0 indicates risk neutrality, while 1 indicates log utility.

$28,000.

7 Counterfactual simulations

I use the estimated value distributions, utility functions, and entry costs to perform counterfactual simulations answering the following questions. First, I ask how the entry process affects current revenue. Second, I quantify the difference selective entry makes (versus non-selective entry) in determining the revenue effect of policies lowering the entry threshold. Finally, I assess the effectiveness of policies that intend to improve competition by expanding the potential-bidder pool. The method used to simulate counterfactual S bid functions is described in Appendix B.

7.1 Entry-induced uncertainty and risk aversion

Fixing lease quality $z'\beta$ and the entry-cost shifter x at modal values, I simulate auction revenue under alternative auction models to learn the implications of the model studied here. As a reference, Figure 6 shows the simulated probability distribution of the number of entrants n. Most auctions have 1-4 entrants, and close to half of all auctions have 2 or fewer entrants; this is a low competition environment. Figure 7 then shows simulated auction revenue, broken out by the realized number of entrants n. The dark blue bars simulate the English (O) auction; the green bars simulate the first-price sealed-bid (S) auction as most commonly modeled, where risk-neutral bidders know n when they bid; and finally, the yellow bars simulate the S auction model postulated in this paper, where risk-averse bidders do not know n when they bid.[35] The red dotted line marks the value of the item to the median bidder conditional on entry.

The dark blue and green bars confirm that a low number of entrants is very damaging to auction revenue; cases with just one entrant are particularly devastating. In light of this, the yellow bars are intriguing. In the case of 1-3 entrants, the yellow bars display a

[35] The dark blue bars and green bars need not be the same because bidders are asymmetric; revenue equivalence does not apply.

large revenue boost similar in size to having one additional entrant in the green bars. Even when there is just one entrant, revenue comes close to the red line, consistent with evidence discussed in section 2.2. Meanwhile at very high realizations of n, the yellow bars eventually become lower than the green bars, but the difference is quite small because risk aversion lifts S bids across the board. In total, Figure 7 paints a picture in which entry-induced uncertainty combined with risk aversion provides meaningful protection against the effects of low competition. This protection applies to the S auction but not to the O auction, where bidders have a dominant strategy for bidding unaffected by the number of entrants or risk aversion. Therefore, there is good reason to favor first-price sealed-bid auctions over English auctions in low competition environments, even absent traditional concerns like collusion.

The overall simulated difference in log revenue between the dark blue (O) and yellow bars, accounting for the probability distribution of n as shown in Figure 6, is 0.36. This is consistent with the revenue pattern seen in Table 1. If expected prices are higher in the S auction as they are here, does this imply entry rates should be lower? It turns out this is not the case. The expected profit from entry - and hence the entry rate - is determined not only by the auction price but also the probability of winning and, for risk-averse bidders, uncertainty regarding the price. Simulations show that subgroup 1 bidders have a higher probability of winning in S than in O, conditional on the same value.[36] An increased probability of winning makes up for a more than proportionate reduction in price-value margins when bidders are risk-averse. Also, the price paid upon winning is naturally unpredictable in O, while it is known with certainty in S to be one's own bid. On balance, available evidence (Table 4) and computational checks[37] point to similar entry rates for S and O in this data.[38]

[36] This is akin to winning patterns observed for small bidders (loggers) against large bidders (mills) in Athey et al. (2011).

[37] I compute the expected profit from entering the O auction for a marginal entrant using the entry threshold estimated from S data, and confirm this is close to zero. That is, for a typical tract (i.e. a tract with modal $z'\beta$ and x), the expected profit for a marginal entrant in O is roughly \$160. Locally, $\Pi_1^o(\bar{s}_1^{o*}|\{\bar{s}_m^{o*}\})$ (see (9)) changes by about that much for every 0.001 change in $\bar{s}_1^{o*}$. This means the $\bar{s}_1$ computed from S data is very close to (within about 0.001 of) the $\bar{s}_1^{o*}$ that satisfies $\Pi_1^o(\bar{s}_1^{o*}|\{\bar{s}_m^{o*}\}) = 0$.

[38] Li et al. (2015) show that if bidders are symmetric and risk aversion takes the form of DARA, CARA, or IARA, DARA leads to higher entry rates in O, CARA leads to equal entry rates, and IARA leads to higher entry rates in S. When bidders are asymmetric this need not apply, one reason being that winning probabilities differ between formats as discussed above.

Finally, the empirical evidence here demonstrate in magnitude how important entry and risk aversion can be in determining revenue, but I qualify that this magnitude will naturally differ from setting to setting. Also, it seems at least theoretically possible that if auction prices are higher in S than in O, the entry rates in S could fall so much that revenue ends up being lower. So the revenue ranking is not absolute, though Li et al. (2015) comment that they were unable to find theoretical examples yielding lower revenue in S given symmetric, risk-averse bidders and selective entry.

7.2 Revenue response to changes in the entry threshold

Low numbers of entrants are often a concern in government auctions generally. In response, governments may pursue policies that try to increase the entry rate, or lower the entry threshold.[39] I quantify the revenue response under selective entry versus nonselective entry when such a policy is pursued. Taking $\bar{s}_1 = 0.943$ as a starting point, I simulate the O revenue change when $\bar{s}_1$ falls to 0.922 in Table 6.[40] Under nonselective entry, a lower entry threshold means more entrants that are just like existing entrants, which would lead to a 34% increase in revenue. However, revenue under selective entry is not as sensitive to the entry threshold, because a lower entry threshold means gaining entrants that are stochastically dominated by existing entrants in value. The selective entry model predicts a 27% revenue increase, compared to the 34% predicted by a nonselective entry model.

7.3 Increasing the number of potential bidders

If there are policies that can increase the number of potential bidders, how effective might they be at increasing revenue? The NMSLO is considering a trial of online auctions with the idea that this would widen the bidder pool. Li and Zheng (2009) show that when entry is endogenous, revenue may or may not increase with the number of potential bidders; the

[39]The entry threshold is one minus the entry rate.

[40][0.922, 0.943] is the approximate range of signals for which $F_1(v|s, z'\beta)$ is identified given the variation of the entry-cost shifter. Results for the S auction are qualitatively similar, as shown in Table 13 of Appendix B.

"competition effect" of having more bidders may be dominated by the "entry effect" of each bidder entering with lower probability. I explore whether this policy will have its intended effect by simulating a counterfactual increase in the number of potential bidders.

In the last row of Table 7, I display simulated counterfactual revenue for a 50% increase in N_1, the number of potential bidders in subgroup 1, from 23 to 34. This row includes both the competition effect and the entry effect. I compare this to the competition-effect-only or exogenous entry scenario where the number of entrants increases in proportion to the number of potential bidders (second row). Judging from the numbers, I conclude that revenue will indeed increase with N_1, but the increase will be smaller than what we could expect under an exogenous entry rate and proportionally smaller than the expansion in N_1.

8 Conclusion

This paper performs a structural analysis of selective entry in an empirical auction market. It takes a nonparametric approach that flexibly estimates the nature of selection in entry. Also, it features a unified auction model allowing for bidders' risk aversion and asymmetry to explore not only the implications of selective entry alone but also its implications in conjunction with these other properties. The estimation procedure uses sieve maximum likelihood to estimate bidders' value distributions, followed by constructive steps to estimate their utility functions and entry costs. The procedure is supported by a step-by-step identification argument, including identification for English auctions with entry where neither losing bids nor number of entrants is observed.

The paper presents both empirical and simulated evidence of selective entry and its effect on auction revenue. A key finding is that, in first-price sealed-bid auctions, entry-induced uncertainty combined with risk aversion tends to moderate the damage from a low number of entrants, often bolstering revenue by an amount similar to having one additional entrant. This effect does not apply to English auctions. An immediate policy implication is to favor first-price sealed-bid auctions over English auctions in low-entry environments.

Appendix A

Proof of Proposition 1 Nowik (1990) looks at a machine made up of N components, each of which have a lifetime X_i with distribution $F_i(\cdot)$. The machine fails when certain designated subsets of the components all fail; for instance, when k out of the N components fail. The statistician observes the lifetime of the machine, the number of total components N, and the set of components that had failed by the time the machine failed. Novik proves that if the X_i are independent, $F_i(\cdot)$ are mutually absolutely continuous (there are no constant sections), each $F_i(\cdot)$ possesses a single positive atom at the common essential infimum, and there is at most one life-supporting component (i.e. there is at most one component such that the machine cannot die without that component dying), then all distributions $F_i(\cdot)$ are identified.

