A Structural Study of a Dynamic Auction Market

Anna Ingster *

October 22, 2009

JOB MARKET PAPER

Abstract

A dynamic model is developed that incorporates two key features of online auctions for standardized goods. First, there is an infinite series of auctions for an identical item. Second, those auctions are naturally ordered over time. I model the buyers' decision of what auctions to bid in and how much to bid. The model is applied to a rich dataset of online auctions for a particular Apple iPod model. I propose an identification and estimation strategy that allows me to evaluate the efficiency of the current mechanism versus a perfect competition benchmark. The estimation reveals significant inefficiency present in the existing mechanism. An alternative second best mechanism is proposed that would allow to reduce market inefficiency.

1 Introduction

Auctions are known to perform well in the sale of collectible, one-off items. However, a high volume of standardized goods are also sold through auctions. For example, in a three month period (Spring 2006)

*Economics Department, New York University, 19 W 4-th str., 6-th Floor, New York, NY 10012, e-mail: ai360@nyu.edu.
I'm thankful to Luis Cabral and John Asker for valuable advice on this paper. I'd also like to thank Victor Archavski, Heski Bar-Isaac, Allan Collard-Wexler, Ariel Pakes, Jorg Stoye, Ali Yurukoglu, Yi Daniel Xu and seminar participants in NYU FAS Economics and NYU Stern for helpful discussion. The remaining errors are all mine.

more than 8,000 Apple 60 GB iPods were auctioned at a major online site. How well do auctions perform in the sale of mass-manufactured items such as this? In terms of number of agents (about 2500 sellers and 30,000 bidders during the same three-month period), the market seems rather competitive. Does the online auction mechanism deliver a competitive outcome? If not, is there an alternative mechanism that would improve efficiency?

In this paper, I first propose a dynamic model of auctions for identical (or close substitute) items which are naturally ordered by auction end time. I then estimate the model, thus obtaining the distribution of buyer valuations. Finally, based on the model and estimation results I evaluate the efficiency of auctions versus a perfectly competitive benchmark.

In my basic setting, the market is modeled as a sequence of standard second price auctions. The auctions are connected through an endogenously evolving set of bidders. Each bidder has unit demand. Bidders enter the entire market, rather than each particular auction. The losers in earlier auctions stay in the market and continue bidding in future auctions. Since low value bidders tend to stay longer in the market, the model implies that the distribution of bidders' valuations differs from the parent distribution of valuations. Each bidder's valuation is determined by the characteristics of a particular item (items are close substitutes) as well as by the bidder's type (the overall valuation of the items, which I assume is fixed for a given bidder). In contrast to a single auction model, I show that it is not optimal for a bidder to bid her valuation. Rather, a bidder should shade her bid by the continuation value of her dynamic optimization problem, which is a function of her type. I provide an explicit characterization of the optimal bid as a function of the bidder's valuation in this environment.

I then extend my baseline model by allowing each auction to be open for a given time interval. In this setting, each bidder must choose in which of the currently open auctions she wants to bid and then how much to bid. This formulation is a better approximation for the actual structure of an online auction market. The model's complete solution is very complicated and highly sensitive to specific informational assumptions. To address these problems, I introduce a reduced-form assumption regarding the bidders' choice of auction in which to bid. I assume that each active bidder considers the open auctions in the order of time remaining until expiration and submits a bid for the first item for which her optimal bid exceeds outstanding item price. The optimal size of the bid as a function of the bidder's valuation is characterized in this environment.

I propose an identification and estimation strategy, that allows me to recover the distribution of the bidder's valuations in the population from the data. First, I parameterize the processes of the item arrival, and of the entry, exit, and participation of the bidders, and estimate the parameters from the data. Given the estimated parameters, I show that there is a one-to-one relationship between the distribution of an order statistic of the bids (I use first or second highest bid) submitted in the auctions and the distribution of the optimal bids in the population. This finding allows me to identify the optimal bid distribution in the population. My theoretical result characterizing the optimal bid as a function of valuation allows me to reconstruct the distribution of the bidders' valuations from the distribution of the optimal bids in the population. My result is parallel to the standard identification result (e.g., Athley and Haile, 2008), namely that the parent distribution can be identified from an order statistic of a sample from the distribution of a given size. However, my result holds in a more complicated dynamic environment. The estimated model parameters play the same role for identification as the known size of the sample in the classic result.

I apply the proposed identification and estimation strategy to a dataset obtained from a major online auction marketplace. The dataset includes all auctions of 60 GB video iPod that were completed in Spring 2006.

One distinctive feature of the dataset I use in my estimation is the availability of the data for the highest bids in each auction. (This data is typically missing in most datasets obtained by direct monitoring the auction's website.) While my model provides identification using only the price data (the data typically used in other studies), the availability of the highest bid data provides a source of external validation.

I then simulate the performance of a competitive market characterized by the demand and supply curves estimated from the data, deriving as a result competitive price and quantity estimates. Since for some of the items only interval estimates of the reserve price are available, I can only provide interval estimates for the supply curve and thus for the competitive outcome. The results suggest that both price and quantity provided by the auction market are below the competitive market values. The difference in price, obtained by alternative estimates, varies between 0.7% and 1.4%. The difference in quantity varies between -0.1% and 7.6%.The total consumer surplus provided by the auction market is estimated to be 13.3% to 40.8% below the potential consumer surplus under perfect competition. Under all assumptions, the distortions in price and consumer surplus are statistically significant, whereas the distortion in quantity may or may not be significant. The average consumer surplus from purchasing an item in the auction market is estimated to be $85.04, or 25% of the average price paid. The consumer surplus per item under perfect competition is estimated to be between $96.36 and $151.99.I also show that, if the dynamic incentives faced by bidders are not taken into account, then estimated values are significantly different. In particular, the estimated consumer surplus per item under this assumption is only $13.93, or 4% of the item price.

I then evaluate the performance of a competitive market characterized by the demand and supply curves estimated from the data, deriving as a result competitive price and quantity estimates. I find significant inefficiency in the operation of the market. In particular, for a given set of assumptions the Consumer Surplus is about 16% below competitive benchmark. I then propose a modification of the auction mechanism that allows bidders to submit bids simultaneously for multiple items without bearing a risk of buying multiple items. My simulations show that the market inefficiency can be reduced significantly by this mechanism. In particular, the loss in Consumer Surplus is reduced to about 11%.

The rest of the paper is organized as follows. In Section 2 I review the related literature. Then, in Section 3, I propose the Baseline Model. Section 4 states the identification results. Section 5 describes the data. Section 6 proposes the estimation strategy and Section 7 provides the estimation results for the Baseline Model. Section 8 describes the Extended version of the model. Section 9 provides the estimation strategy for the Extended Model and Section 10 reports the results. In Section 11 I propose a second best mechanism for trading in a dynamic environment and demonstrate a superiority of the proposed mechanism over the one used in the sample. Section 12 concludes. The description of the market and some of the proofs are placed in the Appendix.

2 Literature Review

In recent years, online auctions have become a topic of a significant and diverse academic literature. The authors have addressed different questions and used various methodologies. Some of this literature is summarized in Bajari and Hortacsu (2004). The authors address such diverse and interesting questions as empirical regularities is bids timing, including "sniping" (Roth and Ockenfels (2000), Ockenfels and Roth (2006), Bajari and Hortacsu (2003), Ely and Hossain (2006), Nekipelov (2007), etc.); the asymmetric information issues (Bajari and Hortacsu (2003), Jin and Kato (2002), Lewis (2007), etc.); the work of reputation mechanisms (Cabral and Hortatcu (2006), Melnik and Alm (2002), Resnick et al. (2006), etc.). The methodology used by different authors also varies. Some use hedonic regressions (Jin and Kato (2002), Melnik and Alm (2002), etc.); some build and estimate structural econometric models: parametric (Bajari and Hortacsu (2003), Lewis (2007), Nekipelov (2007)), or nonparametric (Song (2004), Sailer (2006)). A number of authors performed field experiments using online markets (Lucking-Reiley (1999), Nekipelov (2007)), or check the conjectures about online auctions in the lab (Arieli, Ockenfels and Roth (2005)).

Recently a wide empirical literature emerged that deals with structural estimation of auctions. Much of this literature is reviewed in Paarsh and Hong (2006), Athey and Haile (2008), Hendricks and Porter (2007). Most papers treat the data as a cross-section of auctions that are considered unrelated to each other. In this setting the techniques have been developed to estimate the bidders' valuation distributions in both first-price (e.g. Guerre, Perrigne and Vuong (2000)) and second-price (e.g. Athey and Haile (2001)) environments. Notice, that the second-price case is in general much easier, since the optimal bids in the standard model coincide with the bidders' valuations of the items.

A few papers incorporate the dynamic considerations by the bidders. For example, Jofre-Bonet and Pesendorfer (2003) consider the capacity constraints of the bidders in highway procurement auctions (first price) that make the bidding history-dependent.

On the theoretical front an important contribution is the paper by Peters and Severinov (2006) that analyzes the bidders' optimal behavior in an environment of multiple simultaneous auctions for identical items. Essentially, the market converges to the competitive price in their model. However, as I argue in this paper, the bidders' inter-temporal incentives in an online market are different from Peters and Severinov's model, since the auctions are ordered over time, and the losers of earlier auctions can participate in the later auctions. A recent paper by Jofre-Bonet and Pesendorfer (2008) in a simple two-period environment analyzes the incentives of the bidders in a sequential auctions environment when the auctioned items are either substitutes or compliments and compares different auction designs. My theoretical findings, presented in this paper, are similar to the results obtained by Jofre-Bonet and Pesendorfer for the case of substitutability between items.

Applying the standard estimation techniques to the online auction data introduces certain challenges that have been first addressed by Song (2004).

The standard identification result, widely used in empirical auction literature (see Athey and Haile (2008)), is based on one-to-one relationship between a parent distribution and any order statistic of a sample from this distribution of a given size. This result allows to identify the distribution of the optimal bids in the population of bidders by only observing the first, or the second highest bid in each auction, as soon as the number of the bidders in the auction is known.

The main problem, addressed by Song, is that in an online auction application the number of potential bidders for each auction is not only uncertain, but also unobserved by the researcher. It happens due to several reasons. First, there is no preliminary registration to participate in a certain auction. Any bidder registered on the web-site is eligible for bidding at any time. What is available to a researcher, is typically the bid data (not page-click data), so if the bidder did not submit a bid, she leaves no record in the dataset. On the other hand, a bid may only be accepted, if it exceeds the outstanding price of the item, which is computed as the second highest bid submitted to the moment. Therefore, a bidder who has an intention to submit a bid may not be able to do it, if her preferred bid happens to be below the outstanding price of the item at the time of bidding. This makes many potential bidders unobservable to the researcher. However, as it is shown by Song, for each auction we always observe the first and second highest potential *bidders* in the data. Second highest *bid* is also observable, however for the *first highest bid* we typically only observe the time and identity of the bidder, but not the bid size. It happens due to the "proxy bidding" system implemented by major auction sites (see Appendix A). With unknown number of bidders, one cannot identify the valuation distribution from the second-highest bid data alone. Song proposes a set of conditions under which the *third-highest* bid data also happens to be reliable. Based on the second and third highest bid data she constructs a statistic that does not depend on the number of bidders, and based on this statistic, estimates the bidders' valuation distribution. Notice, however, that while this approach allows us to identify the valuation distribution, the model as a whole is not identified. In particular, the distribution of the number of bidders in each auction remains unknown. Thus one cannot, for example, identify the distribution of highest bids in the auctions (it would depend on both known valuation distribution and unknown number of bidders). Therefore, the results obtained are not sufficient to perform a comprehensive comparative statics analysis. This reasoning illustrates, how far could we get without making additional assumptions. In order to get full identification, one should introduce some additional structure to the model. For example, one could make a parametric assumption on the distribution of the number of bidders and estimate the parameters.

One more important point to be made about Song's paper is that, as the most of the empirical auction literature, it deals with a cross-section of unrelated auctions. This assumption restricts the applicability of the model to a rather narrow subset of goods auctioned online, since many commonly sold goods have close substitutes offered by nearby auctions.

Sailer (2006) extends Song's approach to a dynamic settings. This paper models the auction market as a sequence of second price sealed bid auctions. The main focus of the paper is to estimate the implicit costs of bidding, that measure alternative cost of time and the quality of internet connection, and are bidder-specific. Although the environment in the model is dynamic and bidders have unit demand, the valuations of different items by the same bidder are assumed to be independent from each other. As a result, the continuation value of the problem does not depend on the valuation of the current item by the bidder. The model, I develop in the current paper, is similar, however I relax the independent valuations assumption, allowing the valuations of different items by the same bidder to be correlated.

