Entry and Selection in Auctions

Preliminary, please do not cite or quote without permission.

James W. Roberts* Andrew Sweeting†

May 12, 2010

Abstract

We develop and estimate an independent private value auction model where potential bidders receive an imperfect signal about their value prior to making a costly entry decision. The estimation allows for asymmetries in bidders and unobserved, to the econometrician, object heterogeneity. We illustrate how incorrectly assuming that the current leading models of entry are the data generating processes will (i) cause misestimates of model primitives and (ii) generate bias in important counterfactual analyses. We apply our model to timber auctions to demonstrate how these two consequences affect an important and often studied empirical setting.

JEL CODES: D44, L20, L73

Keywords: Auctions, Entry, Unobserved Heterogeneity

*Department of Economics, Duke University. Contact: j.roberts@duke.edu.

†Department of Economics, Duke University and NBER. Contact: atsweet@duke.edu.

We are extremely grateful to Susan Athey, Jonathan Levin and Enrique Seira for sharing the data with us and for the useful discussions with Doug MacDonald. Excellent research assistance was provided by Christoph Bauner, Yair Taylor and Huihui Wang. Any errors are our own.

1 Introduction

Empirical entry models in industrial organization typically ignore the possibility of selective entry with respect to *ex post* profitability. That is, the notion that actual players chose to participate because they knew prior to paying an entry fee that they were likely to be "competitive" is ruled out. In traditional differentiated product settings this source of "competitiveness" might stem from favorable cost or quality advantages. To address the impact of this source of selection we focus on an empirically important setting: auctions.

In many auctions bidders must make sunk investments to better understand the potential returns to participating in the auction. A few examples (which correspond to the leading applications for much of the empirical auction literature) include timber auctions where bidders need to "cruise" the tract they are bidding on to learn about the types of trees they will harvest, procurement auctions, such as those for highway paving contracts, where bidders must assess the project in order to gauge the likely cost of winning the job and offshore oil auctions where bidders conduct seismic surveys to form expectations about the amount of oil present. Despite their importance, entry costs have only recently been incorporated into empirical auction research. When models include these entry costs, the assumptions made about bidders' information are often extreme. For example, a common assumption is that bidders have no private information about their value prior to paying the entry cost. A common alternative assumption is that bidders know their value perfectly before paying an entry cost. These models are based on Levin and Smith (1994) and Samuelson (1985), respectively. Throughout the paper we refer to these as the LS and S models. These polar cases are rarely going to be correct and a more reasonable modeling assumption lies somewhere in between. This is the type of model we explore in this paper.

Specifically, we estimate an independent private value ascending (or second price) auction model where bidders observe an imperfect signal of their value prior to making a costly investment necessary to learn their true value. We call this the Signal model. We allow for asymmetric bidders and object heterogeneity which is not observed by the econometrician but is observable to bidders. Equilibrium for each type of bidder is characterized by an entry threshold whereby their signal must be sufficiently optimistic about their likely value for the object to justify paying the sunk entry cost.

This matters because it creates a difference between marginal and infra-marginal bidders. Not only will ignoring this difference cause the econometrician to misestimate the distribution of bidder values, the central object of interest in much of the empirical auction literature, there will be direct effects on counterfactual analysis. For example, consider a subsidy to encourage greater participation in an auction. The additional (subsidized) entrants are those whose value for the object is low since they received a signal which forecasted a sufficiently low value that they decided not to incur the sunk investment cost. Therefore, it is likely that the gains from having these additional bidders participate in the auction are overstated since it is doubtful they will be competitive with the former entrants.

We also illustrate the impact utilizing the incorrect entry model has in empirical applications. We begin by focusing on the estimation of the distribution of bidder values and show that, among other biases, models that ignore selection will tend to overestimate the true mean of the value distribution. Intuitively, this results from the econometrician assuming that auction participants are a representative sample of potential entrants. We illustrate and discuss other biases below. Next we turn towards assessing likely biases in counterfactual analysis. To do this we focus on the difference in the marginal participant's value distribution according to competing entry models. It is clear that in the more realistic Signal model this distribution will be first order stochastically dominated by the value distributions non-marginal entrants. We then illustrate how this will directly impact questions such as subsidies and mergers.

We also focus on an artifact of models without selection, noted elsewhere in the literature, that can be used to identify the presence of selection in auction data. In the LS model, as the number of potential entrants increases, the seller's expected revenue is likely to decline.[1] Since a priori, we are not clear whether this should occur in the Selection model, we provide numerical examples to show that a model with selection does not require this to be the case. The intuition as to why a seller's revenue falls as the number of potential entrants increases centers around inefficiencies induced by entry. In models without selection, social inefficiency stemming from high value bidders staying out of the auction increases as the number of potential entrants increases. In contrast, models with selection mitigate this effect enough to overturn the negative impact on revenues. Further, this also suggests a simple test of whether or not the true data generating process involves a model without selection when revenue data is available.

After mapping out the problems resulting from applying the wrong entry model to the data generating process, we develop our estimator for the Signal model. We estimate the model using a nested pseudo-likelihood procedure (Aguirregabiria and Mira (2002)). This differs from the nested fixed point method used elsewhere in the literature (Einav and Esponda (2008)). Our methodology allows for observable asymmetries among bidders and extends to cases where there is a finite mixture of unobserved item heterogeneity. Through a variety of Monte Carlo experiments, we show that our estimator performs well in uncovering the true parameters of the data generating process.

We then turn towards our empirical application. We focus on timber auctions to demonstrate the impact on estimating model primitives when the unlikely polar entry cases are assumed relative to the more realistic Signal model. We find substantial reduced form evidence of selection in these data. We then estimate a structural model of entry and bidding in these auctions with an eye towards counterfactual analysis. We find that the difference in values across types of bidders based only on observed bids greatly understates the true difference in these types' value distributions. Moreover, within bidder type, there is a large difference in the value distributions of marginal and infra-marginal bidders. However, we do not find perfect selection. These findings are important for counterfactual analysis. In particular, we find that the benefits of programs designed to help small bidders are likely overstated when selection is ignored.

The paper proceeds as follows. Section 2 discusses the relevant literature, Section 3 introduces our model and discusses its identification and estimation, Section 4 presents our estimator and

[1] Levin and Smith (1994) note that this happens when bidders enter with some probability less than 1, the case where entry decisions are interesting.

Monte Carlo studies of its effectiveness, Section 5 turns to our empirical application and Section 6 concludes.

2 Literature Review

In the rapidly growing empirical auction literature (see Hendricks and Porter (2007) for a recent survey) bidders' decisions of whether or not to participate in an auction are frequently ignored. When the set of bidders is not considered fixed, the literature often employs the LS model. Examples include Athey, Levin, and Seira (forthcoming) who examine timber auctions, Bajari and Hortaçsu (2003) who consider coin auctions on eBay and Krasnokutskaya and Seim (2007) who analyze procurement auctions (we discuss this last paper in greater detail below). Palfrey and Pevnitskaya (2008) and Ertaç, Hortaçsu, and Roberts (forthcoming) analyze entry in auctions using experimental data and some structural techniques. In a recent paper Li and Zheng (2008) compare the effects on competition in procurement auctions stemming from both the LS and S models. On the other hand, we are interested in a less restrictive model that does not make the limit case assumptions of completely perfect or imperfect information about one's value at the entry stage. Such a model has been proposed in the literature before. For example, Hendricks, Pinkse, and Porter (2003) consider such a model in their testing of competitive equilibrium bidding in offshore oil. In a recent paper Marmer, Shneyerov, and Xu (2007) consider testing whether the LS, S or a general affiliated signal model of entry best explains bidding behavior in procurement auctions. The idea behind their test is, relying on exogenous variation in the number of potential bidders, to examine whether the distributions of entrants' valuations varies with the number of potential bidders. They reject the S model and find some support for the LS model and their affiliated signal model. In this paper we aim to estimate a fully structural model of entry and bidding in independent private value ascending auctions with asymmetric bidders and unobserved heterogeneity. The objects of interest will be bidders' true value distributions (that is the value distribution of any bidder, not only a participant), signal distributions and entry costs. The aim is to illustrate the biases in demand estimation and counterfactual analysis when the wrong model of entry is employed.

In recent work, Einav and Esponda (2008) investigate a model very similar to the one proposed in this paper for highway procurement auctions. Similar to their paper, we adopt a parametric approach. Our work differs from theirs, however, because we consider ascending auctions and allow for unobserved heterogeneity in our estimation. Our estimation strategy of using nested pseudo-liklihood differs from, and is less computationally burdensome than, theirs.

Performing counterfactual analyses is one goal of estimating our structural model. In related work, Brannman and Froeb (2000) use a sample of 51 Oregon timber auctions to evaluate various counterfactuals such as mergers and bidder preference programs. Our analysis differs from theirs not only in estimation strategy but also by our considering the entry decisions of bidders. One counterfactual topic we consider to be particularly important is that of entry subsidies. In ongoing work, Krasnokutskaya and Seim (2007) employ the LS model to highway procurement to assess current bidder preferential treatment programs where small contractors receive subsidies to make them more competitive with their larger rivals. Their analysis finds that the effects of the program and its resulting impact on government costs vary by project, but in general smaller bidders have an increased chance of winning. Assuming the opposite model of entry, Hubbard and Paarsch (2009) evaluate bid preference programs when bidders know their value prior to participation. They find that the effect of preference programs on bidder participation is relatively unimportant.

Our estimation approach will also allow for object heterogeneity that is observed by bidders but not by the econometrician. Recently, the auction literature has begun to control for such unobserved heterogeneity in a variety of ways (see for example Krasnokutskaya (2009), Athey, Levin, and Seira (forthcoming), Roberts (2009), Hu, McAdams, and Shum (2009)). In this paper we take a new approach for the auction literature and allow for a finite mixture of unobserved heterogeneity.

This paper is also related to the literature on estimating incomplete information entry games. In the canonical entry model in this literature, each firm receives a private information shock to its entry cost, and in equilibrium firms choose to enter if their draw is high enough. In our model, potential bidders choose to enter if their signal is high enough. However, an important difference is that in our model the signal is correlated with post-entry behavior (bidding), whereas in the standard model the level of the entry cost has no affect on a firm's behavior or profits post-entry. Instead our model could be viewed as similar to an entry model where firms received imperfectly informative signals about their post-entry marginal costs or qualities.

Finally, in our empirical application we focus on auctions for the right to log federally owned forestland. There is now a long line of empirical auction literature analyzing these auctions (see Paarsch (1997), Baldwin, Marshall, and Richard (1997), Haile (2001) and Athey and Levin (2001) to name a few). We are the first to employ our more general model of entry to these data.

3 Model

We now present our model of entry into auctions. We begin by introducing the model and showing that it is characterized by a cutoff strategy whereby bidders only enter auctions when their signal is sufficiently high. We then discuss identification of the model primitives and then describe how we estimate the model while allowing for unobserved heterogeneity. For much of the discussion we refer to the mechanism as being a second price auction. However, given our informational assumption that bidders have independent private values, the strategies introduced here are akin to those in an English Button auction.[2]

3.1 A General Entry Model with Selection

We consider a series of $t = 1, ..., T$ second price independent private value (IPV) auctions. In any auction there are $\mathbf{N}$ (the set is $\mathcal{N}$) potential bidders who may be one of $\tau = 1, ..., \overline{\tau}$ types. Let the

[2]When we introduce our estimation strategy in Section 4, we extend our methodology to cover more general models of bidding in open outcry auctions which won't require all bidders to bid up to their true value. We are also currently working on extending the analysis to first price auctions.

number of bidders of any type τ be N_τ and the set of these bidders be $\mathcal{N}_\tau$. Thus $\sum_{\tau=1}^{\overline{\tau}} N_\tau = \mathbf{N}$. Consider any auction t. The object being auctioned is characterized by a set of covariates $Z_t = (X_t, U_t)$ where $X_t \in \mathbb{R}^K$ and $U_t \in \mathbb{R}$. The distinction between X and U is that the econometrician and the bidders observe X while only the bidders observe U.[3] Bidder values V are distributed according to $F_\tau^V(V|Z)$ where the dependence on their type and the auction covariates is made explicit. Bidders know the distribution from which their and their competitors' values are drawn. There are two stages to the game, an entry stage and an auction stage.