The English auction model here can be viewed as just such a reliability theory problem: the auction stops ("dies") when $N-1$ out of N private values are exceeded. The case of no entrants is analogous to the case where all N machine components immediately failed. The conditions of Nowik's theorem are satisfied here: (V, S) are independent, $H_m(v)$ is without constant sections and possesses an atoms at $\underline{v}$, and there are no "life-supporting components," since any bidder has positive probability of winning. Hence the $H_m(v)$ are identified by Nowik's theorem.

The entry thresholds $\bar{s}_m$ are identified by solving for the two unknowns $\bar{s}_m, \bar{s}_{-m}$ in the following system of two equations:

$$\begin{aligned}
&\text{P(no one enters)} = \bar{s}_m^{N_m} \bar{s}_{-m}^{N_{-m}} \\
&\text{P(winner's subgroup} = m \,\&\text{ winning bid} = r) \\
&= \text{P(exactly one bidder entered and his subgroup is } m) = N_m(1-\bar{s}_m)\bar{s}_m^{N_m-1}\bar{s}_{-m}^{N_{-m}}.
\end{aligned} \tag{15}$$

The left-hand side probabilities in system (15) are observed in the data. Then the solution to the system can be solved for as follows:

$$\begin{aligned}
&\frac{\text{P(no one enters)}}{\text{P(winner's subgroup} = m \,\&\text{ winning bid} = r)/N_m + \text{P(no one enters)}} \\
&= \frac{\bar{s}_m^{N_m} \bar{s}_{-m}^{N_{-m}}}{(1-\bar{s}_m)\bar{s}_m^{N_m-1}\bar{s}_{-m}^{N_{-m}} + \bar{s}_m^{N_m}\bar{s}_{-m}^{N_{-m}}} = \frac{\bar{s}_m}{1-\bar{s}_m+\bar{s}_m} = \bar{s}_m.
\end{aligned} \tag{16}$$

So the entry thresholds can be computed from observed probabilities, and then $F_m^*(v;\bar{s}_m)$ is identified as $[H_m(v;\bar{s}_m) - \bar{s}_m]/(1-\bar{s}_m)$.

Proof of Proposition 4 Entry thresholds $\bar{s}_m$ can be directly identified separately for the O and S formats as described before. $F_m^*(v;\bar{s}_m)$, the value distribution conditional on entry at threshold $\bar{s}_m$, is identified from the O data, as explained in Proposition 1. Let $J_m(b|\{\bar{s}_m\})$ be the observed distribution of S auction bids conditional on entry at threshold $\bar{s}_m$. Then, analogously to Lu and Perrigne (2008) Proposition 1 - but now conditioning on entry thresholds to account for selective entry - $U_m(\cdot)$ are identified from the S data as follows.

Let $v(\alpha)$ and $b(\alpha)$ be the α-quantiles of $F^*_m(v;\bar{s}_m)$ and $J_m(b|\{\bar{s}_m\})$, respectively; that is, $v(\alpha) \equiv F^{*-1}_m(\alpha;\bar{s}_m)$ and $b(\alpha) \equiv J^{-1}_m(\alpha|\{\bar{s}_m\})$. Then, since bid functions are monotonic, (11) becomes

$$F^{*-1}_m(\alpha;\bar{s}_m) = J^{-1}_m(\alpha|\{\bar{s}_m\}) + \lambda^{-1}_m\left(\frac{W_m(J^{-1}_m(\alpha|\{\bar{s}_m\})|\{\bar{s}_m\})}{dW_m(J^{-1}_m(\alpha|\{\bar{s}_m\})|\{\bar{s}_m\})/db}\right).$$

Hence,

$$\lambda_m(F^{*-1}_m(\alpha;\bar{s}_m) - J^{-1}_m(\alpha|\{\bar{s}_m\})) = \frac{W_m(J^{-1}_m(\alpha|\{\bar{s}_m\})|\{\bar{s}_m\})}{dW_m(J^{-1}_m(\alpha|\{\bar{s}_m\})|\{\bar{s}_m\})/db}, \tag{17}$$

where $J_m(\cdot)$, and $W_m(\cdot)$ are observed in the S data, and $F^*_m(\cdot)$ is identified from the O data. So $\lambda_m(\cdot)$ is identified using (17).

Given A6 and A7, $F_m(\cdot|\cdot)$ is identified from the O auction. As for c_m of each auction format, they are uniquely determined from the respective equilibrium entry conditions given $F_m(\cdot|\cdot)$ and $\bar{s}_m$ for any $U_m(\cdot)$, because $\Pi_m(\bar{s}_m|\{\bar{s}_m\})$ and $\Pi^o_m(\bar{s}_m|\{\bar{s}_m\})$ are monotonic in c_m and c^o_m respectively.

Now it remains to identify $U_m(\cdot)$ from $\lambda_m(\cdot)$. By definition, $\lambda_m(x) \equiv \frac{\tilde{U}_m(x)}{\tilde{U}'_m(x)}$. With a scale normalization such as $\tilde{U}_m(1) = 1$, $\tilde{U}_m(x) = exp \int_1^x 1/\lambda_m(t)dt$. And given the location normalization $U_m(0) = 0$, $U_m(x) = \tilde{U}_m(x + c_m) - \tilde{U}_m(c_m)$.

References

Ackerberg, D., K. Hirano, and Q. Shahriar (2006): "The Buy-it-now Option, Risk Aversion, and Impatience in an Empirical Model of eBay Bidding," http://www.econ.ucla.edu/ackerber/.

Athey, S. and P. A. Haile (2002): "Identification of Standard Auction Models," *Econometrica*, 70, 2107–2140.

Athey, S., J. Levin, and E. Seira (2011): "Comparing Open and Sealed Bid Auctions: Evidence from Timber Auctions," *The Quarterly Journal of Economics*, 126, 207–257.

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

——— (2005): "Are Structural Estimates of Auction Models Reasonable? Evidence from Experimental Data," *Journal of Political Economy*, 113, 703–741.

Bhattacharya, V., J. W. Roberts, and A. Sweeting (2014): "Regulating Bidder Participation in Auctions," *The RAND Journal of Economics*, 45, 675–704.

Bulow, J. and P. Klemperer (1996): "Auctions Versus Negotiations," *The American Economic Review*, 86, 180–194.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," *Handbook of Econometrics*, 6, 5549–5632.

Cox, J. C., B. Roberson, and V. L. Smith (1982): "Theory and Behavior of Single Object Auctions," *Research in Experimental Economics*, 2, 1–43.

Cox, J. C., V. L. Smith, and J. M. Walker (1983): "Tests of a Heterogeneous Bidders Theory of First Price Auctions," *Economics Letters*, 12, 207–212.

Gentry, M. and T. Li (2014): "Identification in Auctions with Selective Entry," *Econometrica*, 82, 315–344.

Gentry, M. L., T. Li, and J. Lu (2015): "Identification and Estimation in First-Price Auctions with Risk-Averse Bidders and Selective Entry," Available at SSRN 2587812.

Guerre, E., I. Perrigne, and Q. Vuong (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525–574.

——— (2009): "Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions," *Econometrica*, 77, 1193–1227.

Holt, C. A. and S. K. Laury (2002): "Risk Aversion and Incentive Effects," *The American Economic Review*, 92, 1644–1655.

Hu, Y., D. McAdams, and M. Shum (2013): "Identification of first-price auctions with non-separable unobserved heterogeneity," *Journal of Econometrics*, 174, 186–193.

Komarova, T. (2013): "A New Approach to Identifying Generalized Competing Risks Models with Application to Second-Price Auctions," *Quantitative Economics*, 4, 269–328.

——— (2017): "Extremum Sieve Estimation in k-out-of-n Systems," *Communications in Statistics - Theory and Methods*, 46, 4915–4931.

Krasnokutskaya, E. (2011): "Identification and Estimation of Auction Models with Unobserved Heterogeneity," *The Review of Economic Studies*, 78, 293–327.

Krasnokutskaya, E. and K. Seim (2011): "Bid Preference Programs and Participation in Highway Procurement Auctions," *The American Economic Review*, 101, 2653–2686.

Krishna, V. (2010): *Auction Theory*, Academic Press, second ed.

Lebrun, B. (1999): "First Price Auctions in the Asymmetric N Bidder Case," *International Economic Review*, 40, 125–142.

Levin, D. and J. L. Smith (1994): "Equilibrium in Auctions with Entry," *The American Economic Review*, 84, 585–599.

Li, T., J. Lu, and L. Zhao (2015): "Auctions with Selective Entry and Risk Averse Bidders: Theory and Evidence," *The RAND Journal of Economics*, 46, 524–545.