I do not adopt Song's approach to the model estimation. With the data on both first and second highest bids in each auction I could achieve even more reliable estimates following Song's approach. However the problem is that I would only be able to estimate the distribution of the bidders' valuation in the *pool* of bidders currently in the market, as opposed to the *population* distribution. In a static model the two distributions would coincide. However in a dynamic setting there is an asymmetry among the bidders that causes the two distributions to differ. Specifically, the bidders with higher types tend to exit the pool faster by winning items and, therefore, comparing to the population distribution there is a bias towards the lower types in the pool. If one is interested in running counterfactual experiments using the model estimates, it is the parent population distribution, not the derived pool distribution that must be evaluated. I provide an alternative estimation strategy that allows me to evaluate the distribution of the bidders' valuations in the population. Figure 1 illustrates the difference between the pool and population optimal bid distributions estimated in my model.

In what follows I will present several versions of my model in the order of increasing complexity. I distinguish between the *Baseline* and *Extended* versions of the model. The Baseline model treats the auction market as a sequence of standard second price sealed bid auctions, thus the auctions in this model are subsequent to each other. In the Extended Model I allow each auction to be open for a prolonged period of time, thus the auctions are overlapping. The Baseline Model is easier to analyze and indeed allows for a closed form solution. The Extended Model is much more complicated, and in order to address this model I have to make simplifying assumption according bidders behavior. Formal results are stated for the baseline model in the text, for expositional ease. Full results for the extended model are found in the appendix except where they differ substantially, in which case they are in text.

3 Baseline Model

For the ease of explanation I will first present a version of my model that assumes all the auctioned items to be identical. Then I will generalize my results to the case of heterogeneous items.

3.1 Homogeneous Items

In this section I introduce a simple version of the Baseline model that will be further generalized.

The time is continuous. Items randomly arrive to the market following a Poisson process with the parameter λ_{item}. Each item is auctioned via a standard second price sealed bid (SPSB) auction with no reserve. (Zero reserve assumption will further be dropped). All the items are identical. I use a simplifying SPSB assumption in this model, thus unlike in the actual auction bidders may only bid once for a given item and cannot react on each other's bids. It would be a good description of a market where all bidders "snipe" (submit bids in the last seconds before item expiration).

Each bidder only has positive demand for one unit of the good auctioned. The bidders are entering the market following a Poisson process with the parameter λ_{bidder}. Each bidder i has a type θ_i that in this simple setting determines her valuation for each item. For the new entrants types are drawn randomly from a population distribution $\mathcal{F}(\cdot)$. Bidders, who have entered the market, form a pool of potential bidders. The winner of each completed auction exits the pool. Bidders also exit the pool randomly. For a bidder currently in the pool at time t the probability to still be in the pool at time $t + \tau$, conditional on non-winning, is determined as δ^τ. With probability $1 - \delta^\tau$ a bidder exits the pool by that time.

If the bidder i wins the item j at price p_j, the bidder's utility from this item is given by

$$U_{ij} = \theta_i - p_j. \tag{1}$$

If a bidder exits the pool without winning the payoff of the bidder is equal to zero. In reality, the exit of the bidder from the market is a complicates event, that may include change of preferences, obtaining the items from a retailer, choosing a different model of an iPod, or buying an item through "Buy It Now". Normalizing the outside option to zero is a simplifying assumption I make in the model.

The bidders in the pool may get active in certain moments of time. Parameter α in the model denotes a probability of a bidder to be active per unit of time. For a time span of the length τ the cumulative probability to get active is $1 - (1 - \alpha)^\tau$. I assume α to be independent of the bidder's type and the time spent by the bidder in the market.

All bidders who happen to be active in the time interval between the auctions $j-1$ and j participate in the auction j. I assume simultaneous bidding for each auction. It means that the bidders do not react to each other's bids even if technically the bidding is sequential. The highest bidder wins the item. The price of the item is determined by second highest bid submitted in the auction[1].

The model does not allow bidders to submit any significant bids early in the auction (before the auction becomes the first one in the list). This assumption makes the model analysis much simpler. Empirically it is justified to some extent. Much of the literature emphasizes the importance of the "sniping" (very late bidding) strategy for online auction markets. In my data set 64% of all winning bids submitted in the last 10 minutes of the auction, 80% within the last hour. The Baseline model concentrates on late bidding in each auction, essentially by ruling out a possibility of early bidding. This assumption will be relaxed in the Extended model (see Section 8).

The reduced form random participation assumption may be a good description for a process like buying an iPod, that is seldom performed on a full-time basis and normally just a minor event in a course of everyday life. A lot of random shocks may divert bidder's attention from the market. I also abstract from the daily and weekly cycles for simplicity.

Although in actual markets bidders may observe the characteristics of the forthcoming auctions, including timing, I assume here that at the moment of choosing the bid for a given auction the bidders *do not observe the timing of the following auctions*. The waiting time until the following auction is evaluated by the bidders as the expected Poisson waiting time.

Consider the problem faced by an individual bidder in this model. The following result generalizes the standard proposition for the optimal bid in a second price auction:

Lemma 1. *Let $V(\theta)$ be the expected payoff for a bidder of type θ from participating in the market. Then the optimal bid for this bidder is determined by*

$$B(\theta) = \theta - \tilde{\delta} V(\theta). \tag{2}$$

[1] I ignore the minimal increment in theoretical model

where $\tilde{\delta}$ *is the expected survival rate until the next item:*

$$\tilde{\delta} = E(\delta^t) = \frac{\lambda_{item}}{\lambda_{item} - \log(\delta)}. \tag{3}$$

To prove this result one should use the exact same reasoning as in the proof of the standard static result (see Krishna (2002)) with the only exception that the payoff of the bidder in case of losing the auction, that is usually normalized to zero, is now given by her outside option to participate in the continuation of the game beginning from the next item[2]. This result is driven by the unit demand assumption. By winning an item today a bidder abandons the opportunity to win another item in the future (possibly at a lower price). This introduces a nontrivial outside option which is a function of the bidder's type. Therefore, in contrast with a single auction environment, the bidding function now becomes a nonlinear function of the bidder's type.

Notice, that the evolution of the auction market described above constitutes a Markov process with the state variable consisting of the number of bidders in the pool and the types of the bidders. Although it might be an interesting question to study the evolution of this process from a given state, I assume that the process is stationary. Under certain regularity conditions (that I assume to be fulfilled) the state distribution of any Markov process converges to the stationary distribution over time. Empirically, since the market operates in unchanged environment for a prolonged period of time, I assume this convergence to have occurred. Therefore, the distributions characterizing the process realizations are time-independent.

Denote by $G^-(\cdot)$ the stationary distribution of the highest bid submitted in an auction by bidders other then i. Then the probability that i wins the auction is $G^-(B(\theta_i))$, and the probability that i loses is $1 - G^-(B(\theta_i))$.

If bidder i indeed wins the auction, her expected payoff is given by

$$W(\theta_i) = \int_0^{B(\theta_i)} (\theta_i - p) dG^-(p). \tag{4}$$

Finally the Bellman equation for the bidder's problem is:

[2] I am ruling out a possibility of inter-temporal coordination between the bidders since the repeated interaction is very limited

$$V(\theta_i) = (1-\alpha)\tilde{\delta}V(\theta_i) + \alpha\left(W(\theta_i) + \tilde{\delta}\left(1 - G^-\left(B(\theta_i)\right)\right)V(\theta_i)\right). \tag{5}$$

Proposition 1. *In the model of an auction market with no reserve the inverse optimal bid function of the bidders is given by*

$$\theta = B^{-1}(b) = b + \frac{\alpha\tilde{\delta}}{1-\tilde{\delta}}\int_0^b (b-p)dG^-(p). \tag{6}$$

The optimal bid function and it's inverse are both one-to-one and increasing.

The equation (6) can be obtained by solving the system (2, 4, 5). Since the integral in (6) is clearly positive and increasing, $B^{-1}(\cdot)$ is one-to-one and increasing and so too is $B(\cdot)$.

The distribution $G^-(\cdot)$ of the highest rival bid remains to be determined. I do not attempt to solve for this distribution analytically, since it can be estimated from the available data.

The model above can be easily generalized for the case of nonzero reserve price. Suppose, that the reserve price in each auction is an independent draw from a distribution $R(\cdot)$. The individual bidding function given by (2) remains unchanged (although only the bids that exceed the reserve price will be accepted). In the equations (4 -5) one should replace the distribution $G^-(\cdot)$ of the highest rival bid by the distribution

$$\check{G}^-(\cdot) = G^-(\cdot) \times R(\cdot) \tag{7}$$

of the maximum between the highest rival bid and the reserve price. Thus the result of Proposition 1 can be generalized:

Proposition 2. *In the model of an auction market with the reserve price randomly drawn from the distribution $R(\cdot)$ the inverse optimal bid function of the bidders is given by*

$$\theta = B^{-1}(b) = b + \frac{\alpha\tilde{\delta}}{1-\tilde{\delta}}\int_0^b (b-p)d(G^-(p) \times R(p)). \tag{8}$$

The optimal bid function and it's inverse are both one-to-one and increasing.

3.2 Item Sellers

In this paper I adopt a reduced form model of sellers' behavior. I assume that each seller who offers an auction for an item j has a reservation value for this item ρ_j that is a random draw from a distribution $\mathcal{R}$. The reservation value determines the minimal price at which the seller would prefer to sell the item rather then keep it. The reservation values for different items are supposed to be independent (even if offered by the same seller).

Let ϵ be a small number representing sellers' sensitivity level. The exact value of this parameter is not important.

Assumption 1. *Let ρ_j be the reservation value of the item j to the seller. Suppose that for an auction with zero reserve the probability for the price of the item to be below reservation value is high enough: $Prob(p_j < \rho_j) > \epsilon$. Then the seller posts a reserve price for the item that equals the sellers reservation value:*

$$r_j = \rho_j \tag{9}$$

The Assumption 1 states, that the sellers for whom the reservation value is binding, meaning there is a significant probability to obtain a price of the item below the reservation value, are submitting the reserve price equal the reservation value, $r_j = \rho_j$ for the items. That is, unlike I do with the bidders, I assume that the sellers post the reserve prices "truthfully". This is a simplifying assumption, that I am making at this stage to concentrate on the demand side of the market. Solving for optimal seller's behavior would be an interesting extension of the current paper. In the empirical part of the paper I discuss the sensitivity of my results to this simplifying assumption.

3.3 Competitive Benchmark

Consider an auction market characterized by the valuation distribution $\mathcal{F}$ of the bidders, sellers' reservation value distribution $\mathcal{R}$, the bidder arrival rate λ_{bidder}, and the item arrival rate λ_{item}.

Definition 1. *The Competitive Benchmark of the auction market $(\mathcal{F}, \mathcal{R}, \lambda_{bidder}, \lambda_{item})$ is a perfectly com-*

petitive market, characterized by the demand function

$$D(p) = (1 - \mathcal{F}(p))\lambda_{bidder} \tag{10}$$

and the supply function

$$S(p) = \mathcal{R}(p)\lambda_{item}; \tag{11}$$

the competitive price is determined as the solution p^* *of the equation*

$$D(p^*) = S(p^*). \tag{12}$$

Assumption 1 made above allows me to use the observed reserve price distribution R to identify the competitive benchmark outcome.

Define

$$S'(p) = R(p)\lambda_{item}. \tag{13}$$

Assumption 1 allows me to use the observed reserve price distribution R to identify the supply curve for high enough levels of price:

$$S(p) = S'(p), p \geq \underline{p}. \tag{14}$$

Assuming the reservation values close to competitive price are binding, that is $\underline{p} < p^*$, I can evaluate the competitive price and quantity from the equation:

$$D(p^*) = S'(p^*). \tag{15}$$

The competitive benchmark represents the first best efficient allocation that might be achieved in the market (this allocation maximizes the total surplus of all market participants). The question I'll be addressing in my estimation is how close does the outcome of the actual market approach to the first best outcome.

3.4 Heterogeneous Items

One might be interested in generalizing the model of Section 3.1 by allowing some level of item heterogeneity. For example, one question of interest is the impact of the seller's feedback characteristics on the auction outcome. There are also other characteristics of the items that may vary, such as item condition, color, various "extras" offered with the item, or just the style of the description.

Indeed, some of the item characteristics (e.g. color) presumably are horizontal. Furthermore, seller's reputation may play a role for participation decisions (making participation endogenous) rather than just altering the valuation. However, as a first step, I put those complications aside and simply assume that the items are differentiated through a vector of vertical characteristics, and, moreover, all bidders are equally sensitive to those characteristics. Further, to make the model more tractable, I assume additive separability of the bidder's utility function with respect to item characteristics. (Multiplicative separability would be much more difficult to work with). Then the valuation of the item j by the bidder i is given by

$$V_{ij} = \theta_i + \beta X_j + \varepsilon_j, \tag{16}$$

and the bidder's utility from purchasing an item is determined as

$$U_{ij} = \theta_i + \beta X_j + \varepsilon_j - p_j, \tag{17}$$

where X_j is the vector of the item's observable characteristics, ε_j stands for those item characteristics that are observable by all bidders, but not observable by the econometrician, p_j is the purchase price. The type θ_i characterizes the bidder and is assumed to be randomly drawn from the population distribution $\mathcal{F}$.