At the start of the entry stage every bidder observes the set of potential bidders and a signal S which is affiliated with their value of the object. For expositional ease, we will focus on the special case where $S = V + \varepsilon$, $V \sim N(\mu_\tau, \sigma_V^2)$ and $\varepsilon \sim N(0, \sigma_\varepsilon^2)$, where μ_τ will be allowed to depend on Z.[4] Therefore $E[S|V] = V$, $\varepsilon \sim F^\varepsilon(\cdot)$ and $E[\varepsilon|V] = E[\varepsilon] = 0$.[5] Given these assumptions, we can characterize a bidder's posterior value distribution after observing a signal s. In particular, the bidder now believes his value is drawn from $N(\alpha\mu_\tau + (1-\alpha)s, \sigma_V'^2)$, where $\alpha = \frac{1/\sigma_V^2}{1/\sigma_V^2 + 1/\sigma_\varepsilon^2}$ and $\sigma_V' = \sqrt{\frac{1}{1/\sigma_V^2 + 1/\sigma_\varepsilon^2}}$. During this entry stage all bidders simultaneously decide whether or not to pay a fixed cost K to observe their true value for the object. Any bidder that does pay K proceeds to the auction stage of the game.[6] In some simulations and in our empirical example we consider an alternative model where $S = VA$, $A = e^\varepsilon$, $V \sim logN(\mu_V, \sigma_V^2)$ and $\varepsilon \sim N(0, \sigma_\varepsilon^2)$.

At the auction stage of the game bidders who paid K to learn their true value participate in the auction. In our second price auction setting bidders submit their true value regardless of how many bidders entered. The object is awarded to the highest bidder at a price equal to the second highest bid or the reserve price (which bidders knew at the entry stage), whichever is greater.

A bidder's strategy in this game consists of a rule for deciding whether to enter as a function of his signal, and a rule for bidding as a function of his value and (potentially) his signal. In a second price auction, an optimal strategy will involve an entrant bidding his value, while the entry rule will involve a simple cut-off rule, where he will enter if and only if the signal is above a threshold. The optimal strategy is established in the following proposition and it follows that equilibrium strategies will satisfy this property.

Proposition 1. *The optimal bidding strategy for entrants is to bid one's value. The optimal entry strategy is given by a signal threshold s'_τ for any type τ bidder such that he pays the entry cost if and only if his signal $s > s'_\tau$*

Proof. Let $\mathcal{M}_{-i}$ be the set of entrants other than bidder i. Entrants submit a bid to maximize their expected profit conditional on entry:

$$(v_i - \mathrm{E}[\max\{v_{-i}, r\}] : v_j \leq b, \forall j \in \mathcal{M}_{-i}) \prod_{\mathcal{M}_{-i}} F(b|\tau)$$

[3] As is standard in the empirical auction literature, we restrict U to be unidimensional.

[4] While we can let σ_V and σ_ε depend on τ and Z, for now we make the simplifying assumptions that they do not.

[5] Many of our results, and indeed the intuition for our modeling contribution, could be established with a more general model of the relationship between a bidder's signal and his value (perhaps they are affiliated, for example). However, since our ultimate goal is to take the model to the data, for clarity sake we present this more specific model.

[6] Here we assume that any bidder must pay K to participate in the auction.

Standard arguments show that participants have the dominant strategy to bid their value regardless of the number of potential bidders or asymmetries among them. This strategy is equivalent in an ascending auction.

Profits are increasing in a bidder's value. Because signals and values are affiliated, a higher signal leads the potential entrant to raise his beliefs about his value. Thus for any signal at which the bidder enters, he would enter for any higher signal and for any signal at which he doesn't he wouldn't for any lower signal. Thus an equilibrium entry rule follows the threshold rule. An equilibrium exists because any bidder's reaction function is continuous in his and his opponents' thresholds. *Q.E.D.*

If $\overline{\tau} = 1$, then there is a unique symmetric equilibrium. This is established in the following proposition.

Proposition 2. *With one type of potential entrant there will be a unique symmetric equilibrium.*

Proof. Suppose that there are two symmetric equilibria, where potential entrants have cutoffs s_1^* and s_2^* respectively and $s_1^* > s_2^*$. Consider a potential entrant i. For any v_i, i's expected profits from entering will be increasing in s_{-i}^*, the cut-off used by all other players as for any set of v_{-i}s an increase in s_{-i} reduces the probability that rivals will enter. From this it follows that i's best response cutoff to s_{-i}^* is decreasing in s_{-i}^*. If so, if s_1^* is i's best response to s_1^*, it cannot also be the case that s_2^* is i's best response to s_2^*, so that s_1^* and s_2^* cannot both be equilibrium thresholds. *Q.E.D.*

To provide a bit of intuition about the model, we now explore some comparative statics, consider the one type case. It is straightforward to see that $\frac{\partial s'}{\partial K} > 0$ and $\frac{\partial s'}{\partial \mathbf{N}} > 0$ since a greater sunk investment and lower chances of winning (all else equal) necessitate a more optimistic signal to justify the sunk investment. With regard to the variance of the signal, either stemming from variability in the true value distribution or in the noise (ε), when N is relatively low $\frac{\partial s'}{\partial \sigma} < 0$ reflecting increased option value of participation. However, as N increases, a noisier signal, which causes a bidder to place more weight on his prior distribution, will not justify entry since having the mean value won't be enough to win when there are many competitors.

3.2 Illustrating the Importance of Selection

In this section we illustrate why incorporating models allowing for selection effects at the entry stage can be important in empirical auction work. To do this we first show how applying the incorrect entry model to the data causes the econometrician to misestimate demand. Second, we consider a relevant counterfactual question involving entry subsidies to show how this type of policy-relevant question may be incorrectly addressed when too restrictive models are taken to the data. Finally, in hopes of providing a preliminary test for selection in auction data, we return to a feature of the LS model noted elsewhere in the literature. In this model expected revenues fall with more potential entrants. Through a variety of numerical examples we show that this result contrasts with the

Signal model where we have yet to find an example in which a similar fall in expected revenues arises from an increase in the number of potential bidders.

Misestimating Demand

As stated above, far and away the most commonly used entry model in empirical auction work is the LS model. Here we consider what impact the incorrect application of such models will have on estimates of model primitives. That is, suppose the true data generating process is the Signal model but the econometrician incorrectly applies the LS model. The results of incorrectly assuming the data generating process is the LS model, when in fact it is the Signal model, appear in Table 1. The table shows the estimated parameters of the value distribution when the family is correctly chosen (that is a normal distribution is correctly assumed) but the LS is incorrectly applied. We generated data using our Signal model for given parameters (shown in the table) and then estimated $\mu_V,\ \sigma_V,\ K$.

[Table 1 about here.]

Intuition for the results in Table 1 is straightforward. Generally speaking, the econometrician wrongly assumes that bids observed in the data are placed by representative bidders since, supposedly, there is no selection effect. In reality, bidders with high signals chose to enter and this is correlated with higher values, thereby biasing the econometrician's estimates of the value distribution. In particular, the estimated distribution under the LS model will first order stochastically dominate and be second order stochastically dominated by the distribution under the Signal model.[7] There will also be consequences for estimates of the entry costs. Since the mean of the value distribution will tend to be overestimated, the entry cost will also be estimated to be larger than it is in order to rationalize why there is relatively low observed entry for such a high mean valuation. This bias in K will fall as N increases since the LS model won't predict as much of an increase in entry as does the Signal model and so the increased entry will be partially attributed to lower and lower entry costs.

We can consider a similar experiment with the S model. The results of incorrectly assuming the data generating process is the S model, when in fact it is the more general Signal model, appear in Table 2. The table shows the estimated parameters of the value distribution when the family is correctly chosen (that is a normal distribution is correctly assumed) but the S is incorrectly applied. We generated data using our Signal model for given parameters (shown in the table) and then estimated $\mu_V,\ \sigma_V,\ K$ and s'.

[Table 2 about here.]

The intuition for the results in Table 2 is the opposite of that for Table 1. The S model will predict even greater selection effects than our model and so its estimated value distribution will be

[7] The second order stochastic dominance follows immediately from $\sigma_V > \sigma_V'$ whenever $\sigma_\varepsilon > 0$.

first order stochastically dominated by and second order stochastically dominate the distribution from the Signal model. Also, the S model should predict less entry than the Signal model and it will justify this by underestimating K.

The results in Tables 1 and 2 clearly illustrate the problems from assuming the wrong entry model. In Table 1 we overestimate μ_V and K and underestimate σ_V. The problem is exacerbated when the sample becomes "more" selected as seen in the s' column. Table 2 illustrates similar problems in estimating demand. In particular, since the model assumes that bidders have full information prior to paying their entry cost, the model underestimates the entry threshold of the true model. These misestimates will lead to a host of problems in counterfactual analysis, which we turn to next.

Biasing Counterfactuals: Subsidies

As in any setting, working with the biased estimates of demand will tend to make the economist err in counterfactual analysis. Consider the question of optimal reserve pricing. Tending to overestimate μ_V will cause the economist to overestimate this key tool of optimal auction design. This bias will be somewhat offset by the underestimate of σ_V (see Roberts (2009) for a discussion of this effect), but in general we will still be left with an incorrect estimate of the optimal reserve price. While there are many examples showing how misestimating demand may cause problems for policy recommendations, we choose to focus on one that directly stems from the selection problem which is ignored by the LS entry model: entry subsidies.

We begin by examining the difference between the value distributions of entrants and those bidders who received a signal just low enough to stay out of the auction, what we term to be the "marginal bidder". This will help us to better understand the magnitude of the error in assuming that the marginal bidder has the identical value distribution to a participant, an implication of the LS model. Figure 1 displays the average value distributions of entrants and marginal bidders in the Signal and LS model.

[Figure 1 about here.]

In the LS model these distributions are the same. In the Signal model, however, the bidder receiving a signal $s = s'$ will have a value distribution that is first order stochastically dominated by the value distributions of entrants. Therefore, when we consider the benefit of the marginal bidder participating, we now need to take into account that they will be less competitive than other participants. For example, in the Signal model when $\sigma_\varepsilon = 5$, the marginal bidder can be expected to win only 13.4% of the time, where as the marginal bidder in the LS model wins 33.0% of the time. While the difference between the marginal and the unsubsidized participants diminishes as σ_ε increases, as seen in the second panel of Figure 1, there is still a first order stochastic dominance relationship and the marginal bidder in the Signal model is less likely to win than in the LS model. Another way to consider the value of the marginal participant in the two models is to find their impact on the seller's revenue (holding constant other potential bidders' entry strategies[8]). Conditional on two bidders participating, (so that the winning bid is determined by a second highest bid), the extra revenue in the LS model is over 14 times that in the Signal model.

Do More Potential Entrants Increase Revenues?

A major policy consideration of competitive bidding processes is that there be enough "competition" (for example see Klemperer (2002)). While most of this paper focuses on the realized "competition" at the auction stage between those bidders that actually participate in the mechanism, another interpretation of "competition" is the set of potential bidders in the auction. In this section we return to the fact that in the LS model, as the number of potential bidders increases, a seller's expected revenue likely falls. This fact has previously been noted by Levin and Smith (1994) in their original work on entry in auctions without selection. A priori it is not clear what should happen in our model. Therefore, we provide numerical examples highlighting this effect across a variety of parameterizations, whereas in each of these parameterizations expected revenue in the Signal model increases with more potential entrants. We provide intuition for the result below. This finding serves as an effective way to test whether the data are generated by a model without selection when seller revenues are available.

Table 3 shows that a seller's expected revenue falls as N increases under a variety of specifications for $F(\cdot)$ in the LS model. The table shows results for just some values of N but the pattern of falling revenues is consistent throughout all N we tried. It is interesting that while the expected number of entrants rises in the LS model, there is sufficient chance of getting fewer entrants, thereby hurting seller revenues, that on average the seller's revenue declines.

[Table 3 about here.]

We now compare the effect of increasing N in the LS model to that in the Signal model. Table 4 displays the results. As shown there, increasing N in the Signal model has the predicted effect of improving seller's revenues. This contrasts with the LS model. We have yet to find an example in the Signal model where expected revenues fall when the number of potential bidders increases.