Li, T. and X. Zheng (2009): "Entry and Competition Effects in First-Price Auctions: Theory and Evidence from Procurement Auctions," *The Review of Economic Studies*, 76, 1397–1429.

——— (2012): "Information Acquisition and/or Bid Preparation: A Structural Analysis of Entry and Bidding in Timber Sale Auctions," *Journal of Econometrics*, 168, 29–46.

Lu, J. and I. Perrigne (2008): "Estimating Risk Aversion from Ascending and Sealed-Bid Auctions: The Case of Timber Auction Data," *Journal of Applied Econometrics*, 23, 871–896.

Marmer, V., A. Shneyerov, and P. Xu (2013): "What Model for Entry in First-Price Auctions? A Nonparametric Approach," *Journal of Econometrics*, 176, 46–58.

Maskin, E. and J. Riley (2000a): "Asymmetric Auctions," *The Review of Economic Studies*, 67, 413–438.

——— (2000b): "Equilibrium in Sealed High Bid Auctions," *The Review of Economic Studies*, 67, 439–454.

Matthews, S. (1987): "Comparing Auctions for Risk Averse Buyers: A Buyer's Point of View," *Econometrica*, 55, 633–646.

McAfee, R. P. and J. McMillan (1987): "Auctions with a Stochastic Number of Bidders," *Journal of Economic Theory*, 43, 1–19.

Mead, W. J. (1967): "Natural Resource Disposal Policy - Oral Auction Versus Sealed Bids," *Natural Resources Journal*, 7, 194–224.

Meilijson, I. (1981): "Estimation of the Lifetime Distribution of the Parts from the Autopsy Statistics of the Machine," *Journal of Applied Probability*, 18, 829–838.

Milgrom, P. R. and R. J. Weber (1982): "A Theory of Auctions and Competitive Bidding," *Econometrica*, 50, 1089–1122.

Nowik, S. (1990): "Identifiability Problems in Coherent Systems," *Journal of Applied Probability*, 27, 862–872.

Reny, P. J. and S. Zamir (2004): "On the Existence of Pure Strategy Monotone Equilibria in Asymmetric First-Price Auctions," *Econometrica*, 72, 1105–1125.

Roberts, J. W. and A. Sweeting (2013): "When Should Sellers Use Auctions?" *The American Economic Review*, 103, 1830–1861.

——— (2016): "Bailouts and the Preservation of Competition: The Case of the Federal Timber Contract Payment Modification Act," *American Economic Journal: Microeconomics*, 8, 257–88.

Samuelson, W. F. (1985): "Competitive Bidding with Entry Costs," *Economics Letters*, 17, 53–57.

Sweeting, A. and V. Bhattacharya (2015): "Selective Entry and Auction Design," *International Journal of Industrial Organization*, 43, 189–207.

Figures and Tables

Figure 1: Histogram of number of bids received in S auctions, Permian Basin 2005-2014

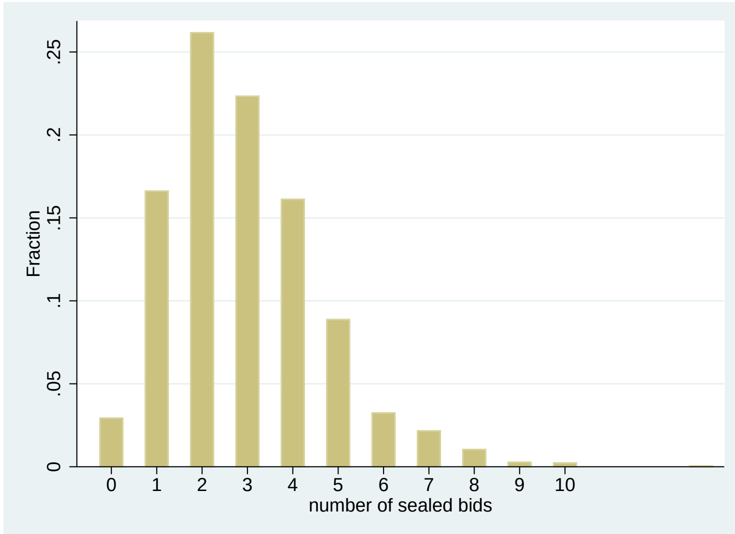
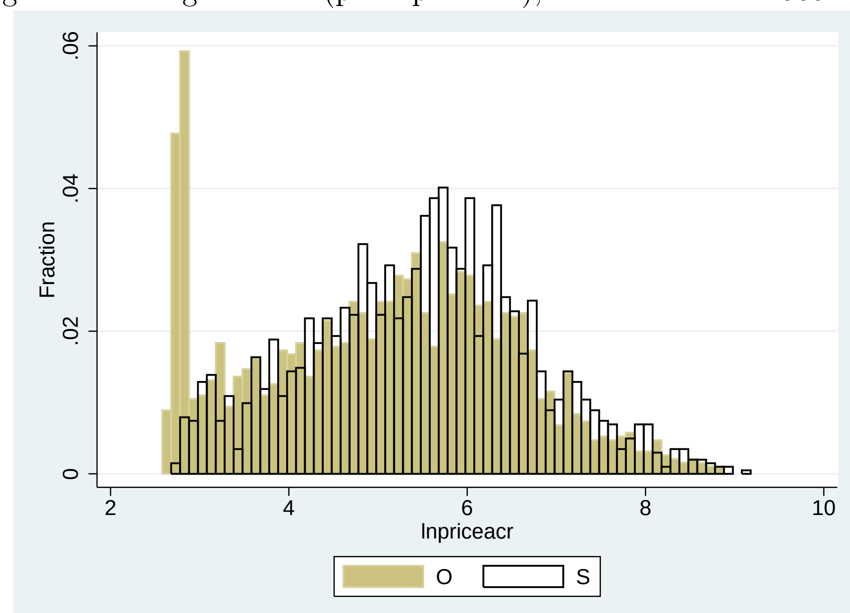


Figure 2: Histogram of ln(price per acre), Permian Basin 2005-2014



Prices are in 2009 dollars, deflated by the GDP implicit price deflator.

Table 1: Auction format and auction price, Permian Basin 2005-2014

	ln(price/acre)
auction format S	0.370
	(0.035)
lease prefix VB ("premium")	0.241
	(0.045)
tract size (acres)	-0.001
	(0.000)
drilled before	-0.026
	(0.040)
ln(production†) 1970-auction date	-0.003
	(0.006)
Township†† FE	Y
Polynomial of geographic coordinates†††	Y
Date FE	Y
Observations	4202
R^2	0.508
Adjusted R^2	0.447

† Oil and gas in barrel of oil equivalents (BOE).

†† A township is a 6 by 6 square mile area.

††† Fourth order polynomial

Heteroskedasticity robust standard errors in parentheses

Table 2: List of covariates z

Contract terms of lease:
Lease prefix V0 or VB*
Time variables:
WTI oil prices, monthly
Natural gas 1 month futures*
Year fixed effects*
Location variables:
Location quartile dummies*
Distance to high-value center relative to own township (for upper quartiles)*
Dummy for having been drilled in the past
Volume of oil produced on tract, 1970-auction date

Figure 3: Acreage offered inside and outside the Permian Basin

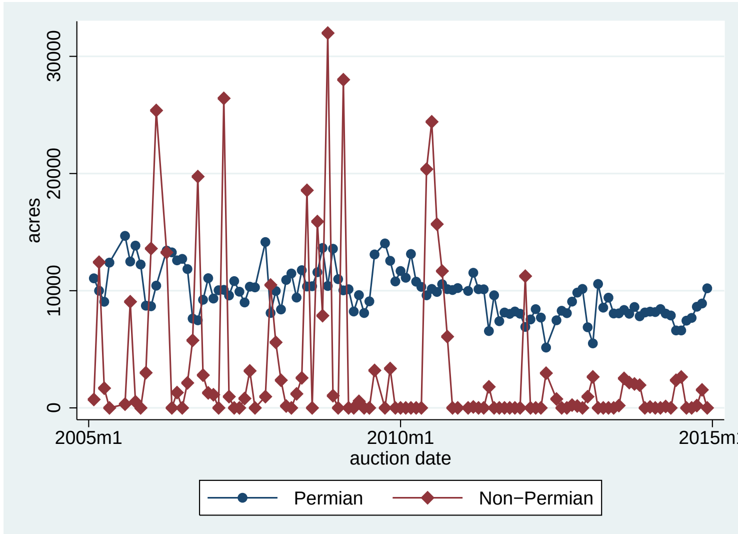


Table 3: Non-Permian acreage and Permian bidding

	(1) GLM† numbids	(2) GLM numbids	(3) OLS lnbid	(4) OLS lnbid
Non P. Basin acreage (1000s)	-0.007	-0.011	0.003	0.003
	(0.003)	(0.003)	(0.003)	(0.003)
Auction covariates††	N	Y	N	Y
Number of bids	-	-	Y	Y
Observations	1039	1039	3083	3083

† Binomial logistic regression. Dependent variable is number of bids.