Unlike I did with the items' timing, I do not have to make any restricting informational assumptions about what bidders know about future items. I can assume more realistically that the information about the characteristics $(X_{j+k}, \varepsilon_{j+k})$ of several forthcoming items is available to the bidders at time of bidding for an item j. It turns out, that under the additive separability assumption, the auctions' characteristics become neutral and do not affect bidders' utility. The intuition is simple: since both the winner's valuation and the price are affected by item characteristics in the same way, the two effects just cancel out, so the winner's utility is not affected by the type of an item purchased. Of course, this reasoning illustrates the restrictive nature of the assumption. However the assumption introduces a useful simplification in the analysis, and still allows me to control for item characteristics in some way.

Exactly the same reasoning as in the homogeneous item case allows me to establish the following Lemma, that generalizes Lemma 1:

Lemma 2. *Let $V(\theta, Z)$ be the expected payoff for a bidder of type θ from participating in the market, computed just before submitting a bid for an item with the characteristics (X, ε); here Z stands for all the information about future items available to the bidder. Then the optimal bid for this bidder is determined by*

$$B(\theta) = \theta + \beta X + \varepsilon - \tilde{\delta} V(\theta, Z). \tag{18}$$

where $\tilde{\delta}$ is the expected survival rate, computed by (3).

Notice, that all the optimal bids by all bidders are affected by the current item's characteristics through the same additive term $\beta X + \varepsilon$. Therefore it is natural to assume, that the distribution G^- of the highest "rival" bid for the item is affected by the same term. I make this assumption explicitly, since I do not derive the distribution G^- in my model. I also have to assume, that the restriction of price being a positive number is not binding for the relevant range of item characteristics.

Assumption 2. *The distribution of the highest "rival" bid for an item depends on the item characteristics in the following way:*

$$G^-(p|X, \varepsilon) = \tilde{G}^-(p - \beta X - \varepsilon) \tag{19}$$

Assumption 3. *For all item characteristics X, ε*

$$supp(\tilde{G}^-(p - \beta X - \varepsilon)) \in [0, \infty) \tag{20}$$

Lemma 3. *Under Assumptions (2 - 3) the expected payoff for a bidder of type θ from participating in the market does not depend on the current or future items' characteristics:*

$$V(\theta, Z) = V(\theta). \tag{21}$$

The optimal bid for a bidder of the type θ is given by

$$B(\theta) = \tilde{\theta} + \beta X + \varepsilon, \tag{22}$$

where

$$\tilde{\theta} = \theta - \tilde{\delta} V(\theta) \tag{23}$$

Proof. See the Appendix B. □

Notice that according to Lemma 3 the optimal bid of each bidder for an item with given characteristics (X, ε) is additively separable with respect to those characteristics. The bid is determined by the item characteristics (X, ε) and $\tilde{\theta}$, a variable determined by the equation (23). As it will be shown in the following Proposition, this variable is a one-to-one function of the bidder's type θ. I will refer to $\tilde{\theta}$ as the bidder's *transformed type.* It is the transformed type of the bidders that determine the optimal bid after controlling for the item characteristics:

$$\tilde{\theta} = B - \beta X - \varepsilon. \tag{24}$$

The result obtained in Section 3.1 is easily generalized to this setting.

Proposition 3. *The inverse optimal bid function of the bidders is given by*

$$\theta = B^{-1}(\tilde{\theta}) = \tilde{\theta} + \frac{\alpha\tilde{\delta}}{1-\tilde{\delta}} \int_{0}^{\tilde{\theta}} (\tilde{\theta} - p) d\tilde{G}^{-}(p), \tag{25}$$

where $\tilde{\theta} = B - \tilde{\beta} X - \varepsilon$ *is the bidder's transformed type. The optimal bid function and it's inverse are both one-to-one and increasing.*

Proof. See Appendix B. □

As before, the result can be easily generalized for the case of non-zero reserve price. Suppose, the reserve price for each item is randomly drawn from the distribution $R(\cdot|X, \varepsilon)$. To make the reasoning straightforward I have to make the following assumption.

Assumption 4. *The distribution of the reserve price for an item depends on the item characteristics in the following way:*

$$R(p|X, \varepsilon) = \tilde{R}(p - \beta X - \varepsilon) \tag{26}$$

I also make a similar assumption concerning the sellers' underlying reservation values:

Assumption 5. *For an item with the vector of characteristics* (X, ε)*, the seller's reservation value is*

$$\tilde{\rho} = \rho + \beta X + \varepsilon; \tag{27}$$

where $\rho \sim \mathcal{R}(\cdot)$.

The Assumptions 4-5 imply that both sellers and bidders are equally sensible to the item's characteristics. Although not being very realistic, those assumptions are needed to maintain the characteristics' neutrality property in the analysis.

Under Assumption 4 one can easily generalize the Proposition 3.

Proposition 4. *Suppose, the reserve price in each auction is randomly drawn from a distribution* $R(\cdot|\cdot)$ *satisfying the Assumption 4. Then the inverse optimal bid function of the bidders is given by*

$$\theta = B^{-1}(\tilde{\theta}) = \tilde{\theta} + \frac{\alpha\tilde{\delta}}{1-\tilde{\delta}} \int_0^{\tilde{\theta}} (\tilde{\theta} - p) d(\tilde{G^{-}}(p) \times \tilde{R}(p)), \tag{28}$$

where $\tilde{\theta} = B - \tilde{\beta}X - \varepsilon$ *is the bidder's transformed type. The optimal bid function and it's inverse are both one-to-one and increasing.*

3.5 Competitive Benchmark

It is easy to generalize the notion of the Competitive Benchmark for the case of heterogeneous items.

Consider an auction market characterized by the valuation distribution $\mathcal{F}$ of the bidders, reservation value distribution $\mathcal{R}$, the bidder arrival rate λ_{bidder}, and the item arrival rate λ_{item}.

Definition 2. *The competitive benchmark of the auction market* $(\mathcal{F}, \mathcal{R}, \lambda_{bidder}, \lambda_{item}$ *is a perfectly competitive market, characterized by the demand function*

$$D(p) = (1 - \mathcal{F}(p))\lambda_{bidder} \tag{29}$$

and the supply function

$$S(p) = \mathcal{R}(p)\lambda_{item}; \tag{30}$$

the competitive price is determined as the solution p^* *of the equation*

$$D(p^*) = S(p^*). \tag{31}$$

Under Assumption 1, the observed reserve price distribution $\tilde{R}$ can be used to identify the competitive price and quantity.

4 Identification

As was discussed in Section 2, due to the institutional structure of the online markets, the only reliable data comes from the first and second highest bids in the auctions. Furthermore, the data on the first highest bids is not available publicly and therefore is missing for many researchers. In my dataset, that have been purchased from a market-maker, I do observe both first and second highest bids in the auctions. However, I try whenever possible to provide identification from the second highest bids only. Therefore my methodology can be used by the researches who only can work with the publicly available data. The data on the first highest bids is used for external validation of the results.

I first provide the identification for the case of homogeneous items, since this case is more transparent. Then I generalize the results to the case of heterogeneous items. I also first concentrate on zero reserve case, and later provide the results for general reserve price distribution.

4.1 Homogeneous items case

Let me introduce some notation. I denote by $\mathcal{F}(\cdot)$ the distribution of the type θ of the bidders in the population. Then $F(\cdot)$ denotes the distribution of the optimal bids for the bidders in the population (given by the bid function $B(\theta)$ determined by (6)). I also refer to the bidder's optimal bid as her *transformed type*, since the optimal bid is a one-to-one function of the bidder's type. $G(\cdot)$ denotes the stationary distribution of the first highest bid in an auction. $H(\cdot)$ is the stationary distribution of the second highest bid in an auction. As before, $G^{-}(\cdot)$ denotes the highest "rival" bid distribution, that is a stationary distribution of the highest bid in the auction when one bidder is randomly removed from the set of bidders.

Lemma 4. *Consider a stationary auction model with homogeneous items, known parameters and zero reserve. Assume, the transformed type distribution function $F(\cdot)$ is strictly increasing and continuous. There exist strictly increasing transformation functions $g : [0,1] \to [0,1]$, $g^{-} : [0,1] \to [0,1]$, and $h : [0,1] \to [0,1]$*

such that for any b

$$G(b) = g(F(b)), \tag{32}$$

$$G^{-}(b) = g^{-}(F(b)), \tag{33}$$

$$H(b) = h(F(b)); \tag{34}$$

The transformation functions $g(\cdot)$, $g^{-}(\cdot)$, and $h(\cdot)$ are determined by the model parameters, but do not depend on the distribution $F(\cdot)$.

Proof. The proof is providing a construction, that is used in my estimation so I present it in the main body of the paper. Suppose, in the original stationary process each bidder's transformed type b is replaces by $b' = F(b)$, while all the probability realizations remain unchanged. Since F is a strictly increasing function, it will not affect the relative order of the bids submitted in any auction, therefore the identity of the first and second highest bidders will remain the same. (Notice, that this result is driven by zero reserve assumption; with nonzero reserve distribution the event of winning an item can be affected by the applied transformation). Each process realization will remain the same, but the bidder's types are now "coded" by different numbers.

Notice that the new bids b' are drawn from the uniform distribution: $P(b' \leq x) = P(F(b) \leq x) = P(b \leq F^{-1}(x)) = F(F^{-1}(x)) = x$.

Let $g(\cdot)$ and $h(\cdot)$ be the first and second highest bid distribution functions in the new process, $g^{-}(\cdot)$ be the highest bid distribution after one active bidder is removed randomly. Since, except the bid "coding", the two processes are identical, it must be $G(b) = g(b')$, $G^{-}(b) = g^{-}(b')$, and $H(b) = h(b')$. Therefore $G(b) = g(F(b))$, $G^{-}(b) = g^{-}(F(b))$, and $H(b) = h(F(b))$, so $g(\cdot)$, $g^{-}(\cdot)$, and $h(\cdot)$ are the desired transformation functions. Those functions are determined by the model structure and the uniform distribution. They do not depend on $F(\cdot)$. The strict monotonicity property is easy to establish. □

This result is analogous to the well-know observation (see Athey and Haile (2001)), that the distribution of an order statistic of a sample of N independent draws from a given distribution can be obtained by applying an increasing transformation function to the parent distribution. Although in my model the relationship between the distributions is more complicated, the parallel result still holds. The crucial point is that there is no need to know the shape of the distribution $F(\cdot)$ in order to obtain the transformation functions.

In a homogeneous item setting the transformed type of a bidder is her bid. As soon as the second highest bid data is available to the researcher, the stationary distribution function $H(\cdot)$ can be identified. Then one could apply the *inverse* of the transformation function to identify the parent distribution: $F = h^{-1}(H)$. Applying another transformation function one can then can identify the first highest "rival" bid distribution: $G^{-} = g^{-}(F)$.

The above construction allows me to establish the following result:

Proposition 5. *In the homogeneous item model with known parameters and zero reserve the population bid distribution $F(\cdot)$, as well as the distribution of the highest "rival" bid $G^{-}(\cdot)$, are nonparametrically identified from the second highest bid data.*

Notice that as soon as the data for the highest bid for each auction is available, based on the results of Lemma 4, I can establish the analogous results identifying the parent distribution $\mathcal{F}(\cdot)$, and the highest "rival" bid distribution from the highest bid data. I can state the following proposition parallel to the Proposition 5.

Proposition 6. *In the homogeneous item model with known parameters and zero reserve the population bid distribution F, as well as the highest "rival" bid distribution $G^{-}(\cdot)$ are nonparametrically identified from the highest bid data.*

The following estimation strategy can therefore be implemented: first, estimate the parameters α, δ, λ_{bidder} and λ_{item} from the data (see Section 6.1 below for details); second, simulate the environment using the uniform distribution to obtain the estimates for the transformation functions h or g, and g^{-}; third, obtain the estimate for $H(\cdot)$, or $G(\cdot)$ from the data and apply the transformation functions to get estimates for F and G^{-}; finally, apply the theoretical result from Section 3.1 (equation (6)) to recover the parent valuation distribution $\mathcal{F}(\cdot)$.

I summarize the findings of this section in the following Theorem:

Theorem 1. *In the homogeneous item model with known parameters and zero reserve the population valuation distribution $\mathcal{F}$, is nonparametrically identified from either highest, or second highest bid data.*

Applying the results of Propositions 5 and 6 I can get two competing estimates of the underlying optimal bid distribution. The results of Theorem 1 allow me to obtain the two competing estimates of the underlying valuation distribution. Comparing the alternative estimates of the same distributions allows me to evaluate the "goodness of fit" between the market structure proposed by the model and the actual auction environment.

4.2 Heterogeneous items case

Let θ denote the bidder's type, so that an item valuation by the bidder is given by (16). The distribution of bidders' types in the population of bidders is denoted by $\mathcal{F}(\cdot)$. For a bidder of the type θ the optimal bid for an item with the characteristics (X, ε) is determined by the item characteristics and the bidder's transformed type $\tilde{\theta}$ given by (23). According to (24) the transformed type can be determined as the bidder's bid for an item cleaned from the item characteristics. I will denote the distribution of the transformed types in the population of bidders as $F(\cdot)$. Although $F(\cdot)$ is not a model fundamental, it is easier to identify then the parent distribution $\mathcal{F}(\cdot)$. In what follows I first provide the identification for $F(\cdot)$ and the highest rival bid distribution $\tilde{G}^{-}(\cdot)$ (defined by (19)). Then, using (25) or (28), the parent distribution $\mathcal{F}(\cdot)$ can be identified.