[Table 4 about here.]

One can also view the finding that expected revenues fall with the number of potential entrants in the LS model as reflecting the inefficiency of the entry process. From a social perspective, bidder entry is inefficient unless it raises the highest value in the auction. But with no means of selection, inefficient entry become more and more likely as the number of potential entrants increases. In contrast, with selection, entry may be less likely to happen, but efficient entry is much more likely.

[8] In general these will change as the other bidders know they will face another competitor, but we ignore this effect for now.

3.3 Identification

Now we briefly discuss the identification of our model. Identification has been examined in Marmer, Shneyerov, and Xu (2007) and Einav and Esponda (2008). Intuitively, identification will stem from exogenous movement in firms' participation decisions. If for some auctions we have very low s' for exogenous reasons, then we can identify the true distribution of values by appealing to standard arguments in the literature (see for example Athey and Haile (2002)). This exogenous variation may come from the number or characteristics of potential competitors.

Once the distribution of values is identified, the importance of signal noise would be identified by seeing how the distribution of observed bids changes when there is more selective entry. For example, if potential entrants receive precise signals, then the distribution of observed bids in a second price auction should look like a truncated distribution of values (with the truncation point at s'). If signals are less informative, then the degree of censoring should be less marked. The entry cost can be identified from how many potential entrants enter and where the distribution of values is censored. For example, if K is large then few potential entrants will enter and observed bids should all be close to the winning bid. An interesting question, which we intend to explore further, is whether it is possible to distinguish noisy signals from unobserved heterogeneity in entry costs across bidders. While it may be difficult, without parametric assumptions, to separately identify the effects of i.i.d. signal noise from i.i.d. heterogeneity in K, we provide some evidence below that persistent heterogeneity in K across bidders does not seem to be an important feature of our data.

4 Estimation

We estimate the model using a nested pseudo-likelihood procedure. Aguirregabiria and Mira (2002) and Aguirregabiria and Mira (2007) show how this procedure can be used to estimate both single agent models and games of incomplete information with a lower computational burden than nested fixed point methods. Applied to games, the procedure involves the iteration of two steps: in the first step, a pseudo-likelihood for each player's action is maximized to give estimates of the parameters, based on each player's best response to a set of beliefs about the actions of other players. In the first step, these beliefs are treated as data. In the second step, the new parameter estimates are used to update the beliefs. The two-step process is iterated until both beliefs and the parameters converge, at which point players will be playing best responses given their beliefs and beliefs will be consistent with players' strategies (i.e., strategies and beliefs will constitute a Bayesian Nash Equilibrium). The method contrasts with a nested fixed point procedure where an equilibrium is found at each step, and it also extends to cases where there is a finite mixture of unobserved heterogeneity.

Throughout much of this section we simplify the model that we are trying to estimate for expositional purposes. One important simplification we make is that we observe data generated by an second price auction. Our methodology is easily extended to cases where the data generating process is an English auction or a more general open outcry auction where bidders may not bid up to their values. We develop these extensions more fully in the appendix. Finally, although we ignore reserve prices here, it is straightforward to include them, and we do in our empirical application.

Before describing the details of the method, we outline some additional notation. We then present the estimation method with multiple (observed) types of potential entrants, but with no unobserved heterogeneity between auctions. Then we describe how the routine changes when there is unobserved heterogeneity between auction types, before making some comments about how multiple equilibria may affect estimation.

4.1 Notation

To keep the notation simple, we will assume that $\overline{\tau} = 2$ (two types of potential entrants) for each auction and that within a type, all bidders are symmetric and use symmetric entry strategies. The type of each potential entrant is observed to the econometrician. Initially, we will also assume that all T observed auctions are identical i.e., for each auction bidders values are drawn from the same distributions F_τ^V (pdf f_τ^V), where $F_\tau^V = N(\mu_\tau, \sigma_V^2)$ and that N_τ is the same across auctions. The distribution of the signal noise around the true value is also normal, i.e., $F^\varepsilon = N(0, \sigma_\varepsilon^2)$. We will also assume that σ_v^2, σ_ε^2 and K are the same across bidder types, although it is straightforward to relax this assumption. The complete set of parameters to be estimated is $\theta \in \{\mu_1, \mu_2, \sigma_v^2, \sigma_\varepsilon^2, K\}$.

To explain the estimation procedure, some additional notation will be useful. When a type τ bidder receives a signal s, his posterior belief is that his true value is distributed $F_\tau'(.|s,\theta) = N(\alpha\mu_\tau + (1-\alpha)s, \sigma_V'^2)$, where $\alpha = \frac{1/\sigma_V^2}{1/\sigma_V^2 + 1/\sigma_\varepsilon^2}$ and $\sigma_V' = \sqrt{\frac{1}{1/\sigma_V^2 + 1/\sigma_\varepsilon^2}}$. $G_\tau^{-i}(.)$ denotes the beliefs that a type τ bidder has about the distribution of the highest bid that will be placed by *other* bidders (hence the $-i$) when he makes his entry decision. $g_\tau^{-i}(.)$ is the associated pdf. $s_\tau'(G_\tau^{-i}, \theta)$ is the signal which makes a type τ bidder indifferent about entering given the parameters and his beliefs. $s_\tau'(G_\tau^{-i}, \theta)$ is defined implicitly by the following equation, because the marginal entrant's expected profits from entering must be equal to zero:

$$\int_0^\infty \left(vG_\tau^{-i}(v) - \int_0^v v' g_\tau^{-i}(v')dv' \right) f_\tau'(v|s_\tau', \theta)dv - K = 0 \tag{1}$$

Given G_τ^{-i}, θ and f_τ', s_τ' can be solved for efficiently using a standard non-linear solver.

For each potential entrant i, we observe either that he does not enter or that he enters and submits a bid equal to b. Denote the observed action $a_{i\tau}$, where $a_{i\tau} = 0$ if there is no entry and $a_{i\tau} = b_{i\tau}$ otherwise. Given (G_τ^{-i}, θ), the probability that a type τ potential entrant i does not enter is:

$$\Pr(a_{i\tau} = 0|G_\tau^{-i}, \theta) = \int_0^\infty F^\varepsilon(s_\tau'(G_\tau^{-i}, \theta) - v|\sigma_\varepsilon^2) f_\tau^V(v)dv \tag{2}$$

The probability (pmf) that a type τ potential entrant i enters and bids $b > 0$ (his value) is:

$$\Pr(a_{i\tau} = b|G_\tau^{-i}, \theta) = \left(1 - F^\varepsilon(s_\tau'(G_\tau^{-i}, \theta) - b|\sigma_\varepsilon^2)\right) f_\tau^V(b|\theta) \tag{3}$$

In the appendix we outline the action probabilities stemming from other, less restrictive modeling assumptions. For example, we can drop the assumption that we observe the winning bidder's bid, which would be the case in an English Button auction. Alternatively, we could loosen the restriction that we observe all bidders' values, as might be the case in an open outcry auction (see for example Haile and Tamer (2003)). These alternative action probabilities can be used instead to form the pseudo-likelihood used in our estimation.

4.2 Estimation with No Unobserved Heterogeneity

Before estimation begins, we specify initial guesses of $G_\tau^{-i}(.)$ for each type of player. With no unobserved auction heterogeneity, it is straightforward to estimate these distributions from the data using either parametric or non-parametric techniques.

The iterative pseudo-likelihood procedure has two steps. For a particular iteration k, the two steps proceed as follows:

Step 1 (maximum pseudo-log likelihood estimation). In this step the parameters θ are estimated using the best response probabilities (2) and (3) given values for $G_\tau^{-i}(.)$. Formally:

$$\widehat{\theta^k} = \arg\max_\theta \sum_{t=1}^{T} \sum_{\tau=1,2} \sum_{i=1}^{N_\tau} \log \Pr(a_{i\tau t} | G_\tau^{-i,k-1}, \theta) \tag{4}$$

For each value of the parameters, the pseudo-likelihood is calculated by solving for $s'_\tau(G_\tau^{-i}, \theta)$ for each player type, and then calculating the probability of the observed action for each potential entrant.

Step 2 (update G). In this step the parameter values and the final values of $s'_\tau(G_\tau^{-i}, \theta)$ from Step 1 are used to update the G_τ^{-i} distributions. For a bidder of type 1, the probability that the highest bid of other bidders is less than some value x is

$$\begin{aligned} \widehat{G_1^{-i,k}(x)} = &\left(\int_0^\infty F^\varepsilon \left(s'_1 - v | \widehat{\theta^k} \right) f_1^V \left(v | \widehat{\theta^k} \right) dv + \int_0^x \left(1 - F^\varepsilon \left(s'_1 - v | \widehat{\theta^k} \right) \right) f_1^V \left(v | \widehat{\theta^k} \right) dv \right)^{N_1 - 1} \\ &\times \left(\int_0^\infty F^\varepsilon \left(s'_2 - v | \widehat{\theta^k} \right) f_2^V \left(v | \widehat{\theta^k} \right) dv + \int_0^x \left(1 - F^\varepsilon \left(s'_2 - v | \widehat{\theta^k} \right) \right) f_2^V \left(v | \widehat{\theta^k} \right) dv \right)^{N_2} \end{aligned} \tag{5}$$

Steps 1 and 2 are iterated until both G and θ converge.[9]

In practice, it was found that the algorithm converged more quickly if G^{-i} was updated only partially in Step 2, i.e., we use a convex combination of $G_1^{-i,k-1}(x)$ and $\widehat{G_1^{-i}(x)}$.[10]

This description assumes that there is no observed heterogeneity between auctions. In practice, one could parameterize μ as a function of observed covariates, e.g., $\mu_1 = X\beta$ and $\mu_2 = X\beta + \gamma$, as is done elsewhere in the literature (see for example Athey, Levin, and Seira (forthcoming)) and could allow for a different number of potential entrants in different auctions. In this case, we alter the computational routine to use a separate G_τ^{-i} and $s'_\tau(G_\tau^{-i}, \theta)$ for each distinct set of auction covariates.

[9] In practice we set the tolerance level to 1.0E-6.

[10] In practice we only update 20% each time, i.e. we place weight 0.2 on the new estimate of the distribution.

4.3 Estimation with Unobserved Auction Heterogeneity

We now consider the case with unobserved heterogeneity. The idea is to specify a set of potential auction types, which for any one auction would not be observed by the econometrician, and recover the frequency that each type appears in the data. For simplicity, suppose that there are two types of auctions $\vartheta = 1, 2$, and the econometrician does not observe the type of auction but the potential entrants do. The mean value of a bidder of type τ in an auction of type ϑ is $\mu_{\vartheta,\tau}$, where $\mu_{1,1} - \mu_{2,1} = \mu_{1,2} - \mu_{2,2} = c > 0$ (i.e., the difference in mean values is common for both entrant types and, as a normalization, auctions of type 1 are more valuable). The payoff-related parameters are now $\theta \in \{\mu_1, \mu_2, c, \sigma_v^2, \sigma_\varepsilon^2, K\}$. The proportion of auctions of type 1 is λ, an additional nuisance parameter. The various distributions (F^V and G^{-i}) are now defined for each auction as well as each bidder type.

Before estimation begins, we specify initial guesses of $G_{\vartheta\tau}^{-i}(.)$ for each type of player and each type of auction. With unobserved auction heterogeneity, estimation of these distributions from the data could be done by estimating parametric finite mixture distributions (e.g., a mixture of Weibull or normal distributions). In practice, this would only be feasible for a relatively small number of unobserved auction types (e.g., less than four), although this could well be sufficient once observable auction heterogeneity is controlled for.