†† Includes all covariates listed in Table 2.

Heteroskedasticity robust standard errors in parentheses

Table 4: Comparing entry statistics between matched S and O leases

	S	O
Count: 0 bidders	10/432	12/432
Count: 1 bidder	66/432	66/432

Figure 4: $\hat{F}_1(v|s, z'\beta)$ and $\hat{F}_2(v|z'\beta)$ at modal $z'\beta$

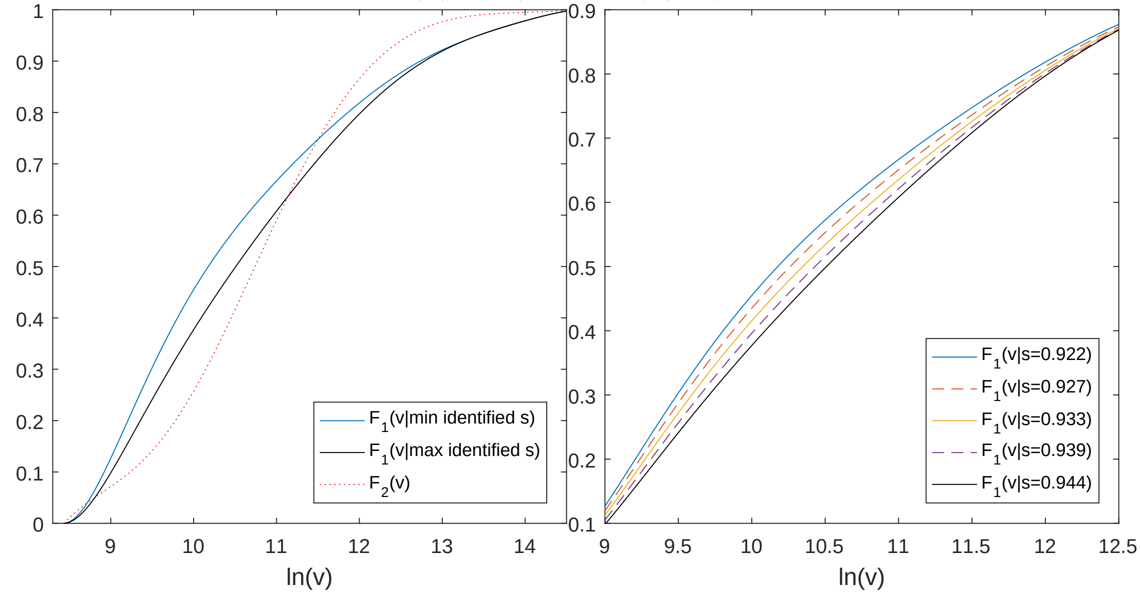


Figure 5: $\hat{U}_1(\cdot)$ (left) and $\hat{U}_2(\cdot)$ (right)

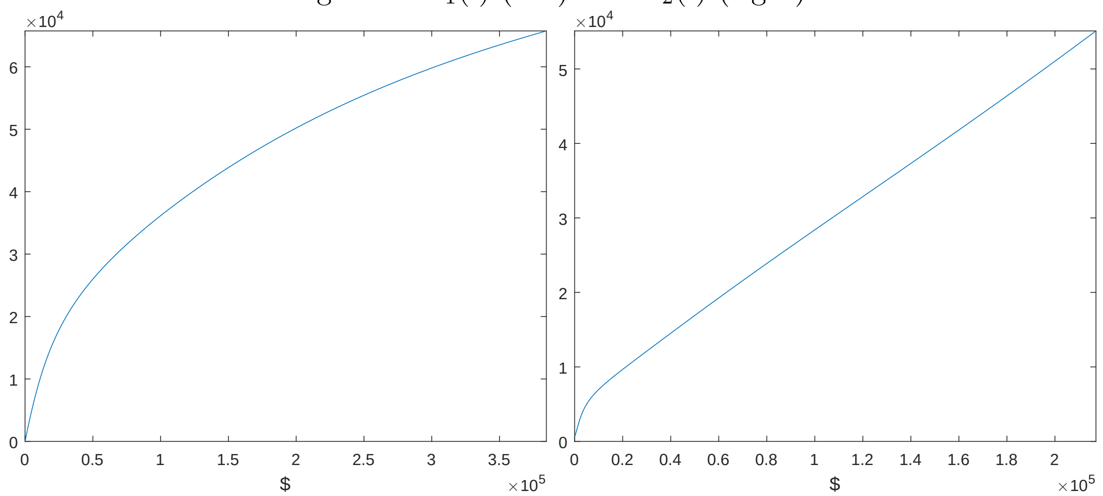


Table 5: HARA utility fitted parameter estimates

	$\hat{U}_1(\cdot)$	$\hat{U}_2(\cdot)$
$\hat{\alpha}$	0.18	0.77
$\hat{\beta}$	12152	0
$\hat{\kappa}$	2893	4.6

Figure 6: Simulated probability distribution of n at $\bar{s}_1 = \bar{s}_1(z'\beta, x)$

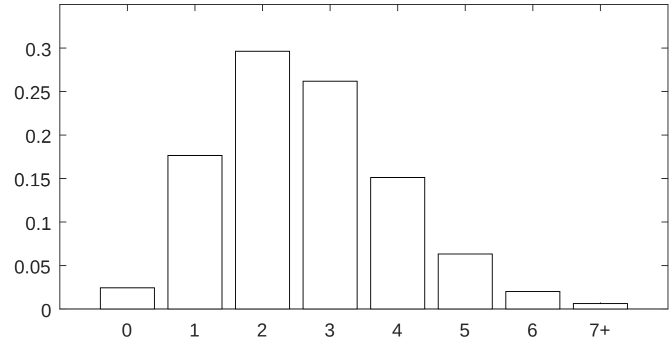


Figure 7: Simulated revenue by realized n

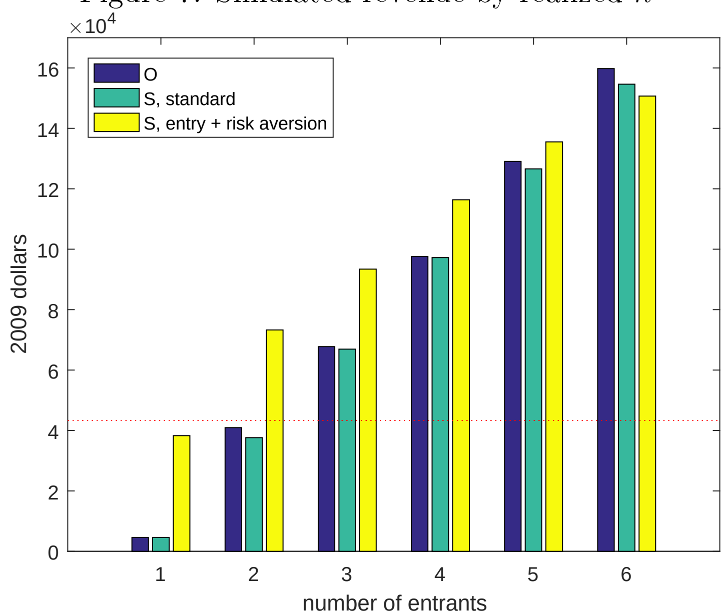


Table 6: Simulated revenue response to drop in entry threshold, at modal $z'\beta$

	Selective Entry		Nonselective Entry	
$\bar{s}_1$	E[O price]	%Δ (logΔ)	E[O price]	%Δ (logΔ)
0.943	46,085	-	46,085	-
0.922	58,419	27% (0.24)	61,751	34% (0.29)

Table 7: Counterfactual revenue with 50% increase in N_1, at modal $z'\beta$ and x

	E[S price]	E[O price]
Current	83,973	58,419
50% larger N_1, proportional increase in entrants	124,214	83,859
50% larger N_1, selective entry model	109,298	74,475

For Online Publication: Appendix B

Assignment of auction items to S and O auction formats I regress the log of post-auction oil and gas production up to December 2014 on auction format, controlling for the lease prefix and the number of months passed since the auction date. To allow auctioned tracts time to start producing, I repeat the same regression in the second column only using leases for which the 5 year lease term has passed (i.e. leases auctioned earlier than 2010). There is no relationship between auction format and realized production.