The data available is coming from the first and second highest bids in the auction. Let me denote by $G(\cdot|X, \varepsilon)$ the distribution of the highest bids in the auctions, and by $H(\cdot|X, \varepsilon)$ the distribution of the second highest bids in the auctions. As I did before for the distribution of the highest "rival" bids, I make now the following assumptions:

Assumption 6. *The distribution of the highest bids in the auction can be represented as*

$$G(b|X, \varepsilon) = \tilde{G}(b - \beta X - \varepsilon). \tag{35}$$

Assumption 7. *The distribution of the second highest bids in the auction can be represented as*

$$H(b|X, \varepsilon) = \tilde{H}(b - \beta X - \varepsilon). \tag{36}$$

I will now concentrate on identifying the parent distribution from the second highest bid data. The identification from the highest bids is similar.

Notice, that while the second highest bids for each auction and the observable auction characteristics can be used in the estimation, the item unobservable term ε makes identifying the distribution H more difficult. One way to address this problem would be by applying the methodology introduced in Krasnokutskaya (2004) to a sample of bidders who repeatedly occupy the second highest position in an auction. However, in the current paper I follow a simpler way of reasoning, by only using the data on "new, sealed" items for the estimation and claiming that for this category of items I can simply ignore the item unobserved term. Thus I assume that the stationary distribution $\tilde{H}$ can be estimated by the empirical cdf of the second highest bids in the auctions cleaned from the observables by linear regression.

After cleaning the second highest bids from the observables the whole reasoning repeats what I have done in Section 4.1. The following results can be established.

Lemma 5. *Consider a stationary, heterogeneous item model with known parameters and zero reserve. Assume, the transformed type distribution function $F(\cdot)$ is strictly increasing. There exist strictly increasing transformation functions $\tilde{g} : [0,1] \to [0,1]$, $\tilde{g}^- : [0,1] \to [0,1]$, and $\tilde{h} : [0,1] \to [0,1]$ such that for any b*

$$\tilde{G}(b) = \tilde{g}(F(b)), \tag{37}$$

$$\tilde{G}^-(b) = \tilde{g}^-(F(b)), \tag{38}$$

$$\tilde{H}(b) = \tilde{h}(F(b)); \tag{39}$$

The transformation functions $\tilde{g}(\cdot)$, $\tilde{g}^-(\cdot)$, and $\tilde{h}(\cdot)$ are determined by the model parameters, but do not depend on the distribution F.

Using the same identification procedure as in Section 4.1 I can establish the following result.

Proposition 7. *In a heterogeneous item model with known parameters and zero reserve the population transformed type distribution F, as well as the distribution $\tilde{G}^-$, are nonparametrically identified from the second highest bid data and item observables.*

If the data on the first highest bids for each auction is available, this data also should be cleaned from the observables to establish the following result:

Proposition 8. *In a heterogeneous item model with known parameters and zero reserve the population bid distribution F, as well as the distribution $\tilde{G}^-$, are nonparametrically identified from the highest bid data and item observables.*

Applying the theoretical result, provides by (25) one can therefore establish the following Theorem:

Theorem 2. *In a heterogeneous item model with known parameters and zero reserve the population valuation distribution $\mathcal{F}$, is nonparametrically identified from either highest, or second highest bid data.*

The two confronting estimates of the distribution F based on the results of propositions 7 and 8, and the two estimates of $\mathcal{F}$ provided by the Theorem 1 allow me to evaluate the "goodness of fit" provided by the model to the data.

4.3 Non-zero Reserve Price

Suppose, that the minimal bids in the auctions (the reserve prices) are drawn randomly from some known distribution $R(\cdot)$. In practice, the reserve prices in the auctions are often observed, so the distribution $R(\cdot)$ can be identified. I formulate the results for the case of homogeneous items. Under assumptions 2-3, 4, 6, and 7 those results can easily be generalized to the heterogeneous items case.

If reserve prices are drawn from a non-zero distribution, the statements of Lemma 4 do not hold any longer: the transformations functions $g(\cdot)$, $g^-(\cdot)$, and $h(\cdot)$ relating the parent distribution $F(\cdot)$ and the distributions $G(\cdot)$, $G^-(\cdot)$, and $H(\cdot)$ of the first, first "rival", and second highest bids in the auctions *depend* on the parent distribution. Indeed, they depend on the relationship between the parent distribution $F(\cdot)$ and the reserve price distribution $R(\cdot)$. In this case the distribution $F(\cdot)$ can be identified as a fixed point of an operator that can be constructed in the following way:

Proposition 9. *Consider a stationary auction model with known parameters and the reserve price randomly drawn from the distribution $R(\cdot)$.*

1. *For each distribution F' of the bidders' transformed type the transformation functions $g(\cdot|F')$, $g^-(\cdot|F')$, and $h(\cdot|F')$ can be identified, that relate the distribution F' and the resulting distributions G', G'^- and H' of the first, first "rival", and second highest bids in the auctions:*

$$G' = g(\cdot|F')$$
$$G'^- = g^-(\cdot|F')$$
$$H' = h(\cdot|F').$$

2. *If the distribution G of the first highest bids in the auctions is identified, then the parent distribution F can be identified, as the fixed point of the operator γ:*

$$\gamma(F) = g^{-1}(G|F). \tag{40}$$

3. *If the distribution H of the second highest bids in the auctions is identified, then the parent distribution F can be identified as the fixed point of the operator η:*

$$\eta(F) = h^{-1}(H|F). \tag{41}$$

The equations (40) and (41) can be solved numerically by iteration to obtain two estimates of the parent bid distribution F: $\hat{F}_1(\cdot)$ and $\hat{F}_2(\cdot)$. Then the highest "rival" bid distribution $G^-(\cdot)$ can be estimated as $\hat{G}_1^-(\cdot) = g^-(\hat{F}_1|\hat{F}_2)(\cdot)$ and $\hat{G}_2^-(\cdot) = g^-(\hat{F}_2|\hat{F}_2)(\cdot)$. Applying the theoretical result (28) allows me to compute two competing estimates of the parent valuation distribution. Comparing the two competing estimates for parent bid distribution F and parent valuation distribution $\mathcal{F}$ allowes me to evaluate the "goodness of fit" of the model.

5 Data

The estimation uses the data purchased from a company affiliated with a major internet auction marketplace[3]. The dataset contains the information on all listed items in the market's category that corresponds

[3]The contract agreement does not allow me to reveal the source in writing

to 60 GB model of Apple iPod Classic. After cleaning the data I ended up with a sample of 8325 listings, each of them auctioning a single 60 GB iPod via standard auction mechanism. (See Appendix A for the description of the auction market.) This model of an iPod have been released in October 2005 and have been the most advanced iPod model available for the period of observation. The price of a new item sold directly by Apple have been $399 plus tax[4]. The nearest substitute for this model in the market have been the analogous 30 GB model, that was available from Apple at a price of $299 plus tax.

For each listing I have data on item characteristics (including start and end time of the auction, minimal bid size, secret reserve flag, seller's feedback score, item condition, item color, and others). I also have a list of all bids submitted for each item accompanied by the bidder's identity, time, and size of the bid. Notice, that unlike most datasets collected by monitoring a marketplace, my data contains *all* bids submitted for the items, including the highest bids that are typically unobservable.

In my estimation I use the available data through two different channels. Although I observe all the bids submitted in each auction, only the two highest bids in each auction make a reliable source for the distribution estimates. However the remaining bids may be used in order to monitor each bidder's activity over time. To pursue those to separate goals I form the two datasets based on my original data. An item-level dataset (dataset 1) summarizes the information available about each item and includes the two highest bids submitted in the auction. A bidder-level dataset (dataset 2) summarizes the activity of each bidder over time as well as the winning of items by bidders.

Let me now discuss each of the two datasets in more detail.

The fields included in dataset 1 for each of the 8325 items are listed in Table 1. Since I do not have a reliable data on the shipping price of the items so I do not include the shipping price in this study. Table 2 provides summary statistics of the whole sample of item level data, while Table 3 summarizes the part of the sample that only consists of 1109 new, sealed items. The average starting price for an item was $77.29, with the minimum of $0.01 and the maximum of $801. The starting price is slightly lower for the new, sealed items: average is $72.12, minimum is $0.01 and the maximum is $445. In the whole sample 941

[4]The amount of tax could vary among states, the estimated amount of tax for New York State is $34

sellers (11.3%) use the Secret Reserve option. This fraction is lower in the new, sealed subsample: only 111 sellers (10.0%) of the new, sealed items use the Secret Reserve. The average success rate (the fraction of the items sold) for the whole sample is 0.87. It is slightly higher for the new, sealed items at 0.90. 56.3% of all items and 64.9% of new, sealed items are black. The sellers feedback[5] for the whole sample varies from -1 to 144011 with the average seller feedback equal 10959. The sellers of new, sealed items are on average less experienced, with the feedback score varying from 0 to 13561 with the average of 447. The average highest bid for an item is $305.79 with the standard deviation of $108[6], the average second highest bid for the sample is $283.79 with the standard deviation of $107, the average item price is $294.85 with the standard deviation of $101.04. Those values are higher for the new, sealed items. The first bid has an average of $326.53 with the standard deviation of $97.65, the second bid is $309.04 with the standard deviation $97.02, and the average price is $316.79 with the standard deviation of $92.94. 13.3% of the sample are the new, sealed items.

The second dataset summarizes the bidding activity of the bidders. I observe 29239 bidders who submitted at least one bid in my dataset. For each of those bidders the dataset consists of the set of moments when the bidder have submitted a bid. I identify those moments as the times when the bidder have been active. Table 4 provides some summary statistics of the bidder activity. A bidder in my sample have been active on average 5.43 times during the observation period with the minimum of 1 bid submitted and the maximum of 536 bids submitted for the items in my sample. Among 29239 bidders, 19873 (68%) were active more than once in the sample. A bidder have participated in the bidding for 2.62 auctions on average, with the minimum of 1 auction and maximum of 346 auctions. Among all bidders 11755 (or 40%) participated in more than one auction. The average number of items won by a bidder is 0.24 with the minimum of 0 and maximum of 59. Among all bidders in my sample, 6417 (22%) were winners in an auction at least once, 525 bidders (1.8%) won more than one auction. The time spent by a bidder in the sample varies from 1 second to 96 days. The mean time spent in the sample is 76 hours with the standard deviation of 230 hours. 9377

[5]Measured as the difference between the number of positive feedback entries and the number of negative

[6]Those numbers are computed for both sold and unsold items, if not bids were submitted for an item the highest bid is assumed to be zero. Same applies to second highest bids and item prices.

bidders (32%) appear in the sample only once. Among the remaining 68% of bidders the mean of the time spent in the sample is 112 hours and the standard deviation is 271 hours. The median time spend in the sample conditional on being observed more than once is 6 hours.

I do the following adjustments in the bidder-level dataset. Although I assume unit demand in my model, there are bidders in the dataset, who have demand for more than one item. Those bidder continue submitting bids even after winning an item. In order to fit those bidders in my model I represent them as separate identities, assuming that the same bidder have independent valuations for each subsequent item purchased. Therefore, for each bidder who wins an item in one of the auctions I create a separate entry in my dataset for the bids submitted after winning. Some bidders generate more than two identities in this process. Overall, through this process I generate 913 new identities of the bidders (3% of the data).

Further, I separate a bidder to multiple identities in case if the same bidder is only active in the periods very distant in time. For this estimation I use the 20 days cutoff level, meaning that if a bidder is not being active in the sample for 20 days I create a separate entry for the bids submitted afterwards. This is a necessary adjustment, to make me able to identify exit of the bidders from the pool. Overall, 878 new identities have been created through this process (another 3% of the data).

This process results in 31269 bidder identities. I summarize all the information available by listing the activity times for each bidder.

One more piece of information I include in this dataset is the variable that indicates for each bidder whether or not she won an item, and, if she did, the item ID. Notice, due to prior adjustment, each bidder in the sample wins at most one item.

6 Estimation Procedure

6.1 Parameters Estimation

For the purpose of parameter estimation I use the sample of all relevant items listed for auction and not sold by Buy It Now (8325 items).

For the need of parameter estimation I discretize time in 10 minutes periods, making 13248 periods of observation.

There are four parameters in the model: λ_{bidder} – the arrival rate of the bidders, λ_{item} – the arrival rate of the items, α – the activity rate of the bidders in the pool, and δ – the survival rate of the bidders in the pool. The three parameters λ_{bidder}, α, and δ are estimated based on the bidder-level data (dataset 2), and λ_{item} is estimated from the item-level data (dataset 1).

I observe a panel of bidder activity that shows for each bidder and each period whether or not the bidder was active in this period.

Although the concept of "activity" in the baseline model (being a potential bidder for the ending auction) and observed "activity" just described (submitting a bid for any item) are quite different, I assume them to be identical for the purpose of parameter estimation.