As before, the iterative procedure involves maximizing a pseudo-likelihood (Step 1) and updating the G distributions (Step 2). Step 2 is exactly the same as before, although it must now be done for each $G_{\vartheta\tau}^{-i}$. The pseudo-log likelihood on iteration k now reflects the mixture of unobserved auction types:

$$\sum_{t=1}^{T} \log \left(\lambda \prod_{\tau=1,2} \prod_{i=1}^{N_\tau} \Pr(a_{i\tau t} | G_{1,\tau}^{-i,k-1}, \theta) + (1-\lambda) \prod_{\tau=1,2} \prod_{i=1}^{N_\tau} \Pr(a_{i\tau t} | G_{2,\tau}^{-i,k-1}, \theta) \right) \tag{6}$$

Maximizing (6) directly can be difficult, so we use an alternative EM-algorithm approach. This involves iterating two steps, constituting an inner loop as part of Step 1. For inner loop iteration l, the first (E-) step involves calculating the conditional expectation of the 'complete information' pseudo-log likelihood given the parameters and the data:

$$\sum_t \pi_t^{\vartheta_t=1,k,l} \left(\log \lambda + \sum_{\tau=1,2} \sum_{i=1}^{N_\tau} \log \Pr(a_{i\tau t} | G_{1,\tau}^{-i,k-1}, \theta) \right) + \tag{7}$$
$$\sum_t \left(1 - \pi_t^{\vartheta_t=1,k,l}\right) \left(\log(1-\lambda) + \sum_{\tau=1,2} \sum_{i=1}^{N_\tau} \log \Pr(a_{i\tau t} | G_{2,\tau}^{-i,k-1}, \theta) \right)$$

where the weights π are calculated by:

$$\pi_t^{\vartheta_t=1,k,l} = \frac{\widehat{\lambda^{k,l-1}} \prod_{\tau=1,2} \prod_{i=1}^{N_\tau} \Pr(a_{i\tau t}|G_{1,\tau}^{-i,k-1}, \widehat{\theta^{k,l-1}})}{\left(\begin{array}{c} \widehat{\lambda^{k,l-1}} \prod_{\tau=1,2} \prod_{i=1}^{N_\tau} \Pr(a_{i\tau t}|G_{1,\tau}^{-i,k-1}, \widehat{\theta^{k,l-1}}) + \\ \left(1 - \widehat{\lambda^{k,l-1}}\right) \prod_{\tau=1,2} \prod_{i=1}^{N_\tau} \Pr(a_{i\tau t}|G_{2,\tau}^{-i,k-1}, \widehat{\theta^{k,l-1}}) \end{array} \right)} \tag{8}$$

The second (M-) step involves finding the parameters that maximize (7). For inner loop iteration l, the estimate of λ is

$$\widehat{\lambda^{k,l}} = \frac{\sum_t \pi_t^{\vartheta_t=1,k,l}}{T}$$

while the values of $\widehat{\theta^{k,l}}$ are found numerically, and the E- and M-steps are iterated until convergence of both θ and λ. A common problem with EM procedures is that converge slowly. However, in our setting, we found that convergence was achieved with a small number of iterations.

4.4 Multiple Equilibria

When there are multiple types of potential entrants, the auction entry game may have multiple equilibria. This is true even if we restrict the strategies to be type symmetric. For example, if there are two types and one potential entrant of each type, and entry is expensive, there may be two equilibria where one potential entrant is much more likely to enter even when the types have similar value distributions.[11] If there is no unobserved auction heterogeneity and the same equilibrium is played in every observation of an auction with the same observables, the existence of multiple equilibria in the game may not present a problem for estimation because the initial estimation of a player's beliefs about the bids of other players (G^{-i}) will help to pick out the equilibrium that is played in the data. This argument is commonly made in the literature on estimating dynamic games using two-step methods. (See for example Aguirregabiria and Mira (2007), Bajari, Benkard, and Levin (2007), and Pesendorfer and Schmidt-Dengler (2009).) With unobserved auction heterogeneity, one could make similar arguments when there are small numbers of types, based on estimates of a first-stage mixture model.

4.5 Methods and Monte Carlos

In this section we present several Monte Carlo experiments to illustrate our estimation method. We begin with the case without unobserved heterogeneity but with asymmetric bidders. We consider two bidder types who differ in their mean value. Table 5 displays the results from several experi-

[11] These equilibria may also be efficient from the perspective of the firms because they minimize the potential for a coordination failure where both firms enter.

ments. The first three rows illustrate how we can recover the parameters of interest as the number of potential bidders varies for a reasonable amount of variation in the signal. We believe that one benefit of the model is that it is general enough to approximately match the two polar entry cases used in the literature so far. The next two rows attempt to illustrate this. They approximate a data generating process akin to the S and LS models, respectively. In the fourth row $\sigma_\varepsilon = 0.55$ and so bidders almost perfectly know their value prior to paying the entry cost. In the final row we set $\sigma_\varepsilon = 55$ and so bidders have almost no information regarding their value prior to entry. In either case we can recover the underlying parameters quite well.

[Table 5 about here.]

We now add unobserved heterogeneity to the model and attempt to recover the model parameters. The results of these experiments appear in Table 6. We show two examples with unobserved heterogeneity. The first has two types and the second has three types. As above, we assume that we know c and estimate the weights. The results show that our estimator works well in recovering all parameters, including the distribution of unobserved heterogeneity, for both cases.

[Table 6 about here.]

5 Empirical Application

We now illustrate how ignoring selection into auctions matters and how to implement our methodology in an empirical setting. We focus on federal auctions of timberland in California. We first describe the data, then discuss why our model applies to the setting, then provide evidence of selection in these auctions, then present our estimation results and finally perform counterfactuals of interest.

5.1 Data and Context

We focus on federal auctions of timberland in California.[12] In these auctions the U.S. Forest Service (USFS) sells logging contracts to individual bidders who may or may not have manufacturing capabilities (mills and loggers, respectively). When the sale is announced, the USFS provides a "cruise" estimate of the volume of timber on the tract. It also announces a reserve price and while bidders must submit a bid of at least this amount to qualify for the auction, it is generally viewed as non-binding.[13] After the sale is announced, bidders are able to perform their own private cruises of the tract to assess its value. These cruises can be informative about the tract's volume, species make-up and timber quality. Finally, bidders must post a deposit of 10% of the appraised value of the tract in order to be eligible to participate in the auction.

As in our model above, we assume that bidders have independent private values. This assumption is also made in other work with similar timber auction data (see for example Baldwin,

[12] We are very grateful to Susan Athey, Jonathan Levin and Enrique Seira for sharing their data with us.

[13] See Haile (1996) for more details.

Marshall, and Richard (1997), Brannman and Froeb (2000), Haile (2001) or Athey, Levin, and Seira (forthcoming)). A bidder's private information is primarily related to its own contracts to sell the harvest, inventories and private costs of harvesting and thus is mainly associated with its valuation only.

In our model we allow bidders to receive an imperfect signal about their value prior to paying an investment cost to fully learn their value. There are multiple reasons why it is likely that in these timber auctions bidders only have imperfect knowledge about their value prior to entry. First, each tract is unique and therefore even if a bidder has previously bid on apparently similar tracts, they must still account for heterogeneity not realized prior to further investigation. Second, the cruise estimates published by the USFS may be imprecise. One industry source told us that they may misestimate the volume of timber by as much as 40%. Finally, bidders must also devote time to planning and organizing their team of harvesters and lining up potential end users for any given tract. These are likely to be only a few of the necessary investments a bidder must make prior to learning its true value for a particular stand of timber. Therefore, a model which at least allows bidders to have a noisy signal about their value, but can still permit this signal to be precise, seems warranted.

We should note here that our selection model differs from Athey, Levin, and Seira (forthcoming) who use similar data but apply the LS model. They allow for two types of bidders (mills and loggers) and their model assumes that mills' value distribution stochastically dominates, according to a hazard rate order, that of loggers. Their proposition 1 states that if the necessary condition for a unique type-symmetric equilibrium is met, then mills must enter with probability 1.[14] In estimation the authors confirm that the necessary condition is met for each tract in their data.

From the original data we focus on the most appropriate auctions. As our methodology so far addresses ascending or second price auctions, we exclude first price auctions during this time. We also eliminate small business set aside auctions, salvage sales and auctions with extremely low or high acreage or volume to acreage ratios as these are likely either outliers or coding errors in the data. Finally, we examine auctions between 1982 and 1989 to reduce resale concerns that might complicate the analysis (see Haile (2001) for an analysis of these auctions with resale). Resale was limited after 1981 because third party transfers (i.e. the winner transferring the right to harvest the timber) were prohibited and speculative bidding was reduced due to shortened contract lengths, larger required deposits, greater penalties for default and increased difficulty of obtaining contract extensions (Mead, Schniepp, and Watson (1983)). This is important because another model with resale like that of Haile (2001) could lead to increased bidding with increased competition, a comparative static also consistent with a selection model. We are left with 988 auctions over this period.

In addition to the USFS data, we add data on (seasonally adjusted, lagged) monthly housing starts and firm locations. The firm location data (NETS data) was purchased from Walls and Associates who obtain the data from Dunn and Bradstreet. For bidders we can identify in the

[14]The other alternative is that loggers enter with probability 0. This is robustly rejected in their data.

NETS data, we obtain their latitude and longitude and this enables us to determine their distance from any auction. We are able to match 43.3% of firms but these firms account for 70.5% of bids and 71.0% of winning bids.

We summarize the data in Table 7. Bids are given in $/mbf (1982 dollars). We see that bids submitted by loggers tend to be lower than those submitted by mills, consistent with the results in Athey, Levin, and Seira (forthcoming), but in our sample loggers win less often than in theirs. We define entrants to be the set of bidders we observe at the auction even if they did not submit a bid above the reserve price. We count the number of potential entrants as those bidders who bid within 50 km of an auction over the next month. Potential *other* entrants are the subset of these that excludes bidders who bid in this auction but not in those within 50 km over the next month. For our estimation we will assume that all entrants paid the fixed cost of entry.

Fewer loggers than mills enter on average, but they are also less likely to enter. Among the set of potential logger entrants, on average 34% enter, whereas on average 66% of potential mill entrants enter. Finally, 4% of tracts failed to sell because they received no bids.

[Table 7 about here.]

Our model assumes that within type, bidders may have heterogenous values but not entry costs. As mentioned in Section 3.3, if we believe that bidders do receive noisy signals about their value prior to entry, we can appeal to the panel nature of our data set to investigate whether there appears to be large, persistent differences in entry costs across bidders. This is because large differences in entry costs would cause those bidders with high entry costs to stay out of auctions unless they receive quite an optimistic signal about their value, in which case they will be likely to win. In the data, then, we can examine whether bidders who rarely enter are more likely to win when they do enter. We begin by focusing on bidders who are often within a reasonable distance from any auction and divide these into two groups based on how often they bid. We then will see whether the group which rarely bids is more likely to win than those who bid often. If this is true, then it is suggestive of persistence in entry costs as those bidders who rarely bid do so because of high entry costs, not low values. To proceed, we isolate a set of bidders to be those who on average are within 76.5 km of an auction, the median distance in the data, and look at the probability of winning conditional on a bidder bidding in more than 15 auctions compared to those who bid in between 5 and 15 auctions. We find that the former wins 61.2% of the time they enter and the latter wins only 14.5% of the time they enter. Therefore, this is suggestive evidence that there is not a great deal of persistent heterogeneity in entry costs across bidders.

We stated earlier that government cruise reports are often inaccurate and thus bidders are incentivized to invest in learning their true value for the tract. We can support this claim in our data because for a sample of "scaled" sales we have data on the timber that was actually cut by the winner.[15] Using this information, we can evaluate the quality of the government's estimates. On average, the government overestimates the amount of timber on the tract 60% of the time. The top panel of Figure 2 displays the distribution of these misestimates in percentage terms.

[15]This is the same data used in Athey and Levin (2001).

[Figure 2 about here.]

It is possible that what matters more to bidders is the government's estimate of the distribution of species type on any tract. One way to gauge their accuracy on this dimension is to compare the share of the volume the most prevalent species supposedly commanded with what it actually did. The bottom panel of Figure 2 displays the distribution of these misestimates in percentage terms. To interpret the figure, 10 means that the share of the (supposedly) most prevalent species was estimated to be 10% higher than it actually was.

Given the potential error in the government's estimates, and the inconsistency of this error, bidders likely to find it valuable to undertake their own investment to more precisely learn about their value.

5.2 Evidence of Selection

In this section we provide evidence that actual bidders in an auction are not a random sample of bidders from the data. We do this by examining the impact of the number of potential entrants on actual entry and submitted bids.