Table 8: Auction format and production, Permian Basin 2005-2014

	(1) log production†	(2) log production† 5 year+
auctionmethod S	-0.039	-0.076
	(0.115)	(0.187)
lease prefix VB	0.718***	0.952***
	(0.121)	(0.188)
months passed since auction by Dec 2014	0.022***	0.028***
	(0.002)	(0.005)
Constant	-0.530***	-1.103**
	(0.096)	(0.457)
Observations	4202	2409

Heteroskedasticity robust standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

† Production of oil and gas between auction date and Dec 2014, in barrel of oil equivalents.

Proof of Proposition 2 $U_m(\cdot)$ and c_m are identified if there is a unique utility function and entry cost for each subgroup that satisfy the subgroup's equilibrium entry condition in (9), given the $F_m(\cdot|\cdot)$ and entry thresholds $\bar{s}_m$ identified from the O data as described in section 4.1. The equilibrium entry conditions are the only restrictions on $U_m(\cdot)$ and c_m, since neither the utility function nor entry cost affect bidding strategy in the O auction.

We see from (7) and (8) that, given $F_m(\cdot|\cdot)$ and $\bar{s}_m$, the expected profit from entering the auction Π^o is monotonically decreasing in c_m for any increasing utility function $U_m(\cdot)$. Also, we see from (7) that if $c_m = 0$, $\Pi^o_m(\bar{s}_m|\{\bar{s}_m\}) > U_m(0)$, and if $c_m = \bar{v}$, $\Pi^o_m(\bar{s}_m|\{\bar{s}_m\}) < U_m(0)$. Then given $F_m(\cdot|\cdot)$ and $\bar{s}_m$, there must exist a $c_m \in (0, \bar{v})$ that satisfies the equilibrium entry condition $\Pi^o_m(\bar{s}_m|\{\bar{s}_m\}) = U_m(0)$ for any increasing utility function $U_m(\cdot)$. Hence, $U_m(\cdot)$ and c_m are not identified.

Proof of Proposition 3 Let $\mathcal{U}$ be the set of $\lambda(\cdot)$ that satisfy $\xi'(\cdot) > 0$, i.e. the set of $\lambda(\cdot)$ that result in strictly increasing bid functions. Let $\mathcal{F}$ be the set of $\{F(v|s) : s \in [0,1]\}$ that are stochastically ordered and have identical support, in accordance with assumptions A2 and A3. The model is identified if there is a unique solution $[F_m(\cdot|\cdot), U_m(\cdot), c_m]$ to the following system of equations and conditions for the subgroups $m = 1, 2$:

C1 $\lambda_m(x) = \frac{U_m(x-c)-U_m(-c)}{U_m'(x-c)}$

C2 $[F_m(\cdot|\cdot), \lambda_m(\cdot), c_m] \in \mathcal{F} \times \mathcal{U} \times \mathbb{R}$

C3 The distribution of $\xi_m(b) \equiv b + \lambda_m^{-1}(\frac{W_m(b)}{dW_m(b)/db})$ is equal to $F_m^*(v; \bar{s}_m) \equiv \frac{1}{1-\bar{s}_m} \int_{\bar{s}_m}^1 F_m(v|t)dt$

C4 $\Pi_1(\bar{s}_1^*|\{\bar{s}_m^*\}) = U_1(0)$ and $\Pi_2(\bar{s}_2^*|\{\bar{s}_m^*\}) = U_2(0)$

C1 restates the definition of $\lambda_m(\cdot)$, a transformation of $U_m(\cdot)$ which is easier to work with. C3 says the first-order condition for bidding must be satisfied. C4 lists the equilibrium entry conditions for the two subgroups.

By the proof of Remark 2 below, $F_m(\cdot|\cdot)$ satisfying C2 and C3 exist for any $\lambda_m(\cdot) \in \mathcal{U}$; therefore, C2 and C3 together reduce to the condition $\lambda_m(\cdot) \in \mathcal{U}$. Now define $\psi_m(b) \equiv \frac{W_m(b)}{dW_m(b)/db}$. The proof of Guerre, Perrigne, and Vuong (2009), Proposition 1 shows that $\lambda(\cdot) \in \mathcal{U}$ reduces to a condition that $\underline{\lambda}' > -\underline{\psi}'$. Then the entire system listed above reduces to $\underline{\lambda'_m} > -\underline{\psi'_m}$ along with the equations C1 and C4.

Now, I show that for any $\lambda_m(\cdot)$ satisfying $\underline{\lambda'_m} > -\underline{\psi'_m}$, there exist $U_m(\cdot)$ and c_m that satisfy C1 and C4. By definition, $\lambda(x) = \frac{\tilde{U}(x)}{\tilde{U}'(x)}$. The solution for $\tilde{U}(x)$ is $\tilde{U}(x) = exp \int^x 1/\lambda(t)dt$, where the starting value for the integral determines the scale normalization. Using this $\tilde{U}_m(\cdot)$, expected stage 2 profit in (2) can be rewritten as $\pi_m(v, b|\{\bar{s}_m\}) = \tilde{U}_m(v-b)W_m(b|\{\bar{s}_m\}) - \tilde{U}_m(c_m)$. This expression is monotonically decreasing in c_m given $\tilde{U}(x)$, $\bar{s}_m$, and the observed bid distribution, so $\Pi_m(\bar{s}_m|\{\bar{s}_m\})$ is also monotonically decreasing in c_m. If $c_m = 0$, $\Pi_m(\bar{s}_m|\{\bar{s}_m\}) > U_m(0)$, and if $c_m = \bar{v}$, $\Pi_m(\bar{s}_m|\{\bar{s}_m\}) < U_m(0)$. Thus there exists c_m that satisfies C4, the equilibrium entry condition $\Pi_m(\bar{s}_m|\{\bar{s}_m\}) = U_m(0)$. Finally, there exists $U_m(x)$ that satisfies $\tilde{U}_m(x) = U_m(x-c) - U_m(-c)$, namely $U_m(x) = \tilde{U}_m(x+c) - \tilde{U}_m(c)$. This means there exists $U_m(x)$ that satisfies C1.

So for any λ satisfying $\underline{\lambda'_m} > -\underline{\psi'_m}$, there exists a structure $[F_m(\cdot|\cdot), U_m(\cdot), c_m]$ that satisfies the conditions C1-C4. Therefore, the model is not identified.

Proof of Remark 1 Assume that $F(\cdot|\cdot)$ does not depend on N or c. Further, assume N and c are determined independently of each other. Since $\bar{s}(c, N)$ is increasing in both N and c, we can think of two structures $\{N, c, U(\cdot), F(\cdot|\cdot)\}$ and $\{N', c', U(\cdot), F(\cdot|\cdot)\}$ with $N' > N$ and $c' < c$ that satisfy $\bar{s}(c, N) = \bar{s}(c', N')$. Since the bidding strategy $b(\cdot)$ is increasing in N and decreasing in c when $\bar{s}$ is held constant, $b(v; N', c') > b(v; N, c)$. Suppose there exists a value of $\bar{s}$, say $\bar{s} = y$, for which we observe two different values of N in the data. Given this setting, we can identify $F^*(v;\bar{s})$ using the identification strategy of Guerre, Perrigne, and Vuong (2009).

Let $b_L(\alpha)$ be the α-quantile of submitted bids (including placeholders) for the structure (N,c), and define $b_H(\alpha)$ similarly for (N',c'), so that $b_H(\alpha) > b_L(\alpha)$. Let $v(\alpha)$ be the α-quantile of the distribution $F^*(\cdot;y)$, where y is recovered directly from the data. Then compatibility conditions are

$$v(\alpha) = b_L(\alpha) + \lambda^{-1}(\frac{W(b_L(\alpha)|y)}{dW(b_L(\alpha)|y)/db}) = b_H(\alpha) + \lambda^{-1}(\frac{W(b_H(\alpha)|y)}{dW(b_H(\alpha)|y)/db}) \tag{18}$$

for all $\alpha \in [0,1]$. Since the distribution of bids follows $G_j(b|y) \equiv F^*(b_j^{-1}(b);y)$ $j \in \{L,H\}$, we have $\frac{W(b_j(\alpha)|y)}{dW(b_j(\alpha)|y)/db} = \frac{G_j(b_j(\alpha)|y)^{N_j-1}}{(N_j-1)G_j(b_j(\alpha)|y)^{N_j-2}g_j(b_j(\alpha)|y)} = \frac{G_j(b_j(\alpha)|y)}{(N_j-1)g_j(b_j(\alpha)|y)} = \frac{1}{N_j-1}\frac{\alpha}{g_j(b_j(\alpha)|y)}$ for $\alpha \in [0,1]$. Then the compatibility conditions can be rewritten as

$$v(\alpha) = b_L(\alpha) + \lambda^{-1}(\frac{1}{N_L-1}\frac{\alpha}{g_L(b_L(\alpha)|y)}) = b_H(\alpha) + \lambda^{-1}(\frac{1}{N_H-1}\frac{\alpha}{g_H(b_H(\alpha)|y)}) \tag{19}$$