In order to estimate the three bidder-side parameters, I compute the size of the pool of the bidders for each period of time. I assume each bidder being in the pool from the first period she is active to the last such period. Notice, that such an estimate for the entry and exit time would be biased: the entry period of a bidder might be any moment before the bidder's first activity, the exit moment is any moment after the last activity period. In order to correct for this bias, I introduce a parameter d to be the average interval between the activity times for a bidder in the pool (this parameter is estimated as $d = 7.5$ hours), and assume the entry time to be $d/2$ earlier and the exit time $d/2$ later, then the observed entry and exit. Given this assumption, I compute the number of observed bidders in the pool for each of the periods. Notice, however, that given my assumptions those estimates are only valid for the central periods of the sample. Since I have no information about bidder's activity for the periods outside of my 3 month data, I cannot properly estimate the size of the pool for the first 20 days of the sample. Indeed, suppose a bidder have submitted a bid just one period before my sample starts and then next time was active one period before day 20. Then, according to my definition, this bidder have been in the pool for the first 20 days, however, I cannot distinguish between this bidder an the one who would only enter the pool a period before day 20. Therefore I do not use the first 20 days of data in my estimation. The same argument holds for the last

20 days. However, another complication arises since the bidders could bid for the items that expire after my sample ends and therefore not included in my sample (however the bidding itself takes place within the period of interest). Since the longest duration of an item is 10 days, this affects the last 10 days of the data. The two effects together force me to exclude the last 30 days from the estimation. As a result, I have to exclude the first 20 and the last 30 days of the data from consideration when estimating the size of the pool and related parameters. In practice, I exclude a little more than that: I exclude the first 3000 periods and the last 4500 periods, therefore the estimation is based on the data of the remaining $T = 5748$ periods of data. The estimated size of the pool varies between 656 and 984 bidders, with the mean of 797 and the standard deviation of 64 bidders.

After computing the size of the pool for each period, I compute the number of bidders being active (submitting a bid) in each period. The number of active bidders for a 10 minutes span varies from 0 to 152 with the mean of 14 and the standard deviation of 12 bidders. I compute the fraction and use the average value as an estimate for the parameter α of bidder's activity rate:

$$\hat{\alpha} = \frac{1}{T} \sum_{t=4501}^{4500+T} \frac{Active_t}{Pool_t},$$

the estimated value is $\hat{\alpha} = 0.0172$.

To estimate the parameter δ, I use the exit data of the bidders from the pool. Aside from winners, the number of bidders who exit the pool each 10 minutes varies from 0 to 16, with the mean of 1.89 and standard deviation of 2.19 bidders. I compute the fraction of the bidders who exit the pool without winning an item each period to the total number of bidders in the pool at this period minus the number of bidders who exit the pool by winning an item this period; and then average this fraction over periods. This forms the exit rate of the bidders. Subtracting this value from one I get the survival rate:

$$\hat{\delta} = 1 - \frac{1}{T} \sum_{t=4501}^{4500+T} \frac{Exit_t}{(Pool_t - Win_t)}.$$

The estimated value is $\hat{\delta} = 0.9976$.

The two events: exit by chance and exit by winning are supposed to be independent, therefore the probability of exit by chance should be the same for the whole pool, for the winners in the period, and for non-winners in the period; since the random exit for winners is not observed, only the proportion of non-winners who exit could be directly measured.

Finally, the bidder entry rate λ_{bidder} is estimated based on the pattern of entry of the bidders into the pool. The number of bidders entering the pool every 10 minutes varies from 0 to 19, with the mean of 2.05 and standard deviation of 2.56 bidder. The variance of the number of entering bidders is 6.58. Notice, that under Poisson assumption the mean and variance of the process both should be equal the process parameter. Thus, the Poisson assumption is indeed violated in my data. At least some part of the variance can be attributed to the day cycle that is present in my data, but which I ignore at this stage. I keep maintaining the Poisson assumption in my estimation for simplicity. I estimate the Poisson parameter as the average entry rate:

$$\hat{\lambda}_{bidder} = \frac{1}{T} \sum_{t=4501}^{4500+T} Entry_t,$$

the resulting estimate is $\hat{\lambda}_{bidder} = 2.5098$.

The parameter λ_{item} is estimated based on the item data. The estimate is rather straightforward:

$$\hat{\lambda}_{item} = \frac{Items}{Periods}.$$

This time there is no need to drop any periods since the item arrival is observed without any distortions. There is, however, one special consideration to be made. Through the course of my estimation there are two different assumptions I make about the reserve price in my simulation. In the initial simulations I assume zero reserve price for all items. It means that any item that gets at least one bid (that happens in simulation with probability close to one) is being sold. For this stage of the estimation it makes sense to estimate the arrival rate of the items based on *sold* items only:

$$\hat{\lambda}_{item,sold} = \frac{Items_{sold}}{Periods}.$$

This approach can be interpreted as substituting the true increasing supply curve by an “inverse L-shaped” one. For the later stages of my estimation I simulate the minimal bids for each item explicitly, allowing for high minimal bids and unsold items. Therefore at this stage it makes sense to use the estimate of the item arrival rate based on the whole sample[7]:

$$\hat{\lambda}_{item} = \frac{Items}{Periods}.$$

The estimated values are: $\hat{\lambda}_{item} = 0.6293$, $\hat{\lambda}_{item,sold} = 0.5462$. Table 6 provides the standard errors for the estimates obtained by model bootstrap (see Section 6.8 below).

6.2 Model Estimation

Provided the estimated parameter set, the model is estimated according to the following algorithm.

First, from the item-level data I estimate the empirical cdfs of first and second highest bids in the auction. This estimation uses only the data on *new, sealed* items in order to minimize the item heterogeneity.

In the model, item characteristics are neutral for the bidding behavior (both price of the item and the item's valuation by the winner are shifted by the same quantity by the item characteristics, holding the difference constant), therefore the *new, sealed* items constitute a random sample of all the items from the bidders' perspective.

My goal is to estimate the distribution of the bidders' valuation for the items net of the characteristics effect $\mathcal{F}$, but first I concentrate on estimation of the distribution of the transformed type F.

I observe the first and the second highest bids submitted for each item in the *new, sealed* sample. I use sold items in the sample for the estimation. First, I clean up this data from the effect of the remaining observables: item color, seller's feedback, and the absolute time of the listing (since the obsolesce rate of my items is high, I have a significant negative trend in the price over time).

To estimate the β coefficient of those observables I use the linear regression of the joint sample of first and second highest bids for each item on the observables.[8]. Although the coefficients of the regression are significant, the explanatory power ($R^2 = 0.097$) is relatively low. See Table 5 for the regression results.

[7]In practice I use a little more sophisticated formula to estimate this parameter. I compute the fraction for each month and then average the three estimates. It provides some variation for the estimate in the Bootstrap below.

[8]The estimates obtained when I use first and second bids separately do not differ significantly

After I estimate β coefficient from the joint sample, I compute the βX vector and subtract it from both second and first bid data. I use the same estimate for the β to clean the minimal bid data as well. This data will be used later to estimate the minimal bid distribution and the market supply.

Based on first and second bid data (FB_j, SB_j) from the "new, sealed" items, cleaned from the observables, I construct the two empirical cdfs:

$$\hat{H}(x) = \frac{1}{N_2} \sum \mathbb{1}(SB_j \leq x) \tag{42}$$

$$\hat{G}(x) = \frac{1}{N_1} \sum \mathbb{1}(FB_j \leq x) \tag{43}$$

Notice, that the number of observations in the two estimates differ from each other: for the estimate of the first bid cdf I use all the sold items for which I observe at least one bid (998 observations in my sample), while for the second bid cdf I only use those items for which at least two bids are observable (987 observation).

6.3 Transformation Functions

Now I have to estimate the transformation functions $h(\cdot)$ and $g(\cdot)$ that relate parent distribution F of the bids and the distributions of first and second highest bids for the items G and H. Since I observe a nonzero reserve price distribution in my data, I have to apply the estimation strategy based on Proposition 9, that requires iteration in order to estimate the parent distribution $F(\cdot)$. For the first iteration I assume that each item has a zero reserve price and simulate the model environment using the bidder's optimal bids drawn from a uniform distribution $U[0, 1]$. It allows me to get initial estimate for the parent distribution $F(\cdot)$ and then to update it by iterations.

I perform the simulation through the random length "periods" corresponding to each item being auctioned. Each time a new item is generated, I draw a random length of a "period" between the current and previous items from the exponential distribution (waiting time distribution of the Poisson process). This approach allows me to discretize the simulation of a continuous time model.

Notice, that I have to wait until my simulation converges to a stationary distribution before I can start estimating the transformation functions. Prior to the convergence, the bids submitted by the bidders in my simulation are not the true optimal bids the bidders would submit in this environment. In a non-stationary environment the optimal bid of each bidder would be time-dependent. However, it does not matter for the state evolution, and upon convergence the simulated bids correspond to the optimal bids in my desired model.

First, my simulation runs for 5000 "periods" to allow the pool size to get to a stationary size[9]. Then I run the simulation for another 20000 "periods" to generate the data.

For each "period" of the simulation I, first, generate the length t_j of the "period" j from the exponential distribution with the parameter $\lambda_{item,sold}^{-1}$ (this is the waiting time distribution for the Poisson process). Notice, that I use the arrival rate $\lambda_{item,sold}$ for this first simulation rather then λ_{item}. I choose to do so in the simulation where the reserve price is zero and therefore all items are sold. This is equivalent to using a step function as a first approximation for the reserve price distribution, assuming that each item either has a zero reserve price and therefore is sold with probability one, or is having a prohibitively high reserve price and thus sold with probability zero. In the following iterations where I draw the reserve price from each item from an estimated distribution, I generate the item arrival from the distribution with the parameter λ_{item}.

Then I generate the new bidders entering the pool this period. The number of new bidders is generated from the Poisson distribution with the parameter $\lambda_{bidder} \cdot t_j$. Each of the new bidders is assigned a unique number and a transformed type that corresponds to her optimal bid. The bidders' transformed types are drawn from the uniform distribution $U[0, 1]$.

For all the bidders in the pool I determine whether or not each bidder is active within the "period". For each bidder the probability to be active is $1 - (1 - \alpha)^{t_j}$.

All active bidders submit the bids for the item according to their transformed types.

The winning bidder is removed from the pool. Each of the remaining bidders remain in the pool with probability δ^{t_j} and exits with probability $1 - \delta^{t_j}$. Then the next "period" starts.

Notice, that the resulting process is a "discrete approximation" to the actual continuous time process,

[9] I do not run any formal test to verify the convergence, however the results do not change substantially as I increase the convergence time from 3000 to 5000 periods

since all entry and exit of the bidders occurs simultaneously.

I record the first and second highest bids for all items through 20000 periods. Based on this information I compute the estimated transformation functions:

$$g(x) = \frac{1}{K} \sum \mathbb{1}(fh_k < x) \tag{44}$$

$$h(x) = \frac{1}{K} \sum \mathbb{1}(sh_k < x) \tag{45}$$

For the estimate of $g(\cdot)$ I use all successfully sold items in my simulation, while for the estimate of $h(\cdot)$ I only select items with at least two bids submitted. Since I apply same selection to my data, this adjustment may seem to be minor, however, in fact, it illustrates a major problem with my Baseline model.

Unlike in the field data, in the simulation I observe a significant fraction of items (around 3%) where only one bid is submitted, and another 3% of the items with no bids submitted despite of low reserve. In the field data I only observe the items with one or no bid in case if the reserve price for the item is very high. It is different in the simulation. Therefore, in my simulation there is positive probability to buy an item at a zero or very low price. The reason why it happens in the Baseline model, is that I ignore the early bidding, therefore decreasing the number of potential bidders for each item. The best I can do within the Baseline model framework to address this issue, is to drop those items from the simulation. That would, however, introduce some bias in the estimate, in general I overestimate the size of optimal bids and valuations because of this bias in this estimation. I address this issue in the Extended version of my model (see Sections 8-9, where I model the early bidding explicitly.

Using the estimated transformation functions I form the two preliminary estimates for the parent distribution F based on the first and the second bids' distributions:

$$\hat{F}_1(x) = \hat{g}^{-1}(\hat{G}(x)) \tag{46}$$

$$\hat{F}_2(x) = \hat{h}^{-1}(\hat{H}(x)) \tag{47}$$

6.4 Reserve Prices

The estimation above assumes zero minimal bids for all items. This is a simplifying assumption I use to get a preliminary estimate of the parent bid distribution. In order to proceed with the iteration, I use the observed data on posted item "minimal bids" to estimate the reserve price distribution from which I then take random draws during the simulation.

To estimate the reserve price distribution I use the minimal bid data for the "new, sealed" items cleaned from the item observables component as described above.