As shown in Section 3.2, in the LS model revenues are likely to decline as the number of potential entrants increases. In the LS model, the revenue falls as potential entry increases due to either the number of entrants falling on average, or there being a greater chance of having sufficiently low entry to lower revenues. We therefore begin providing evidence that selection exists in the data by showing that entry is predicted to rise with potential entry on average and that there is no evidence of a greater chance of low entry numbers with more potential entry.

Table 8 shows regressions of the (log of) entry on the (log of) potential entry and other auction covariates. We are careful to define potential other entry here as those bidders who bid on other auctions nearby within a short amount of time and thus exclude bidders who bid on this auction but never bid on other similar auctions. As shown in the table, as potential other entry increases by 10%, actual entry increases by almost 3%.

[Table 8 about here.]

While the fact that entry rises with the number of potential entrants is consistent with a selection model, it is also possible that this can hold on average in a model without selection. If this is true, it is the greater chance of getting sufficiently low entry as the number of potential entrants increases which drives the falling revenue in the LS model. To test whether there is evidence of an increased likelihood of low entry as potential entry increases, we test whether the likelihood of having fewer than M entrants increases as N potential entrants increases to $N+1$ for $M = 3, 4$ and $N = 5, 6, 7, 8, 9$. In each case the probability falls, thus providing more evidence that a model without selection, which would rely on this probability rising to generate falling revenues, is unlikely to be the true data generating process.

Another test for selection is whether the average valuations of bidders rise as potential entry increases. If there is no selection, then bidders are a random sample from the population regardless of the number of potential entry. This is the essential idea behind the test in Marmer, Shneyerov, and Xu (2007). If we believe that bidders' strategies in the auctions examined here are the same as in a classic English Button Auction with independent private values, then we can consider bids submitted to be bidders' valuations. Thus we can examine whether valuations increase in potential entry by looking at the submitted bids as potential entry increases. These results appear in Table 9. Examining the first four columns, we find that a 10% increase in potential entry leads to a 1.6% increase in the submitted bid (0.9% when we control for distance), thus providing evidence of selection. A confounding factor is that there may be unobserved auction heterogeneity that is driving both increased potential entry and submitted bids. That is there may be factors observable to the bidders but not to the econometrician that affect bidder behavior. To control for this we follow a strategy similar to Haile (2001) and instrument for the number of potential bidders by the number of mills who bid in the same forest during the preceding six months. If we believe mill location and activity is determined well before a particular auction's unobservable (to the researcher) quality is realized, then this is a valid instrument. The estimates employing this instrument appear in the last four columns. We continue to find a positive impact of potential entry on submitted bids. In fact, the impact increases by almost three-fold.[16] While there may be some concern about this instrumental variable strategy, in our structural estimation, we do not rely on this method to control for unobserved heterogeneity. Here we are simply interested in providing suggestive evidence that there is selection at the entry stage for these timber auctions.

[Table 9 about here.]

There is some concern about interpreting bids as values in these auctions since they are not exactly English Button Auctions but rather open outcry auctions (see for example Haile and Tamer (2003)). If we assume that any bid submitted is a value, it is safest to assume that the highest bid is the valuation of the second highest valuation bidder. Therefore, to further push on whether we can find evidence of selection, we repeat the same test using only the winning bid. The results appear in Table 10 and corroborate the evidence just presented using all bids.

[Table 10 about here.]

Another way to investigate the amount of selection in the data is to analyze the impact of potential entry on revenues. We know from above that models without selection will predict that revenues fall as potential entry increases. Given that so few auctions fail to sell, the results are virtually identical to those in Table 10, thus lending support to a model with selection.

Finally, we can test for evidence of selection by applying a Heckman selection model (Heckman (1976)). Given that potential competition effects a bidder's decision to enter an auction but not his bid conditional on entry, we can specify entry as a flexible function of covariates, including potential

[16]If we include the reserve price, following the arguments made in Roberts (2009), to control for unobserved heterogeneity, we also find evidence of selection. We are less enthusiastic about this route since it seems as though reserve pricing function used by the USFS may not always lead to reserve prices being monotonic in unobservable quality.

competition, at the first stage, and exclude competition at the second stage bid regression. The results appear in Table 11. In the first stage entry probit, we include potential other mill and logger entrants and incorporate them through a flexible polynomial. It is clear that we are finding positive selection in that entrants are likely to bid more aggressively. In addition, we see that the difference in logger and mill bids is masked by selection because the difference grows when we control for selection. The results are robust across a variety of specifications and restricting the sample to only winning bids.[17]

[Table 11 about here.]

After establishing the evidence of selection in the data, we now turn to estimating the full entry and bidding model in order to perform counterfactual analysis.

5.3 Estimates

In our empirical specification we assume that values are distributed lognormal. That is $S_\tau = V_\tau A_\tau$, $A_\tau = e^{\varepsilon_\tau}$, where $V_\tau \sim logN(\mu_\tau, \sigma^2_{V\tau})$, $\varepsilon_\tau \sim N(0, \sigma^2_{\varepsilon\tau})$ and $\tau \in \{Mill, Logger\}$. To control for observable heterogeneity, we assume that the Federal government does not take into account the number and types of potential bidders when setting reserve prices. We begin by collecting the residuals from a first stage regression of the log of the reserve price on tract characteristics. We then take these, add them to the mean (across auctions) of the projected log values from this regression and difference them from the log of the observed bid to homogenize our sample. We then exponentiate.

It is well known (Haile and Tamer (2003)) that interpreting bids as values is a strong assumption in open outcry auctions. In light of this, when forming the likelihood, we make the following assumptions about bidder values and entry. The winning bidder had a value greater than the second highest bid. The second highest bidder had a value equal to his bid. All other bidders had a value greater than the reserve but less than the second highest bid. Finally, all potential bidders that did not submit a bid had a signal less than s'. Alternative assumptions could be made. For example, we might assume that the second highest bidder has a value less than the winning bid. In practice, however, 96% of the time the difference between these bids is less than 1% of the high bid.

We further restrict the sample to exclude outliers and so we only focus on observations where the winning bid was between \$10/mbf and \$240/mbf and where there were fewer than 20 potential bidders. Our estimates of the lognormal parameters $\{\mu_\tau, \sigma_{V\tau}\}$, $\sigma_{\varepsilon\tau}$ and K_τ, for $\tau \in \{Mill, Logger\}$ appear in Table 12. We calculate the standard errors following Aguirregabiria and Mira (2007) Proposition 2. Our estimates suggest a substantial difference in values between mills and loggers and that there is a fair amount of selection in the data.

[Table 12 about here.]

The value distributions for both types are appear in Figure 3.

[17] Although not shown in the table, the results are also robust to changing the regressions from logs to levels.

[Figure 3 about here.]

The mean of the logger values ($30/mbf) is less than half that of mills ($64/mbf). Recall that the mean submitted logger bid was only 22% less than that of mills. Therefore we confirm the intuition that selection at the entry stage masks true differences based on bidding data alone.

With our parameter estimates, if we assume a set of potential bidders and reserve price, we can compare the entrants' and marginals' (those who observed s') value distributions. For the case where the reserve and the number of potential mill and logger entrants are set to their respective means of $37/mbf, four and five, this comparison for mills and loggers appears in Figure 4.

[Figure 4 about here.]

As predicted by the Signal model, there is a stark difference in the marginal and infra-marginal bidders for each type. For mills, the mean value of entrants in these auctions is $101/mbf. For marginal mills it is $34/mbf. For loggers, the mean value of entrants in these auctions is $66/mbf. For marginal loggers it is $38/mbf. These findings suggest that appreciating the differences between participating and marginal bidders will be important for policies like entry subsidies and set-asides. To highlight this point, we now investigate such counterfactual scenarios.

5.4 Counterfactuals: Entry Subsidies

As a preliminary illustration of how selection will affect counterfactual analysis in the data, we analyze the value of an entry subsidy. Specifically, we evaluate the impact of encouraging the marginal logger to participate in our model. If we force this bidder, who receives as his signal s', to enter an auction with 4 mills, 5 loggers and a reserve price of $37, then we find that revenues rise approximately 1.38% on average. If, on the other hand, we believed that the marginal logger's value distribution was the same as the participants', then we would predict a much higher boost to revenues on average: 10.73%. This accords with the intuition that the true marginal bidder is one who received a signal predicting a much lower value than an actual entrant. This finding suggests that the value of entry subsidies and small bidder programs may be much less than previously appreciated when selection is ignored.

6 Conclusion

In most of the empirical auction literature, the econometrician assumes that the bids observed in the data are submitted by a random sample of bidders. In this paper we take this assumption to task. Specifically, we develop and estimate an independent private value auction model where potential bidders receive an imperfect signal about their value prior to making a costly entry decision. The estimation allows for asymmetries in bidders and unobserved, to the econometrician, object heterogeneity.

We illustrate how incorrectly assuming that the current leading models of entry are the data generating processes will (i) cause misestimates of model primitives and (ii) generate bias in important counterfactual analyses. We also highlight that a feature of models without selection, namely that seller revenues decrease in the number of potential bidders, can be used to test for the presence of selection. This is because seller revenues can increase with potential entry in the more general Signal model.

After explaining out estimation method and exploring its effectiveness via Monte Carlo studies, we turn to our empirical application. We apply our model to timber auctions to demonstrate how ignoring selection in bidder participation can affect an important and often studied empirical setting. Specifically, we show that the set of participating bidders appears to be a set of bidders for whom paying the entry cost was attractive because they were likely to have high valuations. That is the set of participating bidders does not appear to be a representative sample of the set of potential bidders.

Our structural estimates suggest that potential bidders have quite precise estimates of their value prior to paying an entry cost. This results leads us to conclude that the value of entry subsidies and small bidder assistance programs may be overstated.

Appendix

Here we outline how we can adapt the estimation procedure to loosen our restrictive assumptions about the data generating process.

English Button Auction

If we assume that the data generating process is the classic English Button Auctions where an auctioneer continuously raises the price and bidders press a button signaling their willingness to participate, we can interpret bidders' bids as their dropout points, which are their values. Thus any bid we observe is a bidder's value. However, we do not observe the winning bidder's value since we wins when the bidder with the second highest value removes his finger from the button. Therefore, while the probabilities of a bidder not entering and submitting an observed bid are as in Equations 2 and 3, respectively, we now lose the information that observing the winner's bid provides. Instead, we incorporate the probability of the winning bidder having a value in excess of the second highest bid b_2 by:

$$\Pr(a_{i\tau} = win|G_{\tau}^{-i}, \theta) = \int_{b_2}^{\infty} \left(1 - F^{\varepsilon}(s_{\tau}'(G_{\tau}^{-i}, \theta) - v|\sigma_{\varepsilon}^2)\right) f_{\tau}^{V}(v)dv \tag{9}$$

General Open Outcry Auction

The English Button auction, as imagined by Milgrom and Weber (1982), is obviously a simplification of what often are open outcry auctions. In these auctions we may worry that submitted bids are not actually bidders' values. Instead, a very general set of assumptions one can make upon observing a set of bids from a set of potential bidders, where the second highest bid b_2 are:

A1 The winning bidder had a value greater than b_2.

A2 The losing bidder that bid the most had a value equal to b_2.

A3 All participating bidders had values less than b_2 but greater than the reserve price r.

Probability $\Pr(a_{i\tau} = win|G_{\tau}^{-i}, \theta)$ is as in Equation 9. The probability that the highest losing bidder had a value equal to b_2 is as in Equation 3. However, for those bidders for which we observe a losing bid, we now say:

$$\Pr(a_{i\tau} = b \in [r, b_2)|G_{\tau}^{-i}, \theta) = \int_{r}^{b_2} \left(1 - F^{\varepsilon}(s_{\tau}'(G_{\tau}^{-i}, \theta) - v|\sigma_{\varepsilon}^2)\right) f_{\tau}^{V}(v|\theta) \tag{10}$$

References

Aguirregabiria, V., and P. Mira (2002): "Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models," *Econometrica*, 70(4), 1519–1543.

——— (2007): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75(1), 1–53.

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

Athey, S., and J. Levin (2001): "Information and Competition at U.S. Forest Service Timber Auctions," *Journal of Political Economy*, 109(2), 375–417.