Define $R_j(\alpha) \equiv \frac{1}{N_j-1}\frac{\alpha}{g_j(b_j(\alpha)|y)}$ for $\alpha \in [0,1]$. Now we see that these are the exact compatibility conditions used in GPV 2009. Hence $\lambda(\cdot)$ and $F^*(\cdot;y)$ are identified. Once $\lambda(\cdot)$ is identified, we can use the inversion of Guerre et al. (2000) to identify $F^*(\cdot;\bar{s})$ for all observed $\bar{s}$. Arguments made previously using an exclusion restriction and continuous variation assumption allow identification of $F(\cdot|\cdot)$ from $F^*(\cdot;\cdot)$. And the entry equilibrium equations, along with the definition $\lambda(x) = \frac{U(x-c)-U(-c)}{U'(x-c)}$, identify $U(\cdot)$, c, and c'.

Remark 2 For any cdf, there exist stochastically ordered $F_m(v|s)$ with identical support such that $\frac{1}{1-\bar{s}_m}\int_{\bar{s}_m}^{1} F_m(v|t)dt$ is equal to that cdf.

Proof. There exists for any cdf $F^*(\cdot;\bar{s})$ a function $k(\cdot)$ defined on $[\underline{v},\bar{v}]$ that satisfies the following conditions:

1. $k(\cdot)$ is continuously differentiable
2. $\frac{\partial}{\partial v}[F^*(v;\bar{s}) + k(v)] > 0$ and $\frac{\partial}{\partial v}[F^*(v;\bar{s}) - k(v)] > 0$; i.e. $k'(v) \in (-f^*(v;\bar{s}), f^*(v;\bar{s}))$
3. $\forall v \in (0,1)$ $k(v) < 0$
4. $k(\underline{v}) = 0$
5. $k(\bar{v}) = 0$

(Condition 2 just says $k'(v)$ must lie in some band around zero which is continuously changing in v, since $f^*(v;\bar{s})$ is continuous.) $k(\cdot)$ must exist because it can be freely chosen from the infinite number of functions that satisfy conditions 1-3, subject to only two point equalities, conditions 4 and 5. For instance, I can construct one such $k(\cdot)$ this way:

Let $k(\underline{v}) = 0$, satisfying condition 4.

For each $v < \frac{\underline{v}+\bar{v}}{2}$, pick $k'(v) \in (max\{-f^*(v;\bar{s}), -f^*(\bar{v}-(v-\underline{v});\bar{s})\}, 0)$ and let $k'(\frac{\underline{v}+\bar{v}}{2}) = 0$, while maintaining continuity of $k'(v)$. This satisfies conditions 1 and 2.

For each $v > \frac{\underline{v}+\bar{v}}{2}$, let $k'(v) = -k'(\underline{v} + (\bar{v}-v))$ (i.e. $k'(v)$ to the left and right of $\frac{\underline{v}+\bar{v}}{2}$ are mirror images of each other)

Then $k(v) = \int_{\underline{v}}^{v} k'(v)dv + k(\underline{v})$.

The construction of $k'(v)$ above gives $k(\bar{v}) = \int_{\underline{v}}^{\bar{v}} k'(v)dv + k(\underline{v}) = 0$, satisfying #5. As for #3, we know $k(v) < 0$ for $v \leq \frac{\underline{v}+\bar{v}}{2}$ since $k'(v) < 0$ there. Also, since $k(v)$ is monotonically increasing for $v > \frac{\underline{v}+\bar{v}}{2}$ and $k(\bar{v}) = 0$, it must be that $k(v) < 0$ for $v > \frac{\underline{v}+\bar{v}}{2}$ as well. Hence #3 is satisfied.

Given such a $k(\cdot)$, let $F(v|s) = F^*(v;\bar{s}) + \frac{2s-\bar{s}-1}{1-\bar{s}}k(v)$. Then we see that

$$F(\bar{v}|s) = F^*(\bar{v};\bar{s}) + \frac{2s-\bar{s}-1}{1-\bar{s}}k(\bar{v}) = 1 + 0 = 1$$

Since $\frac{2s-\bar{s}-1}{1-\bar{s}} \in [-1,1]$ and $k'(v) \in (-f^*(v;\bar{s}), f^*(v;\bar{s}))$ from #2, $\frac{2s-\bar{s}-1}{1-\bar{s}}k'(v) \in (-f^*(v;\bar{s}), f^*(v;\bar{s}))$. Hence

$$\frac{\partial}{\partial v}F(v|s) = \frac{\partial}{\partial v}[F^*(v;\bar{s}) + \frac{2s-\bar{s}-1}{1-\bar{s}}k(v)] > 0 \text{ (monotonicity)}$$

$$\frac{\partial}{\partial s}F(v|s) = \frac{2}{1-\bar{s}}k(v) \leq 0 \text{ (stochastic ordering)}$$

$$\begin{aligned}
\frac{1}{1-\bar{s}}\int_{\bar{s}}^{1} F(v|t)dt &= \frac{1}{1-\bar{s}}\int_{\bar{s}}^{1} F^*(v;\bar{s}) + \frac{2t-\bar{s}-1}{1-\bar{s}}k(v)dt \\
&= F^*(v;\bar{s}) + k(v)\frac{1}{(1-\bar{s})^2}[\int_{\bar{s}}^{1} 2tdt + (-\bar{s}-1)(1-\bar{s})] \\
&= F^*(v;\bar{s}) + k(v)\frac{1}{(1-\bar{s})^2}[1-\bar{s}^2-(1-\bar{s}^2)] \\
&= F^*(v;\bar{s})
\end{aligned}$$

So we have found stochastically ordered cdfs $F(v|s)$ such that $\frac{1}{1-\bar{s}}\int_{\bar{s}}^{1} F(v|t)dt = F^*(v;\bar{s})$. □

Criterion for selecting polynomial degrees in the sieve maximum likelihood estimator

Just as kernel methods require choosing a bandwidth, sieve methods require choosing degrees for the polynomial bases being used. One way to do this is to choose the polynomial degrees that minimize a relevant criterion.

The estimator estimates the unknown value distributions $F_m(v|s,z)$ using the observed distribution of O prices conditional on entry thresholds and auction covariates, $F(p|\bar{s},z)$. Therefore, $F(p|\bar{s},z)$ is a good candidate to use as the basis for building a mean squared error (MSE) criterion. Specifically, I choose the polynomial degrees that yield the best approximation of observable $F(p|\bar{s},z)$ by minimizing the mean squared error between observed and approximated $F(p|\bar{s},z)$. The observed distribution can be computed with kernel methods using rule of thumb bandwidths.

With a trivariate polynomial, the number of parameters $\alpha_{p,q,r}$ to be estimated grows steeply with the choice of polynomial degrees. When using Bernstein polynomials, the number of polynomial coefficients needed to approximate subgroup 1's $F_1(v|s,z)$ is $(m+1)(n+1)(l+1)$, and for subgroup 2's $F_2(v|z)$ it is $(m+1)(l+1)$ (refer to sections 5.2 and 6.4). Even if I exploit $F(\underline{v}) = F(r) = 0$, which means pinning down $\alpha_{0,q,r} = 0$, the total number of parameters to be estimated is $m(n+1)(l+1)+m(l+1)$. As an illustration, allowing a degree of just 5 for each of the three variables leads to 210 parameters being estimated. Thus, minimizing the MSE criterion without any constraints could lead to polynomial degrees that yield an impractically high number of parameters to estimate.

On the other hand, the parameters being estimated are not free and independent. As the polynomials are approximating cdf's, all Bernstein polynomial parameters must lie within [0,1]. Furthermore, since I restrict $F_m(v|s,z)$ to be nondecreasing in v, as all cdf's should be, $\alpha_{p,q,r}$ is bounded by $[\alpha_{p-1,q,r},\alpha_{p+1,q,r}]$; each parameter is bounded by other parameters. Therefore, estimating parameters in this context is less demanding than estimating the same number of parameters when they are independent and unbounded.

For my estimation, I choose polynomial degrees that minimize the MSE criterion subject to a constraint that the number of parameters to be estimated must not exceed 200, i.e. $m(n+1)(l+1)+m(l+1) \leq 200$. The criterion minimizing degrees are $m = 10$, $n = 2$, $l = 4$.