One difficulty is introduced by the "Secret Reserve" price that is used by some sellers. If this option is used by a seller, the item is only sold if the outstanding price exceeds a certain amount posted secretly by the seller. For each item I have data on whether the "Secret Reserve" option was used by the seller or not, but the value of reserve price is not revealed in my data. In my dataset 941 items among the total of 8574 items, and 111 among the 1162 "new, sealed" items are using "Secret Reserve". Although the exact value of "Secret Reserve" is unknown, I can always construct an interval that contains the unknown value. If the item is sold (and the "Secret Reserve" option is on), the "Secret Reserve" must be between the minimal bid posted by seller and the sale price. If the item is unsold the "Secret Reserve" is above the outstanding price and below some high value, that I set equal to \$1000[10].

Since I need to use a specific reserve price distribution in simulation, I estimate the model under three different assumptions about the reserve price distribution. First, I assume, that I use the upper bound for the unobserved reserve price in each case, then I use the average between the upper and lower bound, and, finally, I use the lower bound for the estimate.

The types of the bidders are drawn from the estimate $\hat{F}_2$ obtained on the previous stage. Using $\hat{F}_1$ in the iterations provides an alternative estimate.

I proceed with the iterations until the maximum difference between the two subsequent estimated of $F(\cdot)$ gets below 0.02.

The estimation results are presented at Figure 2. The Figure presents the empirical cdf of the second

[10] The choice of this value does not affect the equilibrium price and quantity in any of the estimate

and first bids $(\hat{H}, \hat{G})$, the two estimates of the parent bid distribution $(\hat{F}_2, \hat{F}_1)$ and finally an estimate for the first bid distribution constructed from the second bid data: $\hat{G}_2 = \hat{g}(\hat{F}_2)$.

6.5 Valuation estimation

Based on my simulations I compute an estimate for the transformation function g^- that allows me to estimate the distribution of the highest rival bid in the auction. Notice, that when estimating this function I face once again the major problem introduced by the simple structure of my model: in my simulation there is a positive probability of getting no rival bidder and thus buying an item at zero or very low price. This property is not matched in my data. If I preserve the possibility of zero price in my estimation it would increase significantly the estimated valuation distribution of all bidders. I prefer to be conservative in my estimation by dropping all the simulations with only one bidder when estimating g^-. The resulting estimate of the highest rival distribution $G^-(\cdot)$ turns out to be very similar to the estimated distribution $G(\cdot)$. Given the estimates for the parent bid distribution $F(\cdot)$, the highest bid $G^-(\cdot)$, and the reserve price distribution $R(\cdot)$ (I use the average estimate for the reserve price distribution in my estimation), I apply the theoretical result obtained in Proposition 4 to get the estimate of the valuation distribution $\mathcal{F}$. I compute two competing estimates of the valuation distribution based on the two estimates of $F(\cdot)$ and $G(\cdot)$ obtained from the first and second price data.

The two estimates are illustrated in the Figure 3. The figure presents the two estimates of the parent bid distribution $F(\cdot)$ based on the first and second bid data, and the two resulting estimates of the valuation distribution $\mathcal{F}(\cdot)$. One can see that the corresponding estimates are reasonably close to each other except for the low values of the price. The Bootstrap section below will present a more precise characterization of the proximity between the estimates.

6.6 Supply Estimation and the Competitive Counterfactuals

The estimate of the supply curve in my model relies on the Assumption 1, that states that the sellers with high enough reservation values for the items state the reserve price “truthfully” to be equal their reservation values. To estimate the market supply curve I use the minimal bid data for the "new, sealed" items cleaned from the item observables. For each value of the price I compute the number of items with the reserve price less than or equal to the price value to estimate quantity supplied. An estimate for the market supply curve could be constructed as an estimate of the reserve price distribution multiplied by the item arrival rate λ_{item} and scaled to match the size of the market (the total supply equals 8325 items).

As I have discussed in sections 3.2-3.3, the curve estimated in such a way would differ from the true supply curve for low values of the price, however, under Assumption 1 I made, it will coincide with the true supply curve for high enough values of the price, that is sufficient to identify the perfectly competitive price and quantity. However, since the true valuation amounts for the sellers with low valuations are not observable, I cannot estimate the "sellers surplus" in my welfare analysis.

As discussed in Section 6.4 I do not observe the reserve price for some of the items, whose sellers post a "secret reserve" for the item. For those items I can provide the upper and lower bound for the reserve price. In model simulation I've been using three different assumptions about the reserve price distribution, by estimating the unobserved secret reserve either by upper bound, or by lower bound, or by the average between the two. I make the matching assumption in estimation of the supply curve when analyzing competitive counterfactuals.

The estimate of market demand is constructed based on the estimated population valuation distribution $\mathcal{F}(\cdot)$ multiplied by the Poisson rate of bidders' arrival, λ_{bidder} and then scaled to match the size of the market. For each price the quantity demanded is equal the number of the bidders with the valuations at or above the given price.

For comparison I also provide the estimates of the market demand in case if the dynamic incentives of the bidders are not taken into account (or if we assume the bidders to be "myopic"). This estimate is based on the assumption that the population valuation distribution coincides with the population bid distribution, therefore I can use the estimated distribution $F(\cdot)$ for constructing the estimated demand.

In my model I assume the sellers to post their reserve prices "truthfully", to be equal the actual reservation values they assign to the items. What would happen, if this assumption is violated? Notice, first, that for a seller with the binding reservation value (with a significant probability to receive a price below the reservation value in case of posting zero reserve), it could be only optimal to set a reserve price at or above the true reservation value. Setting a lower reserve price would result in some cases in selling the item below the reservation value thus bringing a negative utility. On the contrary, posting a price above the true reservation value could be justified in some cases, if the seller has a possibility to relist the item, in case if the auction is unsuccessful. If this is the case, my estimate for the supply curve would underestimate the quantity supplied at each price. That would mean, that my estimate of the perfectly competitive quantity is below the actual value, while the perfectly competitive price is overestimated. There is, however, a limit on those distortions. The highest possible estimate for the competitive quantity and the lowest estimate for the price would be obtained, if I assume, that, indeed, none of the reservation values of the sellers in the sample is binding. That would bring an estimated competitive quantity of 8325 items (all the items listed in the sample). I estimate the competitive outcome under this assumption as well.

6.7 Welfare Analysis

As I discussed above, I do not have reliable data on seller's costs (or reservation values) for the whole pool of sellers, therefore I cannot estimate the "sellers' surplus" in my model.

I provide the estimates for the consumer surplus in the auction market and under perfect competition. For comparison, I also present the results computed under assumption of "myopic" bidders, when the dynamic incentives of the bidders are not taken into account.

The most reliable estimate of the consumer surplus utilizes both first and second bid data. For each item sold I observe the bid, submitted by the winner and can therefore imply the winner's item valuation from the model. I then subtract the observed price and sum over the items in the sample to get the total surplus, or average among the items to get the surplus per item.

I can also propose an alternative estimate that only relies on the second bid data. I compute this estimate in a following way. I use the same estimate as above to determine the average price. However to compute the average value of a winner I draw 10000 values from the estimated distribution $\hat{G} = \hat{g}(\hat{F}_2)$ of the first bids.

Notice, this estimate only uses the second price data. Then for those bid values I compute the corresponding valuations and average between observations.

Under perfect competition, the consumer surplus is computed using the standard methodology. I report the estimates based on the alternative estimates of the demand function.

6.8 Bootstrap

In order to get the confidence intervals for the estimates I use the bootstrap procedure described below.

I treat the item-level data (dataset 1) as an independent sample of items. This assumption is not strictly correct, since the size of the pool has autocorrelation over time and therefore the number of bidders for each item and the item prices may be auto-correlated. In practice one observes significant day cycles in the data. I ignore those effects at the current stage. Similarly, I treat the bidder-level data (dataset 2) as an independent sample.

The arrival processes of items and bidders are independent in my model. However the exit of the bidders from the pool through winning is closely related to the item expiration. This is why I always link an item and the bidder who wins this item.

The bootstrap procedure is as follows. First I draw a sample of 8325 observations from the items sample with replacement. Together with each item I draw bidder observation corresponding to the winner of the item. Then I draw an additional independent sample of 23455 bidders from the sample of non-winning bidders. This procedure allows me to obtain two samples from the datasets, such that each sample is independent but the two are interconnected in a way consistent with the model. Using bootstrap re-sampling data I repeat the estimation procedure described above for 1000 times.

7 Baseline Model Estimation Results

The estimated distribution for the Baseline Model are presented at Figure 2[11]. The Figure presents the empirical cdf of the second and first bids $(\hat{H}, \hat{G})$, the two estimates of the parent bid distribution $(\hat{F}_2, \hat{F}_1)$ and finally an estimate for the first bid distribution constructed from the second bid data: $\hat{G}_2 = \hat{g}(\hat{F}_2)$.

Figure 3 presents the two estimates of the parent bid distribution $F(\cdot)$ based on the first and second bid data, and the two resulting estimates of the valuation distribution $\mathcal{F}(\cdot)$. One can see that the corresponding estimates are reasonably close. Below I present a more precise characterization of the proximity between the estimates.

Figure 4 provides the goodness of fit test that compares the two competing estimates of the parent bid distribution based on the second and first bid data. On this Figure I present the 95% confidence interval[12] for the estimate $\hat{F}_2$ obtained from the second bid data together with the estimate $\hat{F}_1$ obtained from the first price data.

Figure 5 provides similar test that compares the two competing estimates of the parent valuation distribution based on the second and first bid data. I present the 95% confidence interval for the estimate $\hat{\mathcal{F}}_2$ obtained from the second bid data together with the estimate $\hat{\mathcal{F}}_1$ obtained from the first price data.

One can see that with minor exceptions (for the prices between \$363 and \$393) the first-bid-based estimate for the bid distribution falls into the confidence interval for the second-bid-based estimate. Same is true for the valuation distribution with the minor deviation in the interval between \$394 and \$578. (This deviation explains the difference in the mean estimates of the consumer surplus, but does not affect the competitive outcome).

Table 7 reports the estimated results for the competitive counterfactuals comparing with the auction market performance. The Table provides the results for the three different assumptions on the reserve price distribution: the estimates based on the upper bound of the unobserved reserve price, the lower bound,

[11] Here I report the figures, constructed by the estimate using the average assumption for the reserve price distribution. The other assumptions produce similar results

[12] Those are 2.5% and 97.5% Bootstrap quantiles, computed point-wise. The actual size of the test is still to be established.

or on the average between the two. I also report the results for the "Limit" case, that corresponds to an assumption that the sellers do not submit reserve prices truthfully and indeed none of the sellers' reservation values is binding. Under each of the assumptions I provide two estimates based on the first and second bid data in each auction. For the auction market I report the mean price, quantity sold, the estimated consumer surplus (I use the more precise estimate, that exploits both first and second bid data), and the estimated consumer surplus per item sold. For the competitive counterfactuals I report the estimated competitive price and quantity, the estimated consumer surplus, and the estimated consumer surplus per item sold. For each model of perfectly competitive market I report the percentage deviation of the auction market outcome from the competitive counterpart. For each of the percentage difference I indicate the significance level.

The first panel of the Table presents the results under the assumption of strategic behavior by the bidders. My estimation indicates that both price and quantity delivered by the auction market are below the competitive counterparts. The distortion in price varies between 3.6% and 4.3% and is statistically significant. The distortion in quantity varies between 1.5 and 9.0%. The most conservative estimate based on the higher bound for an unobserved "Secret Reserve" price indicates the difference in not significant (or marginally significant if I use second bid data). The two less conservative assumptions indicate significant bias in quantity. The total consumer surplus obtained in the market sector over 3 month period is estimated to be \$569,849. The total consumer surplus provided by the auction market varies between \$557,443 and \$831,426. The difference varies between 0.2% and 26.4%, most of the estimates indicate the difference to be insignificant. (The estimates of the consumer surplus based on the second bid data are very noisy). I also report the numbers for the average consumer surplus per item sold, since those numbers are easier to interpret. The average surplus per item obtained by the auction market is \$78.76 or around 23% of the average price. The competitive values vary between \$75.90 and \$104.57 per item. Notice, that in general the estimates for the competitive surplus obtained from the second bid data are higher and more noisy.

I also analyze the limit case that may be achieved if I drop the assumption of the "truthful" behavior by the sellers. In this case the distortion in quantity gets larger and may rich 13.1%, the distortion in price get lower but remains statistically significant. The distortion in the consumer surplus is getting higher and becomes significant.

The second panel of the Table report the estimates for the case of "myopic" bidders.

The most important observation is that the numbers provided by this model are significantly different from the estimates that take dynamic incentives into account. The estimated price for the perfect competition get lower and now the price obtained by the auction market is significantly *higher* then the perfectly competitive counterpart. The quantity distortions get smaller. The estimates for the consumer surplus under each model, including the auction market get significantly lower. Now the consumer surplus of the whole auction market is estimates to be just \$100,617. The average surplus per item is estimated to be \$13.91 or just 4% of the average price.

While the Baseline model provides consistent estimates of the parent distributions, there is a major problem with this model, that is likely a source of a bias in the estimates. The problem is, that becouse of the simmultaneous bidding assumption the model generates a significant number of items with one or zero bids, despite of low reserve price. The best I can do within the model is just to drop those items from the estimation. However it does introduce a bias. In practice, those items with no late bids still receive some early bids, that would form the observed first and second highest bids in the data. Since the early bids are on average lower, then the late bids, droppong those items couse me to overestimate the bid (and therefore valuation) distribution. I address this issue in the remaining part of the paper, by introducing a model that incorporates the early bids explicitly.