Athey, S., J. Levin, and E. Seira (forthcoming): "Comparing Open and Sealed Bid Auctions: Theory and Evidence from Timber Auctions," *Quarterly Journal of Economics*.

Bajari, P., L. Benkard, and J. Levin (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75(5), 1331–1370.

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

Baldwin, L., R. Marshall, and J.-F. Richard (1997): "Bidder Collusion at Forest Service Timber Sales," *Journal of Political Economy*, 105(4), 657–699.

Brannman, L., and L. Froeb (2000): "Mergers, Cartels, Set-Asides, and Bidding Preferences in Asymmetric Oral Auctions," *Review of Economics and Statistics*, 82(2), 283–290.

Einav, L., and I. Esponda (2008): "Endogenous Participation and Local Market Power in Highway Procurement," *Working Paper*.

Ertaç, S., A. Hortaçsu, and J. Roberts (forthcoming): "Entry into Auctions: An Experimental Analysis," *International Journal of Industrial Organization*.

Haile, P. (1996): "Auctions with Resale Markets: An Application to U.S. Forest Service Timber Sales," *SSRI Working Paper 9702*.

——— (2001): "Auctions with Resale Markets: An Application to U.S. Forest Service Timber Sales," *American Economic Review*, 91(3), 399–427.

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

Heckman, J. (1976): "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic aand Social Measurement*, 5, 475–492.

Hendricks, K., J. Pinkse, and R. Porter (2003): "Empirical Implications of Equilibrium Bidding in First-Price, Symmetric, Common Value Auctions," *Review of Economic Studies*, 70(1), 115–145.

Hendricks, K., and R. Porter (2007): *An Empirical Perspective on Auctions*vol. 3 of *Handbook of Industrial Organization*, chap. 32. North-Holland.

Hu, Y., D. McAdams, and M. Shum (2009): "Nonparametric Identification of Auction Models with Non-Separable Unobserved Heterogeneity," *Working Paper*.

Hubbard, T., and H. Paarsch (2009): "Investigating Bid Preferences at Low-Price, Sealed-Bid Auctions with Endogenous Participation," *International Journal of Industrial Organization*, 27, 1–14.

Klemperer, P. (2002): "What Really Matters in Auction Design," *Journal of Economic Perspectives*, 16(1), 169–189.

Krasnokutskaya, E. (2009): "Identification and Estimation in Highway Procurement Auctions Under Unobserved Auction Heterogeneity," *Review of Economic Studies, forthcoming*.

Krasnokutskaya, E., and K. Seim (2007): "Bid Preference Programs and Participation in Highway Procurement Auctions," *Working Paper*.

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

Li, T., and X. Zheng (2008): "Entry and Competition Effects in First-Price Auctions: Theory and Evidence from Procurement Auctions," *Review of Economic Studies, forthcoming*.

Marmer, V., A. Shneyerov, and P. Xu (2007): "What Model for Entry in First-Price Auctions? A Nonparametric Approach," *Working Paper*.

Mead, W., M. Schniepp, and R. Watson (1983): "Competitive Bidding for U.S. Forest Service Timber in the Pacific Northwest 1963-83," Washington, DC: U.S. Department of Agriculture, Forest Service.

Milgrom, P., and R. Weber (1982): "A Theory of Auctions and Competitive Bidding," *Econometrica*, 50(5), 1089–1122.

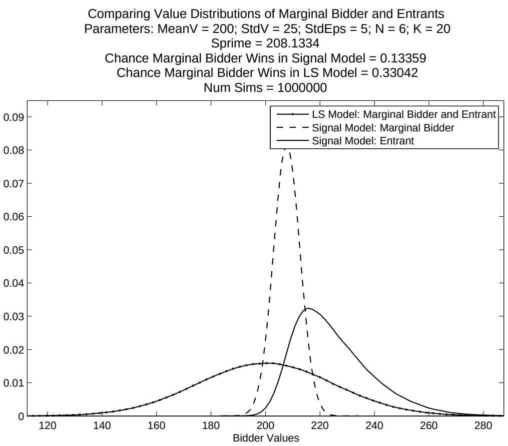
Paarsch, H. (1997): "Deriving an Estimate of the Optimal Reserve Price: An Application to British Columbia Timber Sales," *Journal of Econometrics*, 78, 333–357.

Palfrey, T., and S. Pevnitskaya (2008): "Endogenous Entry and Self Selection in Private Value Auctions: An Experimental Study," *Journal of Economic Behavior and Organizations*, 66, 731–747.

Pesendorfer, M., and P. Schmidt-Dengler (2009): "Sequential Estimation of Dynamic Discrete Games: A Comment," *Econometrica, forthcoming*.

Roberts, J. (2009): "Unobserved Heterogeneity and Reserve Prices in Auctions," *Working Paper*.

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



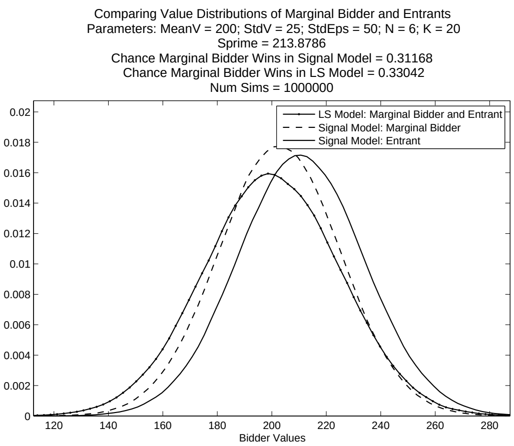


Figure 1: Comparing the value distributions for "marginal" bidders according to LS and Signal models.

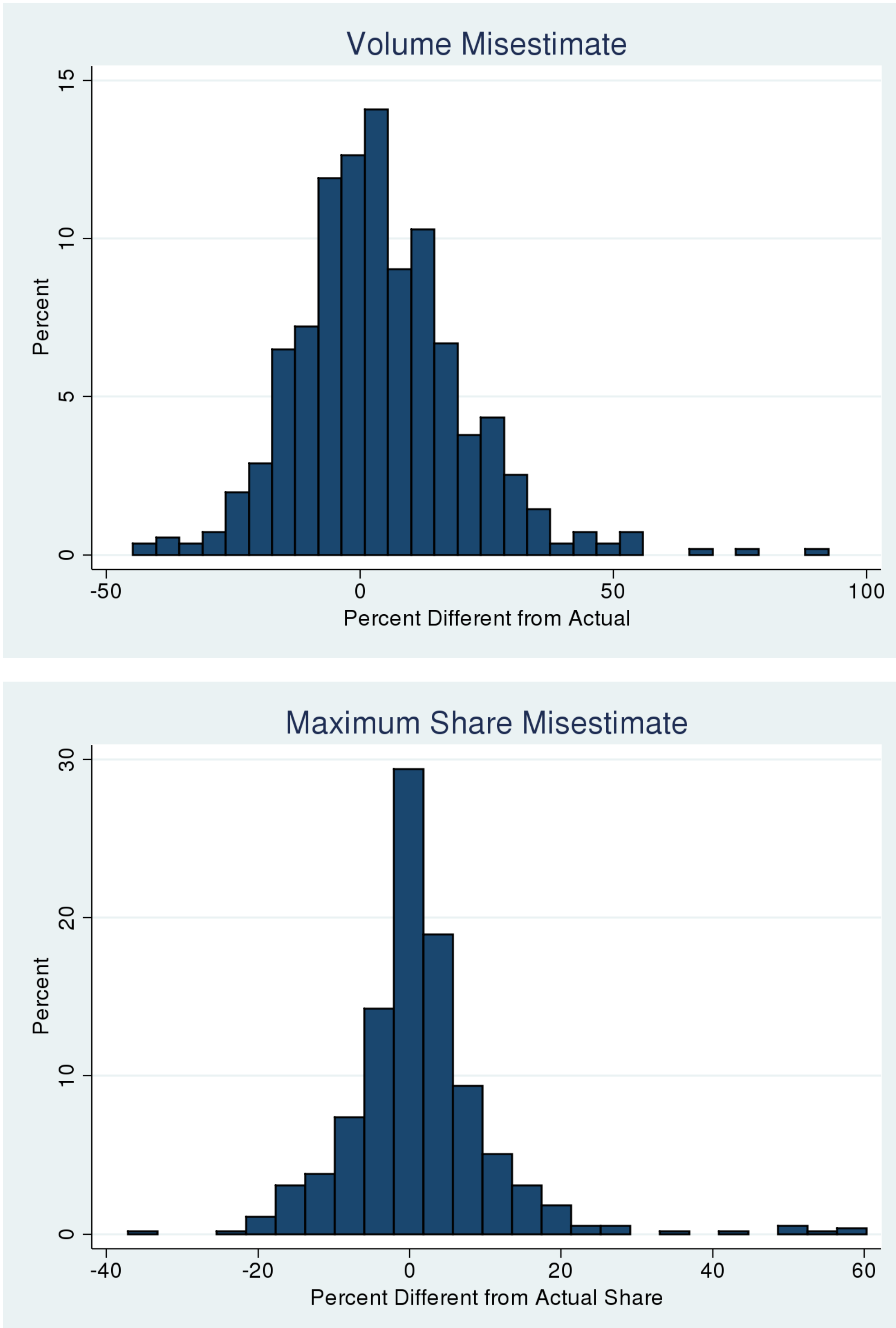


Figure 2: Evaluating the quality of the government's predictions.

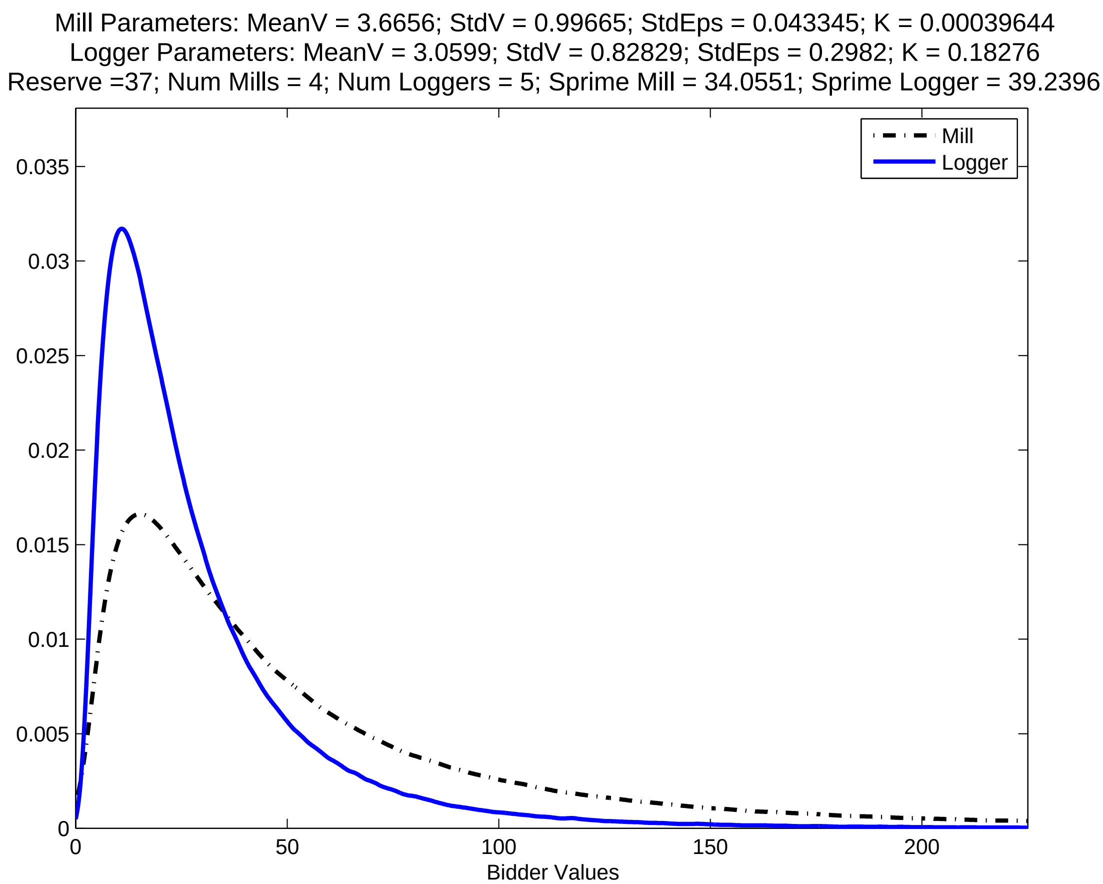


Figure 3: Comparing the value distributions for Mills and Loggers.