Simulations of the sieve maximum likelihood estimator

I test the estimator using simulated data. As shown in Table 9, the simulated environment mimics the environment of the actual auction, with the goal of testing how well the estimator would perform in this application. In choosing the shapes of the distributions $F_m(v|s,z)$ from which to draw the simulated data, I rely on the distribution of observed sealed bids rather than choosing an arbitrary shape. Specifically, I construct the "true" value distribution for the simulation by approximating the observed bid distribution within the space of Bernstein polynomials. Then, using these value distributions and the settings in Table 9, I simulate the data (transaction prices and winners' subgroup identities) that would be generated by the auction model of this paper. Finally, the sieve maximum likelihood estimator is employed on this simulated data. This process is repeated 100 times.

Simulation results for $\hat{F}_1(v|s,z)$ are shown in Figure 8. In the left column I hold z fixed at the median value while varying s, and in the right column I hold s fixed at the median value while varying z. The solid dark line is the true $F_1(v|s,z)$ from which simulated data were drawn, and the light dashed lines represent the 100 Monte Carlo runs of the estimator. As s gets lower, the estimated $F_1(v|s,z)$ becomes less precise. This is as expected; the lower a bidder's signal, the less likely he will enter the auction, so bids for bidders with lower s are less frequently observed. Figure 9 shows results for $\hat{F}_2(v|z)$, varying z.

Table 9: Simulated data and estimator settings

items auctioned	1000
N_1	23
N_2	1
fixed entry rate of subgroup 2	0.84
$\bar{s}_1$	$U[0,1]$
s	$U[0,1]$
z	$U[0,1]$

Bernstein polynomial degree:	
v	10
s	2
z	4

Figure 8: Monte Carlo simulations of sieve maximum likelihood estimator, subgroup 1

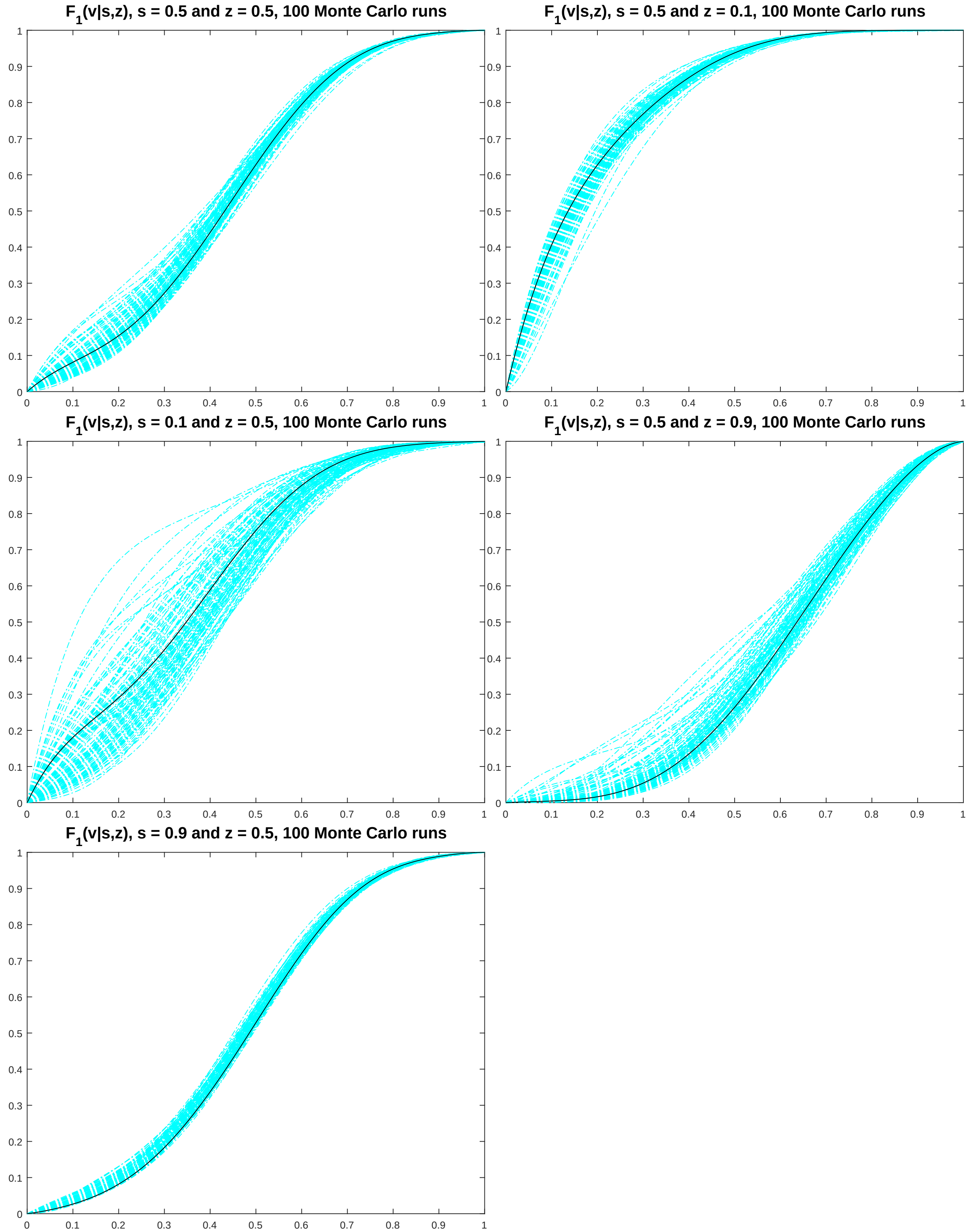
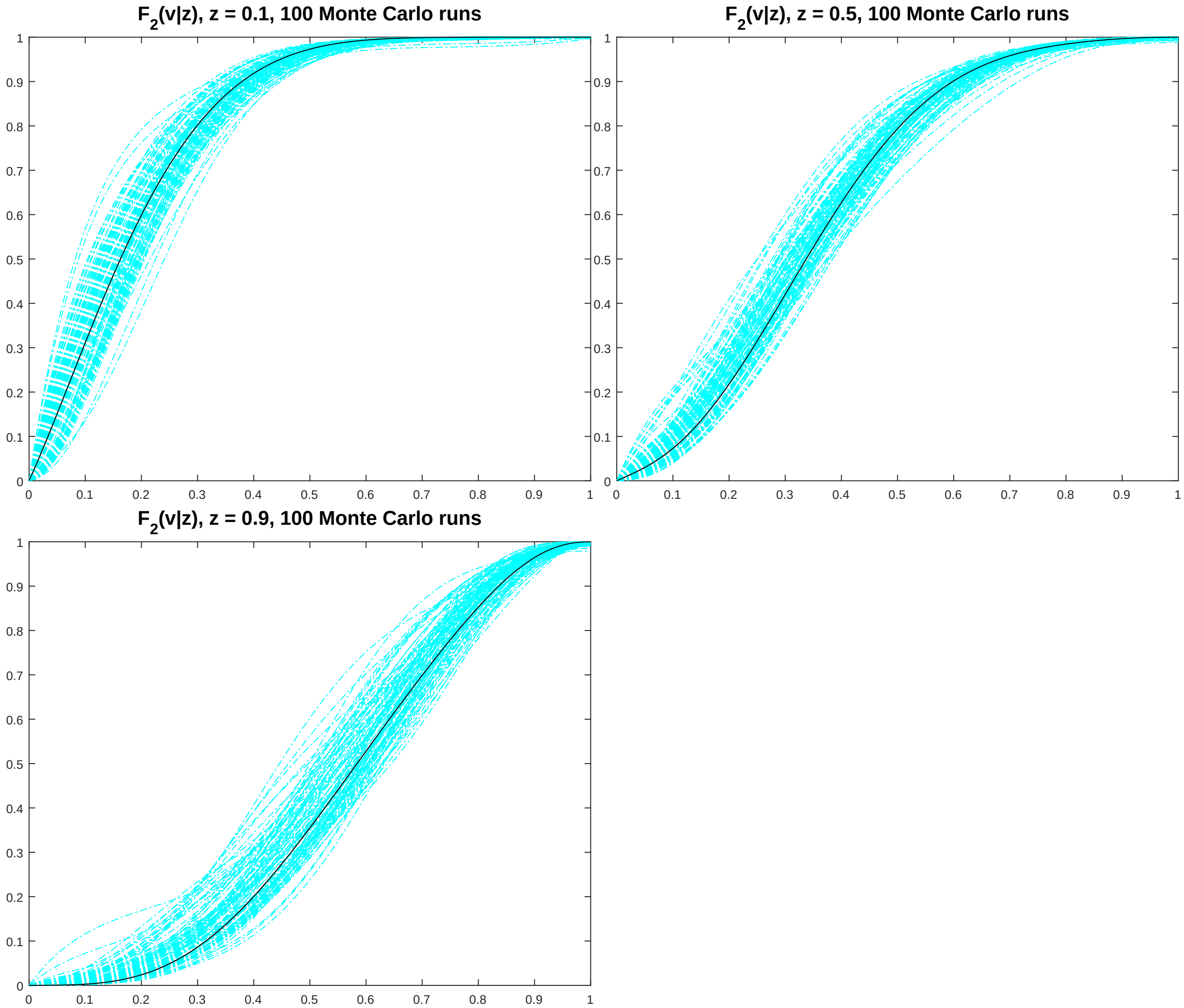