8 Extended Model

Consider the following extension of the Baseline Model of Section 3.

Time is continuous. Items randomly arrive to the market following the Poisson process with the parameter λ_{item}. Each auction lasts for time T. The reserve price in each auction might be assumed to be zero, or is randomly drawn from a distribution $R(\cdot)$.

Each bidder has unit demand. Bidders enter the market randomly over time following the Poisson process with the parameter λ_{bidder}. Bidders, who have entered the market, form a pool of potential bidders. Upon item expiration the winner of the auction exits the pool. Bidders also exit the pool randomly. For a bidder currently in the pool the probability to stay in the pool for another t units of time conditional on non-winning is determined as δ^t, with probability $1 - \delta^t$ a bidder exits the pool. Each bidder i has a type θ_i that is drawn randomly from a population distribution $\mathcal{F}(\cdot)$. Given item j's characteristics (X_j, ε_j), the bidder's valuation of this item is

$$V_{ij} = \theta_i + \beta X_i + \varepsilon_j$$

where the vector of coefficients β is common for all bidders. The utility of the winner is determined as the difference between item valuation and the price paid:

$$U_{ij} = \theta_i + \beta X_i + \varepsilon_j - p_j$$

Each bidder currently in the pool gets "active" at certain moments, following another Poisson process. I denote by α the probability for a bidder to get active in a unit of time. Then for a time period of the length t the probability for a bidder in the pool to get active is $1 - (1 - \alpha)^t$. I assume this probability to be independent on the bidder's type and the time spent by the bidder in the market.

Any active bidder has an opportunity to submit bids for any of the open auctions. I restrict bidders who are currently the highest bidders for any of the auctions from submitting bids for other auctions. So, effectively, a bidder is only bidding for one item at a time. This is a reasonable assumption when we assume unit demand, at least for those bidders whose bids have a high enough probability to become winning bids. A bid may only be submitted for an item, if it exceeds the current *outstanding price*, that is determined as the second highest bid currently submitted for an item (I ignore the minimal increment).

Each active bidder who is not the highest bidder for any auction has to make two choices:

1. Which item she is going to bid for, and

2. How much she is going to bid for this item.

Assumption 8. *If a bidder exits the pool before winning an item her payoff equals zero.*

The Assumption 8 allows me to get a straightforward answer to the second of the two questions above. I will discuss the role of the Assumption after stating the result.

Lemma 6. *Consider a bidder i of the type θ_i and an item j with the characteristics (X_j, ε_j). Subject on a decision of the bidder i to bid for the item j and not to bid for any of the earlier items, the optimal bid is computed as*

$$B_{ij} = \theta_i + \beta X_j + \varepsilon_j - \tilde{\delta} V(\theta_i) \tag{48}$$

where $\tilde{\delta}$ is the expected probability to survive between two subsequent item's expiration times; $Value(\theta_i)$ is the bidder i's continuation value.

Proof. Suppose δ_1 is the probability for the bidder i to survive until the expiration time of the auction j. Subject on the decision to submit a bid for an item j and not to bid for any of the earlier items the payoff of the bidder i in case of losing the auction j is equal $\delta_1 \tilde{\delta} V(\theta_i)$. If the bidder wins the auction at price p_j the payoff is $\delta_1(\theta_i + \beta X_j + \varepsilon_j - p_j)$. The bid of the bidder i does not alter any of the two payoffs, but determines whether or not she wins the item. The bidder prefers to win the item if and only if $\delta_1(\theta_i + \beta X_j + \varepsilon_j - \tilde{\delta} V(\theta_i)) < \delta_1 p_j$. Therefore (48) provides the optimal bid of the bidder. □

Notice, how the Assumption 8 is used in the reasoning above. It allows me to compute the payoff of the bidder in case of winning an item. Specifically, it states that in case of exiting the market before the auction is completed (probability $1 - \delta_1$) the bidder gets a payoff of zero. An alternative assumption would be that the bidder gets the payoff $(\theta_i + \beta X_j + \varepsilon_j - p_j)$ even if she is not in the pool at the moment of the auction completion. This assumption might seem reasonable, however it implies a counterintuitive result: the earlier the bid is submitted for an item, the *higher* should be the bid. In contrast, as it is shown in Lemma 6, under the Assumption 8 the optimal bid of a bidder for a given item does not depend on the bid timing and is only determined by the bidder's and item's characteristics and the parameters of the market. In practice, an exit of a bidder from the market may be caused by a lot of different reasons, such as financial shock, acquiring an item from another source, change of preferences etc. I do not want to model those issues explicitly. The reduced form assumption of zero payoff upon exit while not being very realistic *per se*, creates reasonable incentives for the bidders in the model and therefore is adopted by the analysis in this paper.

The second question remains to be answered: which item should a bidder select to submit a bid. Addressing this question in terms of bidder's optimal equilibrium strategy happens to be an extremely difficult task. The main challenge is that the exact answer would depend on the information available to each bidder, including the relative timing of the forthcoming auctions, outstanding prices of the items, and the bidder's believes about the optimal strategies of the rivals. Taking all of this information into account would be extremely difficult computationally. Even if this computation is performed, it still remains an open question, if the bidders actually follow the proposed optimal strategy, or do they just follow some simple "rule of thumb" while selecting an item.

Instead of solving the model for the optimal bidding strategy, I propose a reduced form behavioral assumption that would describes the bidders' selection of the items. Under this assumption I estimate the model, based on the data available for the first and second highest bids in each auction and compare the estimation results[13].

Assumption 9. *Each bidder considers the items one by one, on the order of time remaining until expiration. A bidder submits a bid determined by* (48) *for the first item for which the value of optimal bid (determined by* (48)*) exceeds the outstanding item price.*

8.1 Identification in the Extended Model

The identification results for the Extended Model are parallel to those for the Baseline model and can be found in the Appendix C. Notice, that unlike the Baseline Model I do not have to identify the distribution $G^{-}(\cdot)$ since it is not used in my estimation of the Extended Model.

[13] I have also tried an alternative Assumption: *Each bidder considers the items one by one, on the order of time remaining until expiration. A bidder submits a bid determined by* (48) *for the first item for which the value of optimal bid (determined by* (48)*) exceeds the outstanding item price, and the outstanding price is not exceeding the expected item price given the item characteristics.* However, it turned out to provide inferior results, comparing to the Assumption 9, so I do not report the results in this paper.

8.2 Valuation Distribution

Unlike the Baseline Model, the Extended Model does not allow for a closed form solution. The following proposition provides a general result that allows me to evaluate the Valuation Distribution numerically.

Proposition 10. *The bidder's valuation can be determined as*

$$\theta = \frac{\tilde{\theta} - \tilde{\delta} P(win|\tilde{\theta}) E(p|win, \tilde{\theta})}{1 - \tilde{\delta} P(win|\tilde{\theta})} \tag{49}$$

where $\tilde{\theta}$ *in the bidder's transformed type;* $P(win|\tilde{\theta})$ *is the probability for a bidder with the transformed type* $\tilde{\theta}$ *to win an item before exiting the market; and* $E(p|win, \tilde{\theta})$ *is the expected price (cleared from the item's characteristics) that the bidder with the transformed type* $\tilde{\theta}$ *should pay for an item.*

Proof. See Appendix D. □

The two values $P(win|\tilde{\theta})$ and $E(p|win, \tilde{\theta})$ can be estimated during the simulation.

8.3 Item Sellers

The model of item sellers in the Extended model repeat the one of the Baseline model (see Section 3.2).

8.4 Competitive Benchmark

The notion of Competitive Benchmark is the same in the Extended model as in the Baseline Model, see Section 3.3, 3.5.

9 Estimation Procedure: Extended Model

In this section I provide the details of the estimation methodology for the Extended model, where they differ for the Baseline model.

9.1 Parameter Estimation

The parameter estimation is exactly the same for the two models.

The identification of the model parameter α as the probability of a bidder to submit a bid in a given period is more relevant in the Extended Model then it was in the Baseline Model, since according to the Extended Model each bidder submits a bid whenever active.

9.2 Empirical cdfs

The estimates for the distributions G and H are exactly the same for the extended model as for the Baseline Model.

9.3 Transformation Functions

Now I have to estimate the transformation functions $h(\cdot)$ and $g(\cdot)$ that relate parent distribution F of the bids and the distributions of first and second highest bids for the items G and H. I do it first by simulating the model environment using the bidder's optimal bids drawn form a uniform distribution $U[0,1]$ and assuming zero reserve price for each item. (Later I obtain an estimate of a parent bid distribution that I substitute for the uniform distribution in my simulation, while drawing the reserve price from the estimated distribution; then I iterate until convergence.)

I run the simulation through the random length "periods" corresponding to the expiration times of each item. As before, the "period" is an artificial construction corresponding to each new item arriving and designed to discretize the simulation procedure for a continuous time model.

First, my simulation runs for 5000 periods to allow the pool size to get to a stationary size. Then it runs for another 20000 periods to generate the data.

In the first stage simulation each item has a zero starting bid. I allow each item to exist for 150 "periods", however the lifetime of the item is never binding in my simulation: bidders never submit bids for items closer than 50 "periods" to expiration. I start with a set of 150 items.

Each period the following events take place. First, I generate the length of the period, meaning the time lag between the previous and current items' expiration times. The period length t_j is randomly drawn from the exponential distribution with the parameter $\lambda^{-1}_{item,sold}$. Since the optimal bids in my model do not depend on exact timing of items, I do not need to generate the information about the item's expiration times early in the item's life. Each period one new item appears in the end of the list. I generate a new item number for each item that will allow me to track the bids submitted for this item in the future.

Then I generate the new bidders entering the pool in this period. The number of new bidders is generated from the Poisson distribution with the parameter $\lambda_{bidder} \cdot t_j$. Each of the new bidders is assigned a unique number and a type that corresponds to her optimal bid. The bidders' types are drawn from the uniform distribution $U[0, 1]$.

Then for all the bidders in the pool I determine whether or not the bidder is active in this "period". For each bidder the probability to be active is $1-(1-\alpha)^{t_j}$. Then for the bidders active in this period I determine the random order in which they are going to submit bids this period. This is a reduced form setting that replaces the original Poisson process of my theoretical model.

Then I allow bidders to submit bids in the determined order. Each bidder, given an opportunity to bid considers the open items in order of their starting date (therefore consider first the items closest to expiration). For each item the outstanding price is computed as the second highest bid, previously submitted for this item if there are at least two bids, or zero otherwise. I ignore the minimal bid increments in my simulation. The bidder starts from the item closest to expiration and compares her optimal bid with the outstanding price of the item. If the optimal bid is above the outstanding price the bidder submits a bid and stops bidding for the period. Otherwise she considers the next item. In my simulation I never observe bidders considering more than 20 items.

After all bidders have submitted their bids I compute the outcome for the expiring item. I record first and second highest bids submitted and then remove the item from the list. The winning bidder is also removed from the pool. Each of the remaining bidders remain in the pool with probability δ^{t_j}, and exits with probability $1 - \delta^{t_j}$. Then the next "period" starts.

I record the first and second highest bids for all items through 10000 "periods". Based on this information I compute the estimated transformation functions:

$$g(x) = \frac{1}{K} \sum \mathbb{1}(fh_k < x) \tag{50}$$

$$h(x) = \frac{1}{K} \sum \mathbb{1}(sh_k < x) \tag{51}$$

Using the estimated transformation functions I form the two estimates for the parent distribution $F(\cdot)$ based on the first and the second bids' distributions:

$$\hat{F}_1(x) = \hat{g}^{-1}\hat{G}(x) \tag{52}$$

$$\hat{F}_2(x) = \hat{h}^{-1}\hat{H}(x) \tag{53}$$

The non-zero Reserve Price is introduced by the same procedure as in the Baseline Model. I use the three different assumption according the reserve price as discussed in Section 6.4.

I reevaluate the parent bid distribution estimate for the both models using the estimated minimal bid distribution to draw minimal bids for the items. This time I draw the period lengths from the exponential distribution with the parameter λ_i^{-1}. I have to iterate the process until convergence.

9.4 Valuation estimation

It was shown that the bidder's valuation can be determined as

$$\theta = \frac{B - \tilde{\delta}P(win|B)E(p|win, B)}{1 - \tilde{\delta}P(win|B)} \tag{54}$$

where B is the bidder's type measured by her optimal bid, $P(win|B)$ is the probability for a bidder of type B to win an item before exiting the market, and $E(p|win, B)$ is the expected payoff of this bidder given she wins an item, $\tilde{\delta}$ is the expected survival probability until the next item. The parameter $\tilde{\delta}$ can be estimated as $\tilde{\delta} = \frac{\lambda_i}{\lambda_i - log(\delta)} = 0.9963$. Given B, both $P(win|B)$ and $E(p|win, B)$ can be estimated during the simulation.