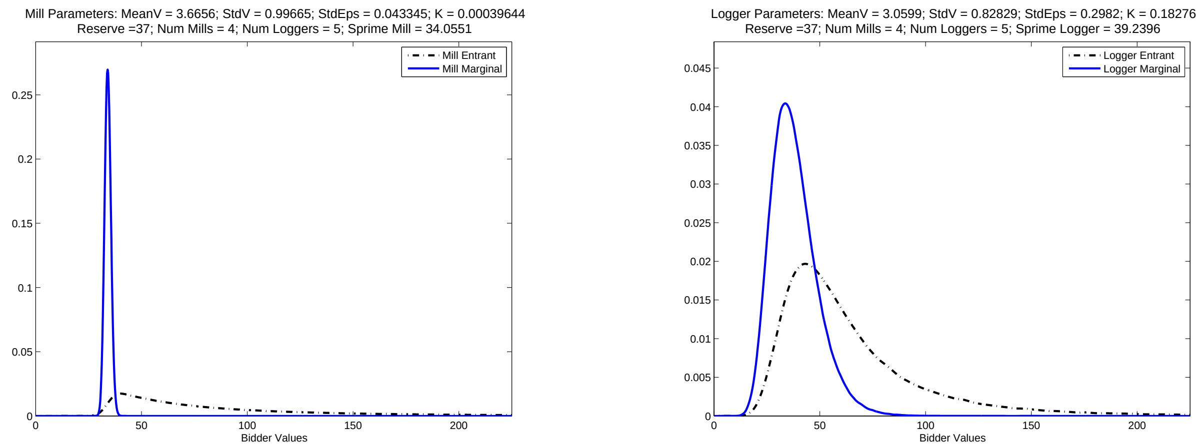


Figure 4: Comparing the value distributions for entrant and "marginal" bidders by type.

N	$\widehat{\mu}$	$\widehat{\sigma_V}$	$\widehat{K}$	s'
3	132.0222	17.8262	20.8270	106.8737
4	137.3831	16.2657	18.2699	116.1196
5	141.0022	15.4720	17.5082	122.1247
6	143.6077	14.5615	17.0921	126.4939
7	146.2489	14.2386	16.8575	129.8908
8	147.8688	13.4926	16.2295	132.6493
9	149.4981	13.2966	15.7235	134.9597
10	150.9570	12.9902	15.3284	136.9395

Table 1: Misestimating Demand, LS Model. The table displays estimates of model parameters assuming the LS model. Based on generating $T = 5000$ auctions where $\mu = 120$, $\sigma_V = 25$, $\sigma_\varepsilon = 5$, $K = 10$. We solve for s' as the equilibrium outcome for this game for each value of N.

N	$\widehat{\mu}$	$\widehat{\sigma_V}$	$\widehat{K}$	$\widehat{s'}$	s'
3	115.9863	29.7016	7.0017	97.5955	106.8737
4	113.1916	31.7532	7.4736	106.1968	116.1196
5	111.6691	32.5119	8.0142	112.9868	122.1247
6	109.6941	33.6069	8.2450	117.1782	126.4939
7	107.8251	34.5973	8.2982	120.2677	129.8908
8	106.2405	35.2924	8.3522	123.0219	132.6493
9	103.2874	36.9834	8.3002	124.0900	134.9597
10	103.3601	36.5966	8.2759	126.7540	136.9395

Table 2: Misestimating Demand, S Model. The table displays estimates of model parameters assuming the S model. Based on generating $T = 5000$ auctions where $\mu = 120$, $\sigma_V = 25$, $\sigma_\varepsilon = 5$, $K = 10$. We solve for s' as the equilibrium outcome for this game for each value of N. $\widehat{s'}$ is the equilibrium outcome for the S model for each value of N.

	Uniform		Log Normal		Weibull	
N	$E[n]$	$E[R]$	$E[n]$	$E[R]$	$E[n]$	$E[R]$
5	3.25	192.14	4.25	217.20	3.81	213.00
8	3.45	190.37	4.59	216.14	4.16	211.84
12	3.55	189.34	4.78	215.52	4.35	211.08

Table 3: Effect of Potential Entrants on Expected Revenue in LS Model. Though only some N are shown, the revenues decline consistently for all N in between. Expected revenue based on 1,000,000 simulated auctions. Uniform: $U[150, 250]$ (i.e. $E[V] = 200.00$, $Std[V] = 28.87$) and $K = 10$. Log Normal: $\mu = 5.34$, $\sigma = 0.1245$ (i.e. $E[V] = 210.00$, $Std[V] = 26.25$) and $K = 5$. Weibull: Scale = 220, Shape = 10, (i.e. $E[V] = 209.30$, $Std[V] = 25.18$) and $K = 5$.

		LS Model			Signal Model			
$r = 0$	N	p	$E[n]$	$E[R]$	s'	p	$E[n]$	$E[R]$
	3	0.9051	2.7154	191.7136	181.4393	0.7667	2.3001	178.3281
	4	0.7503	3.0013	190.0088	191.6932	0.6277	2.5109	183.2315
	5	0.6313	3.1567	188.8675	198.2535	0.5273	2.6365	186.8642
	6	0.5422	3.2535	188.0835	202.9833	0.4534	2.7205	189.6069
	7	0.4742	3.3191	187.5808	206.6376	0.3973	2.7811	191.6811
	8	0.4208	3.3666	187.1215	209.5913	0.3534	2.8271	193.4651
	9	0.3780	3.4024	186.7867	212.0559	0.3182	2.8634	194.9804
	10	0.3430	3.4303	186.6583	214.1617	0.2893	2.8929	196.4757
$r = 170$	N	p	$E[n]$	$E[R]$	s'	p	$E[n]$	$E[R]$
	3	0.7888	2.3664	192.7922	202.9425	0.4541	1.3622	192.9382
	4	0.5848	2.3392	192.2166	208.7208	0.3662	1.4646	197.4935
	5	0.4627	2.3136	191.9409	212.8566	0.3070	1.5352	200.7010
	6	0.3823	2.2939	191.7288	216.0356	0.2647	1.5881	203.0942
	7	0.3256	2.2789	191.6409	218.5976	0.2329	1.6300	205.1570
	8	0.2834	2.2672	191.5515	220.7321	0.2081	1.6645	206.8232
	9	0.2509	2.2578	191.4522	222.5549	0.1882	1.6935	208.2750
	10	0.2250	2.2502	191.4411	224.1410	0.1718	1.7185	209.5457

Table 4: Comparing Effects of Potential Entrants on Expected Revenue in LS and Signal Models. Expected revenue based on 1,000,000 simulated auctions. $F(\cdot)$ is Normal, $\mu_V = 200$, $\sigma_V = 25$, $K = 10$ and $\sigma_\varepsilon = 5$.

σ_ε	N_1	N_2	$\widehat{\mu}_1$	$\widehat{\mu}_2$	$\widehat{\sigma}_V$	$\widehat{\sigma}_\varepsilon$	$\widehat{K}$
5.00	1	5	209.4751	200.2048	24.8229	4.8584	10.1479
			(0.393)	(0.447)	(0.212)	(0.177)	(0.196)
5.00	2	5	209.6216	200.2784	25.0191	5.0690	10.2776
			(0.206)	(0.373)	(0.171)	(0.266)	(0.291)
5.00	3	5	209.7727	199.9918	25.0671	4.8339	10.2015
			(0.220)	(0.389)	(0.169)	(0.183)	(0.259)
0.55	2	4	209.5822	200.3087	25.0717	0.6748	10.3577
			(0.236)	(0.277)	(0.122)	(0.125)	(0.241)
55.00	2	4	209.5344	197.4346	25.1296	50.8864	9.7548
			(0.220)	(3.413)	(0.214)	(7.426)	(0.646)

Table 5: Recovering Parameters, No Unobserved Heterogeneity. The table displays estimates of model parameters assuming the correct Signal model. Based on generating $T = 5000$ auctions where $\mu_1 = 210$, $\mu_2 = 200$, $\sigma_V = 25$, $K = 10$. The cases when $\sigma_\varepsilon = 0.55$ and 55 reflect the S and LS models, respectively. Standard deviations in parentheses.

Case	N_1	N_2	$\widehat{\mu}_1$	$\widehat{\mu}_2$	$\widehat{\sigma}_V$	$\widehat{\sigma}_\varepsilon$	$\widehat{K}$	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\lambda}_3$
1	2	5	119.2551	109.3541	24.9287	4.3378	4.8222	0.2913	0.7087	N/A
			(0.221)	(0.383)	(0.384)	(0.725)	(0.127)	(0.008)	(0.008)	
2	2	5	119.4149	109.7010	24.8160	4.4372	4.8855	0.2944	0.2320	0.4736
			(0.243)	(0.427)	(0.381)	(0.839)	(0.138)	(0.008)	(0.006)	(0.008)

Table 6: Recovering Parameters, Unobserved Heterogeneity. The table displays estimates of model parameters assuming the correct Signal model. Based on generating $T = 5000$ auctions where $\mu_1 = 120$, $\mu_2 = 110$, $\sigma_V = 25$, $\sigma_\varepsilon = 5$, $K = 5$. Case 1: $(c_1, \lambda_1) = (0, 0.3)$, $(c_2, \lambda_2) = (25, 0.7)$. Case 2: $(c_1, \lambda_1) = (0, 0.3)$, $(c_2, \lambda_2) = (25, 0.2)$, $(c_3, \lambda_3) = (50, 0.5)$. Standard deviations in parentheses.

Variable	Mean	Std. Dev.	25^{th}-tile	50^{th}-tile	75^{th}-tile	Min	Max	N
WINNING BID ($/mbf)	92.54	150.43	39.22	69.89	125.01	2.04	4255.73	945
BID ($/mbf)	77.46	92.82	29.5	58.37	106.36	2.04	4255.73	3944
LOGGER	64.98	58.01	23.14	48.37	90.53	2.04	723.96	1072
MILL	82.11	102.46	32.55	62.25	113.55	5.29	4255.73	2872
LOGGER	0.27	0.44	0	0	1	0	1	3944
LOGGER WINS	0.18	0.38	0	0	0	0	1	945
ENTRANTS	3.99	2.47	2	4	5	0	12	988
LOGGERS	1.09	1.32	0	1	2	0	10	988
MILLS	2.91	1.89	1	3	4	0	10	988
POTENTIAL ENTRANTS	9.98	6.45	5	9	14	1	38	988
LOGGER	5.26	4.71	2	4	8	0	27	988
MILL	4.73	2.78	3	4	6	0	14	988
TOTAL OTHER	9.86	6.4	5	9	14	0	38	988
PREVIOUS MILLS (6 mos)	4.57	2.85	2	4	7	0	14	988
SPECIES HERFINDAL	0.55	0.23	0.35	0.51	0.72	0.20	1.00	988
DENSITY (acres/mbf)	0.21	0.21	0.07	0.16	0.28	0.02	1.81	988
VOLUME (hundred mbf)	75.17	45.41	41.05	68.6	103.1	5	275.4	988
HOUSING STARTS	1610.74	267.8	1580.5	1628	1782	843	2260	988
RESERVE ($/mbf)	37.95	32.36	16.38	27.3	47.89	2.04	221.87	988
SELL VALUE ($/mbf)	291.05	64.18	259.51	292.29	326.03	0	518.95	976
ROAD CONST ($/mbf)	12.23	14.33	1.06	7.49	17.76	0	91.55	976
LOG COSTS ($/mbf)	116.53	33.13	98.4	113.12	133.78	0	252.46	976
MFCT COSTS ($/mbf)	134.46	24.09	127.07	136.22	146.15	0	227.55	976
MISSING APPRAISAL	0.01	0.11	0	0	0	0	1	988
FAIL	0.04	0.20	0.00	0.00	0.00	0.00	1.00	988
DISTANCE (KM)	86.91	122.58	37.7	60.2	86.48	0.86	1017.84	2702

Table 7: Summary Statistics for California ascending auctions from 1982-1989. We exclude SBA set asides, salvage sales, auctions with very high or low volume to acreage ratios and failed sales. For calculating when a logger wins, we focus only on auctions where the tract sold. Entrants are those bidders bidding more than the reserve price. We count the number of potential entrants as those bidders who bid within 50km of an auction over the next month. Potential other entrants are the subset of these that excludes bidders who bid in this auction but not in those within 50 km over the next month. Statistics about the potential mill bidders over the previous 182 days that bid in the same forest district is also given by PREVIOUS MILLS.