Figure 9: Monte Carlo simulations of sieve maximum likelihood estimator, subgroup 2

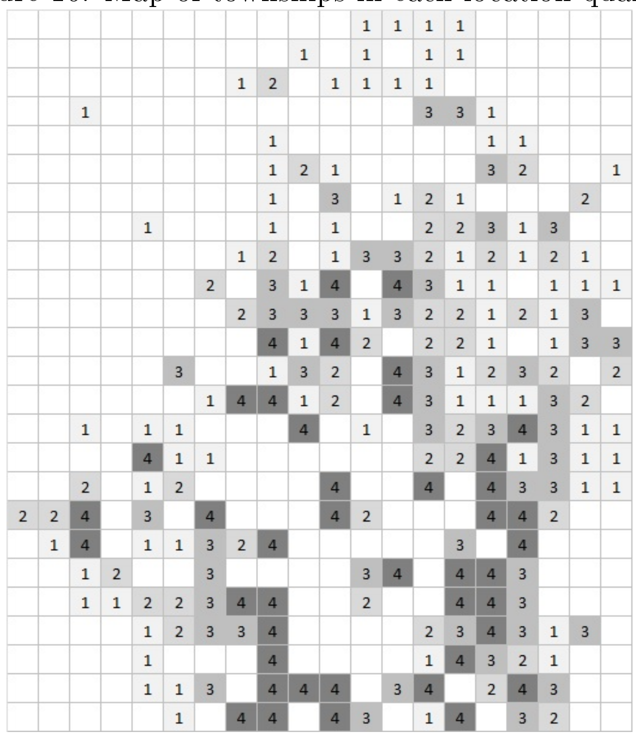


Covariates z

Observable characteristics of auctioned leases fall into three major categories: terms of the lease (royalty rate, annual rental), time of auction (industry, economic, local conditions of that time), and location of the tract (encompassing geological features). The royalty rate is indicated by the lease prefix, V0 or VB. The rental rate ($0.50 or $1) is determined by whether the tract is located in a township north ($0.50) or south ($1) of a horizontal geographic line; thus it is subsumed by the location variables. Year fixed effects represent the time of auction, supplemented by oil and gas prices.

Location contains important information and is observed at a detailed level. To reduce this information to a smaller number of variables while retaining flexibility, I first regress all submitted sealed bids on the lease prefix and time variables, along with township (6-by-6 square mile) fixed effects, of which there are more than 200. I then sort the township coefficients into quartiles, and assign a dummy variable for each quartile. Mapping the quartiles (Figure 10) shows a pattern in which townships of the south-central area are highest value and values decline moving further out and away from this "center."

Figure 10: Map of townships in each location quartile



To allow variation within township, I supplement the quartile dummies with a distance-to-center variable that is computed relative to own township: taking the average x-y coordinates of all top-quartile townships as the "center" of high value, I compute how much farther each auctioned lease is from this "center" relative to the centroid of the township it is located in. The conjecture is that within each township, tracts that are located closer to the "center" will have relatively higher value. This relative position within township seems to matter only for townships within some radius of the center's influence, so I add this variable only for the upper quartiles. The drilling and production history associated with each geographic location is also observed.[41]

Table 10: Summary statistics of covariates

variables	mean	min	max
lease prefix VB (dummy)	0.389		
drilled before (dummy)	0.439		
log production 1970-auction date (boe)	1.099	0	16.91
WTI oil price†	79.59	39.07	135.2
nat gas 1 mo futures†	6.101	1.952	14.45

† Deflated by GDP implicit price deflator.

Not summarized: year fixed effect dummies and geographic location descriptors.

[41] The information contained in geographic location could be organized in alternative forms.

To ascertain which characteristics of the lease most affect its value to bidders, I regress the log of submitted sealed bids on these covariates. Table 11 shows the results. The coefficient on the dummy variable for lease prefix VB is highly significant and positive because the NMSLO assigns the VB prefix to premium tracts. Also, the location quartile fixed effects are higher for higher quartiles, and being further away from the "center" leads to lower bids. Controlling for year fixed effects, gas prices seem to explain bids better than oil prices. The incremental explanatory power of the remaining variables on bids is negligible, perhaps because this information is subsumed in the lease prefix and location variables. As such, I take the covariates in column (3) to form the single index.

Table 11: Regression of ln(sealed bid) on observable characteristics

	(1)	(2)	(3)
lease prefix VB	0.257	0.253	0.256
	(0.042)	(0.043)	(0.042)
location quartile 2	0.552	0.552	0.551
	(0.051)	(0.051)	(0.051)
location quartile 3	0.986	0.989	0.989
	(0.053)	(0.053)	(0.053)
location quartile 4	1.743	1.746	1.745
	(0.057)	(0.057)	(0.057)
relative dist. to center (if upper qrtl)	-0.054	-0.056	-0.054
	(0.018)	(0.018)	(0.018)
nat gas 1 mo futures	0.036		0.030
	(0.016)		(0.014)
WTI oil price	-0.001	0.001	
	(0.002)	(0.001)	
drilled before	-0.023	-0.018	
	(0.041)	(0.041)	
log production 1970-auction date (boe)	0.003	0.003	
	(0.006)	(0.006)	
Year FE	Y	Y	Y
Observations	3083	3083	3083
R^2	0.352	0.351	0.352
Adjusted R^2	0.348	0.347	0.348

Heteroskedasticity robust standard errors in parentheses

Method for simulating counterfactual bid functions

With asymmetric bidders and other model features, we do not have analytical solutions for the bid function $b_m(v)$ in most counterfactual scenarios. They need to be computed numerically in some way.

For each of the counterfactual scenarios I consider, the relevant inverse bid function $\xi_m(b)$ can be written in terms of b and the distribution of b, analogously to (11). Of course, the function $\xi_m(b)$ depends on N_m, $U_m(\cdot)$, whether bidders know the number of realized bidders, and other details that can change for each scenario. Now, let *vgrid* denote a grid of values that spans the support of v. Given the bid function of subgroup $-m$, I can compute the bid function for subgroup m by searching for the $b_m(\cdot)$ that minimizes the sum of log differences between v and $\xi_m(b_m(v))$ over this grid, $\sum_i |ln(vgrid_i) - ln(\xi_m(b_m(vgrid_i)))|$. I conduct the search for $b_m(\cdot)$ in the space of monotonically increasing Bernstein polynomials of degree 10. Starting values for the search are set at approximations of the $b_m(\cdot)$ observed in actual data.

As there are two asymmetric subgroups, I implement an iterated best response procedure, in which this minimization is iterated in turns for each group while holding the other group's strategy fixed at the last iteration's $\hat{b}_m(\cdot)$. The final $\hat{b}_m(\cdot)$ are obtained when they stabilize for both groups. In practice, I observe that the $\hat{b}_m(\cdot)$ converge quickly, stabilizing by the 2nd or 3rd iteration.

Table 12: Binomial logistic regression of n_1 on x and $z'\beta$

Non P. Basin acreage (1000s)	-0.011
	(0.003)
$z'\beta$	-43.561
	(9.747)
$(z'\beta)^2$	3.983
	(0.909)
$(z'\beta)^3$	-0.120
	(0.028)
Constant	154.937
	(34.704)
Observations	1039

Standard errors in parentheses

Table 13: Revenue response to drop in entry threshold, at modal $z'\beta$

	Selective Entry		Nonselective Entry	
$\bar{s}_1$	E[S price]	%Δ (logΔ)	E[S price]	%Δ (logΔ)
0.943	63,440	-	63,440	-
0.922	83,973	32% (0.28)	88,077	39% (0.33)

* Everything other than selection in entry is kept the same between "Selective" and "Nonselective" columns, including utility functions and entry costs.