	(1)	(2)	(3)
log P.O. ENTRANTS	0.292*** (0.025)	0.289*** (0.025)	0.258*** (0.024)
SCALE		0.149* (0.079)	0.137* (0.077)
SPECIES HERFINDAL		0.072 (0.076)	0.094 (0.077)
DENSITY (acres/mbf)		-.233*** (0.081)	-.174** (0.078)
VOLUME (mbf)		-.0008** (0.0004)	-.0006 (0.0004)
HOUSING STARTS		-.00002 (0.0001)	-5.01e-06 (0.0001)
log SELL VALUE ($/mbf)			0.393*** (0.077)
log ROAD CONST ($/mbf)			-.039*** (0.013)
log LOG COSTS ($/mbf)			-.574*** (0.075)
log MFCT COSTS ($/mbf)			0.141* (0.073)
MISSING APPRAISAL			-.617*** (0.238)
YEAR DUMMIES	YES	YES	YES
QUARTER DUMMIES	YES	YES	YES
COUNTY DUMMIES	YES	YES	YES
N	945	945	945
R^2	0.334	0.344	0.404

Table 8: Impact of Potential Entrants on Actual Entry. Dependent variable is log of actual entry. P.O. ENTRANTS (Potential other entrants) are the subset of these that excludes bidders who bid in this auction but not in those within 50 km over the next month. OLS Results are presented.

	OLS				IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log P.O. ENTRANTS	0.157***	0.155***	0.083***	0.092***	0.361***	0.381***	0.221***	0.281***
	(0.024)	(0.024)	(0.022)	(0.026)	(0.077)	(0.078)	(0.076)	(0.09)
LOGGER	-.159***	-.145***	-.121***	-.100***	-.166***	-.154***	-.125***	-.101***
	(0.029)	(0.029)	(0.027)	(0.032)	(0.029)	(0.03)	(0.027)	(0.032)
SCALE		0.246***	0.178***	0.082		0.231***	0.167***	0.064
		(0.061)	(0.056)	(0.067)		(0.061)	(0.056)	(0.067)
SPECIES HERFINDAL		-.663***	-.367***	-.359***		-.675***	-.381***	-.379***
		(0.063)	(0.06)	(0.071)		(0.063)	(0.06)	(0.071)
DENSITY (acres/mbf)		-.052	0.107	0.039		0.0001	0.135**	0.096
		(0.071)	(0.066)	(0.079)		(0.074)	(0.067)	(0.083)
VOLUME (mbf)		-.0007**	-.001***	-.0007*		-.0007**	-.001***	-.0006*
		(0.0003)	(0.0003)	(0.0003)		(0.0003)	(0.0003)	(0.0003)
HOUSING STARTS		0.0002**	0.0002**	0.0003***		0.0001	0.0002*	0.0003***
		(0.00009)	(0.00008)	(0.0001)		(0.00009)	(0.00009)	(0.0001)
log SELL VALUE ($/mbf)			1.418***	1.462***			1.367***	1.393***
			(0.066)	(0.08)			(0.071)	(0.086)
log ROAD CONST ($/mbf)			0.057***	0.04***			0.059***	0.043***
			(0.011)	(0.013)			(0.011)	(0.013)
log LOG COSTS ($/mbf)			-1.536***	-1.617***			-1.488***	-1.554***
			(0.063)	(0.078)			(0.068)	(0.083)
log MFCT COSTS ($/mbf)			-.061	-.111			-.047	-.089
			(0.065)	(0.084)			(0.065)	(0.084)
MISSING APPRAISAL			0.411**	0.034			0.395**	0.013
			(0.186)	(0.215)			(0.186)	(0.215)
log DISTANCE (KM)				-.001				-.005
				(0.019)				(0.019)
YEAR DUMMIES	YES	YES	YES	YES	YES	YES	YES	YES
QUARTER DUMMIES	YES	YES	YES	YES	YES	YES	YES	YES
COUNTY DUMMIES	YES	YES	YES	YES	YES	YES	YES	YES
FIRST STAGE IV COEFFICIENT					0.079***	0.077***	0.075***	0.075***
					(0.004)	(0.004)	(0.004)	(0.005)
N	3944	3944	3944	2702	3944	3944	3944	2702
R^2	0.227	0.256	0.379	0.382				

Table 9: Impact of Potential Entrants on All Bids. Dependent variable is log of bid per volume. P.O. ENTRANTS (Potential other entrants) are the subset of these that excludes bidders who bid in this auction but not in those within 50 km over the next month. Instrument for potential entrants is the number of unique mill bidders in a forest during the previous 6 months. Observations drop when we include distance because we are missing that data for some bidders.

	OLS				IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log P.O. ENTRANTS	0.279***	0.273***	0.19***	0.174***	1.119***	1.130***	0.924***	1.086***
	(0.037)	(0.037)	(0.032)	(0.039)	(0.144)	(0.147)	(0.136)	(0.203)
LOGGER	-.222***	-.192***	-.074	-.026	-.280***	-.285***	-.155**	-.111
	(0.07)	(0.071)	(0.062)	(0.076)	(0.086)	(0.089)	(0.077)	(0.103)
SCALE		0.416***	0.32***	0.169		0.327**	0.246*	0.071
		(0.119)	(0.102)	(0.126)		(0.148)	(0.126)	(0.168)
SPECIES HERFINDAL		-.618***	-.332***	-.368***		-.598***	-.360***	-.340**
		(0.113)	(0.102)	(0.124)		(0.14)	(0.125)	(0.165)
DENSITY (acres/mbf)		-.066	0.065	-.037		0.117	0.199	0.294
		(0.12)	(0.103)	(0.127)		(0.152)	(0.129)	(0.182)
VOLUME (mbf)		-.013**	-.015***	-.011*		-.019***	-.020***	-.012
		(0.005)	(0.005)	(0.006)		(0.007)	(0.006)	(0.007)
HOUSING STARTS		0.0002	0.0002	0.0002		0.00007	0.00003	0.00009
		(0.0002)	(0.0001)	(0.0002)		(0.0002)	(0.0002)	(0.0002)
log SELL VALUE ($/mbf)			1.276***	1.135***			0.952***	0.843***
			(0.102)	(0.119)			(0.139)	(0.17)
log ROAD CONST ($/mbf)			0.047***	0.03			0.065***	0.05*
			(0.018)	(0.022)			(0.022)	(0.029)
log LOG COSTS ($/mbf)			-1.766***	-1.552***			-1.437***	-1.242***
			(0.1)	(0.122)			(0.136)	(0.174)
log MFCT COSTS ($/mbf)			0.315***	0.229*			0.376***	0.282*
			(0.096)	(0.125)			(0.119)	(0.165)
MISSING APPRAISAL			0.131	-.113			-.052	-.187
			(0.317)	(0.399)			(0.391)	(0.527)
log DISTANCE (KM)				0.076**				0.027
				(0.032)				(0.044)
YEAR DUMMIES	YES	YES	YES	YES	YES	YES	YES	YES
QUARTER DUMMIES	YES	YES	YES	YES	YES	YES	YES	YES
COUNTY DUMMIES	YES	YES	YES	YES	YES	YES	YES	YES
FIRST STAGE					0.094***	0.092***	0.087***	0.075***
IV COEFFICIENT					(0.009)	(0.009)	(0.009)	(0.012)
N	945	945	945	671	945	945	945	671
R^2	0.36	0.396	0.563	0.551				

Table 10: Impact of Potential Entrants on Winning Bids. Dependent variable is log of the winning bid per volume. P.O. ENTRANTS (Potential other entrants) are the subset of these that excludes bidders who bid in this auction but not in those within 50 km over the next month. Instrument for potential entrants is the number of unique mill bidders in a forest during the previous 6 months. Observations drop when we include distance because we are missing that data for some bidders.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CONSTANT	2.941***	2.690***	4.030***	3.071***	3.335***	3.333***	3.043***	3.037***
	(0.421)	(0.421)	(0.785)	(0.768)	(0.424)	(0.423)	(0.72)	(0.713)
LOGGER	-.154***	-.388***	-.202***	-.638***	-.118***	-.211***	-.054	-.285***
	(0.029)	(0.047)	(0.072)	(0.096)	(0.027)	(0.041)	(0.063)	(0.081)
SCALE					0.184***	0.187***	0.34***	0.339***
					(0.056)	(0.056)	(0.104)	(0.103)
SPECIES HERFINDAL					-.359***	-.351***	-.324***	-.292***
					(0.06)	(0.06)	(0.104)	(0.103)
DENSITY (acres/mbf)					0.09	0.081	0.031	0.0008
					(0.066)	(0.066)	(0.105)	(0.104)
VOLUME (mbf)					-.001***	-.001***	-.001***	-.002***
					(0.0003)	(0.0003)	(0.0005)	(0.0005)
HOUSING STARTS					0.0002**	0.0002**	0.0002	0.0001
					(0.00008)	(0.00008)	(0.0001)	(0.0001)
log SELL VALUE ($/mbf)					1.448***	1.466***	1.360***	1.397***
					(0.066)	(0.066)	(0.104)	(0.103)
log ROAD CONST ($/mbf)					0.055***	0.051***	0.042**	0.035*
					(0.011)	(0.011)	(0.018)	(0.018)
log LOG COSTS ($/mbf)					-1.565***	-1.608***	-1.851***	-1.946***
					(0.063)	(0.065)	(0.101)	(0.102)
log MFCT COSTS ($/mbf)					-.070	-.049	0.299***	0.349***
					(0.065)	(0.065)	(0.098)	(0.098)
MISSING APPRAISAL					0.421**	0.437**	0.178	0.221
					(0.186)	(0.186)	(0.323)	(0.32)
$\widehat{\lambda}$		0.342***		0.658***		0.136***		0.341***
		(0.054)		(0.098)		(0.046)		(0.077)
YEAR DUMMIES	YES	YES	YES	YES	YES	YES	YES	YES
QUARTER DUMMIES	YES	YES	YES	YES	YES	YES	YES	YES
COUNTY DUMMIES	YES	YES	YES	YES	YES	YES	YES	YES
ONLY WIN BID	NO	NO	YES	YES	NO	NO	YES	YES
N	3944	3944	945	945	3944	3944	945	945
R^2	0.219	0.227	0.32	0.352	0.377	0.378	0.546	0.555

Table 11: Heckman Selection Evidence. Dependent variable is log of the bid per volume. The table displays the second stage results of the two step selection model. The first stage probit is of entry where the exogenous shifters are potential other mill and logger entrants, incorporated as a flexible polynomial. $\widehat{\lambda}$ is the estimated inverse Mills ratio from the first stage.

	(1)	(2)
μ_{Mill}	3.719	3.666
	(0.0129)	(0.0154)
μ_{Logger}	2.778	3.060
	(0.0172)	(0.0387)
$\sigma^2_{V,Mill}$	0.8916	0.997
	(0.0113)	(0.0187)
$\sigma^2_{V,Logger}$	same as mill	0.828
		(0.0227)
K_{Mill}	0.0042	0.0003
	(0.0003)	(0.0000)
K_{Logger}	same as mill	0.1827
		(0.0350)
$\sigma^2_{\varepsilon,Mill}$	0.112	0.043
	(0.0030)	(0.0026)
$\sigma^2_{\varepsilon,Logger}$	same as mill	0.183
		(0.0130)
N	873 auctions	873 auctions

Table 12: Parameter Estimates. Estimates of parameters for lognormal parameterization. The two specifications differ on their restrictions regarding signal variance. The latter model allows it to differ across types due to differences in value dispersion and noise dispersion.