During the simulation I collect the data on each bidder's type, success status (win an item or not) and the price paid (if win). Then I use kernel estimate to evaluate the functions $P(win|B)$ and $E(p|win, B)$[14].

9.5 Supply Estimation, Competitive Counterfactuals, and Welfare Analysis

I use same methodology as for the Baseline model.

10 Extended Model Estimation Results

The estimated distributions for the Extended model are presented at Figure 6[15]. The graph represents the empirical cdf of the second and first bids $(\hat{H}, \hat{G})$, the two estimates of the parent bid distribution $(\hat{F}_2, \hat{F}_1)$ and finally an estimate for the first bid distribution constructed from the second bid data: $\hat{G}_2 = \hat{g}(\hat{F}_2)$. Figure 7 illustrates the two estimates of the population bid distribution together with the two estimates of the population valuation distribution. The corresponding estimates for the parent distributions seem to be very close. I will give qualification to this proximity below.

Figure 8 provides the 95% confidence interval[16] for the estimate $\hat{F}_2$ obtained from the second bid data together with the estimate $\hat{F}_1$ obtained from the first bid data. One can see that the first bid estimate falls in the 95% confidence interval for most of the bid range, with minor deviations in the interval of price between \$373 and \$393.

Figure 8 provides the 95% confidence interval for the estimate of the valuation distribution $\hat{\mathcal{F}}_2$ obtained from the second bid data together with the estimate $\hat{\mathcal{F}}_1$ of the valuation distribution obtained from the first bid data. One can see that the first bid estimate falls in the 95% confidence interval for most of the bid range, with minor deviations in the interval of valuation between \$392 and \$670.

Table 8 reports the estimated results for the competitive counterfactuals comparing with the auction

[14]The probability $P(win|B)$ could have been evaluated during the simulation already on the first stage, however the expected payoff $E(p|win, B)$ can only be measured if the bidders' bids are generated from an estimate of parent bid distribution.

[15]Here and below I report the figures, constructed by the estimate using the average assumption for the reserve price distribution. The other assumptions produce similar results

[16]I report here the Bootstrap quantiles. The exact size of the test is still to be established.

market performance. The Table provides the results for the three different assumptions on the reserve price distribution: the estimates based on the upper bound of the unobserved reserve price, the lower bound, or on the average between the two. I also report the results for the "Limit" case, that corresponds to an assumption that the sellers do not submit reserve prices truthfully and indeed none of the sellers' reservation values is binding. Under each of the assumptions I provide two estimates based on the first and second bid data in each auction. See Table 9 for the notation. For the auction market I report the mean price, quantity sold, the estimated consumer surplus (I use the more precise estimate, that exploits both first and second bid data), and the estimated consumer surplus per item sold. For the competitive counterfactuals I report the estimated competitive price and quantity, the estimated consumer surplus, and the estimated consumer surplus per item sold. For each model of perfectly competitive market I report the percentage deviation of the auction market outcome from the competitive counterpart. For each of the percentage difference I indicate the significance level.

The first panel of the Table presents the results under the assumption of strategic behavior by the bidders. My estimation indicates that both price and quantity delivered by the auction market are below the competitive counterparts. The distortion in price varies between 0.7% and 1.4% and is statistically significant on at least 10% level for all the estimates. The distortion in quantity varies between -0.1% and 7.6%. The most conservative estimates based on the higher bound for an unobserved "Secret Reserve" price indicates the difference in not significant. The two less conservative assumptions indicate significant bias in quantity. The total consumer surplus obtained in the market sector over 3 month period is estimated to be $615,406. The total consumer surplus provided by the auction market is significantly below perfectly competitive level under any assumption. The difference varies between 13.3% and 40.8%. Notice that the estimates based on the second bid data are very noisy. They also provide higher average estimates for the competitive total consumer surplus. If we rely on less noisy estimates based on the first bid data, the estimated difference varies between 13.3% and 19.9%. I also report the numbers for the average consumer surplus per item sold, since those numbers are easier to interpret. The average surplus per item obtained by the auction market is $85.04 or around 25% of the average price. The competitive values vary between

$96.36 and $151.99 per item. Notice, that in general the estimates for the competitive surplus obtained from the second bid data are higher and more noisy.

Notice, that the consumer surplus provided by the auctions is significantly below the competitive surplus despite the fact that the price paid by the bidders is significantly lower. There are two sources of the distortion in consumer surplus. First, the quantity delivered by the auction market is lower then the competitive level. Second, there is a significant distortion in the identities of the bidders who are served through the auctions as opposed to the competitive market. Indeed, in my model even the highest value bidders can exit the market without purchasing, just by random. On the other hand, a large number of the bidders are winning in an auction, who would not be served under perfect competition, since their valuations are below perfectly competitive price.

I also analyze the limit case that may be achieved if I drop the assumption of the "truthful" behavior by the sellers. In this case the distortion in quantity gets larger and may rich 13.1%, while the distortion in price get lower and may even switch sign. The distortion in total consumer surplus gets higher (23.2% for the first bid estimate) and remains significant.

The second panel of the Table report the estimates for the case of "myopic" bidders.

The most important observation is that the numbers provided by this model are significantly different from the estimates that take dynamic incentives into account. The estimated price for the perfect competition gets lower and now the price obtained by the auction market is significantly *higher* then the perfectly competitive counterpart. The quantity distortions get smaller. The estimates for the consumer surplus under each model, including the auction market, get significantly lower. Now the consumer surplus of the whole auction market is estimates to be just $100,787. The average surplus per item is estimated to be $13.93 or just 4% of the average price.

11 Second Best Mechanism

As was show in Section 10 there is a significant inefficiency present in the market. In this section I will analyze the market inefficiency in more detail and will provide an alternative mechanism that would improve market performance.

What is the source of inefficiency in a dynamic auction market, and what would be an efficient allocation achieved by a social planner?

If there are no inter-temporal constraints, and the central planner is allowed to redistribute items among bidders arriving at different times, the first best allocation would be provided by a competitive outcome. That would correspond to the case, when all items and all bidders are present simultaneously in the market.

Assume, however, that the central planner is facing the same inter-temporal constraint as the market participants: each of the items and each of the bidders is only available for a limited period of time. In this case, the second best solution for the central planner could be characterized as follows. Keep track of the pool of all bidders currently in the market. As soon as a new item arrives, allocate the item to a bidder in the pool with the highest valuation.

Notice, that the second best allocation could be achieved if all the bidders were active all the time while being in the pool. In this case, each of the bidders in the pool would submit a bid for each item arriving, and therefore the item would be allocated to the bidder with the highest valuation.

Therefore, we can conclude, that the main source of inefficiency is the inability of the bidders to monitor the market all the time. Furthermore, while it is allowed by market rules to submit bids simultaneously for multiple items, that would bring a risk of winning in multiple auctions – an undesirable outcome.

I therefore propose the following mechanism, that would approach as close as possible to second best, given the inter-temporal constraints faced by the bidders. Whenever the bidder is active in the market, allow her to submit *conditional bids* for as many open auctions as desired. The conditional bid remains active until the bidder wins an item and is being automatically withdrawn as soon as the bidder wins in one of the auctions[17]. Notice, that the bidder is allowed to submit different bids for different items to control for

[17] That could be easily extended to the case of multiple item demand, by allowing the bidder to specify the desired number

market heterogeneity. I will refer to this mechanism as the "second best mechanism" in the rest of the paper.

In the remaining part of this section I will provide simulation results that allow me to evaluate the performance of the second best mechanism with respect to the existing mechanism.

To be specific, I concentrate on one model estimate based on a particular set of assumptions about the model parameters. I estimate market demand (the distribution of bidders' valuations in the population) using my Extended Model based on the first bids. I estimate market supply as the average between upper and lower bound estimates. Those assumptions corresponds to model AV1 in Table 8[18].

Under my estimate the performance of the existing market characterized by 15.8% lower consumer surplus then the first best perfectly competitive outcome.

Without additional assumptions about sellers' costs it is hard to evaluate the seller surplus and total social surplus provided by the market. For the sake of this exercise I am making a simplifying assumption that the reserve price posted by each seller indeed represents her opportunity cost of the item. Under those assumptions I evaluate the Seller Surplus and Social Surplus provided by alternative mechanisms.

Provided the distributions of the reserve prices and bidder valuations, as well as the arrival, activity, and departure parameters estimated, I simulate the market performance under second best mechanism. I allow each bidder active in a particular point of time to place her bid for all items in the list, with bids automatically withdrawn when a bidder wins an item in any auction. While I have information about each bidder's item valuation, I still have to estimate the optimal bidding function in this environment. I use the optimal bid formula (48) for the estimation, where the value function $V(\theta)$ is evaluated for each type through simulation. I iterate my simulation until convergence in value function.

I then evaluate the consumer surplus as a difference between the value and price paid for each bidder, and the seller surplus as the difference between the price received and the opportunity cost. In order to control for simulation error, I repeat the simulation 100 times and report the mean and standard deviation for each variable of interest. The estimation results are presented in Table 10.

of items, after wining which the remaining bids are automatically withdrawn

[18]However, since I use a single estimate, not a Bootstrap mean, the results I report for the existing model will be slightly different from those in Table 8

The first observation is that the average price under the second best mechanism (p_{SB} = \$352.88) is significantly higher than either competitive price (p_C = \$343.55) or the existing average price (p_A = \$341.78). That is natural since the competition for each item is very intense under second best mechanism. Despite the high price, the consumers are better off under the second best mechanism then under the existing mechanism. While the existing mechanism provides the value of Consumer Surplus that is %15.82 below competitive level, the second best mechanism decreases the difference to just %10.73 of the competitive value. Further, because of higher prices the sellers are actually 4.01% better off under second best mechanism then under perfect competition (they were 13.62% worse off under the existing mechanism then under perfect competition). The qualitative result is robust to the simplifying assumptions I've made about costs. Since change in costs affects the Seller Surplus similarly under any mechanism, the benefit driven by the higher price will remain even if the actual costs are higher.

The negative impact on the bidders is almost offset by the positive impact on the sellers, with the Social Surplus under second best mechanism being just 0.24% below the competitive value (comparing to 14.18% loss in Social Surplus under existing mechanism). Notice, that this last finding is sensitive to the simplifying assumption on cost, I am making for this exercise. If the Sellers' opportunity costs are higher, then the posted reserve price, the relative impact on the Sellers surplus may be lower, resulting in a less impressive improvement in Social Surplus.

12 Conclusion

The central research question I address in this paper is how close does the auction market approximate perfectly competitive outcome. I discover that both average price and quantity delivered by the auction market are below the perfectly competitive level, and the consumer surplus delivered by the auction market is significantly below competitive level. Let me now provide some qualification to those findings.

Consider first the quantity. Since I only have limited data on the supply side of the market, the estimates for the competitive quantity vary in a wide range. The most conservative estimate indicates no significant difference in quantity between the auction market and the perfectly competitive market. This prediction sounds too optimistic: in the auction market that does not operate simultaneously one should expect certain inter-temporal frictions between sellers' and bidders' arrival to the market. Less conservative estimates of the supply curve, that maintain the assumption of the "truthful" behavior by the sellers indicate statistically significant deviation in price, that ranges between 4.2% and 7.6%. The upper bound of a possible bias in a model assuming strategic sellers is 13.1% of the quantity.

Now consider the price. The bias in price is statistically significant under any of the estimates in the Extended model. The main source of the price bias in the model is the inter-temporal considerations by the bidders that leads to bid shading. It is important to notice, that the estimates for the competitive price change significantly when I take the dynamic incentives of the bidders into account. In this context, it is important to verify whether or not the dynamic shading by the bidders actually takes place. The field data does not provide such a possibility: I only observe the bids submitted by each bidder and should use theoretical findings to recover the underlying valuations. It is possible, however, to address this question in a laboratory experiment, where both valuations and bids are observed to a researcher. I propose such an experiment as a part of my research agenda. If I take a possibility of strategic behavior by the sellers into account, the bias in price gets smaller and may even change sign.

The consumer surplus produced by the auction market is significantly lower, then the potential competitive level, despite the fact that the average price paid by a consumer is lower. The distortion in the consumer surplus can be attributed to two factors: first, the distortion in quantity sold, and second, the distortion in the identities of the winning bidders. The impact of the last factor is indicated by a significant difference between the consumer surplus per item sold. The reason why this distortion occurs is that unlike in perfect competition it is possible in the auction market, that a bidder with a high valuation exits the market without purchasing an item. Moreover many of the bidders who would not be served under competitive market (they value an item below competitive price) are actually being served through the auction. One should be careful interpreting the size of the distortion, since the estimates of the competitive surplus are very noisy. Another concern is that the estimate only takes into account the surplus obtained on the auction sector of the market ignoring the posted price ("Buy It Now") sector of the market.

The following extensions of the current approach may be suggested as the directions for further research.

The simplicity of the current model estimation is in a large extend due to the stationarity assumption made in the paper. It would be useful to relax this assumption by introducing "states" in the environment, such as the time of the day, items' characteristics and timing, etc. A related issue is incorporating more flexible model of item heterogeneity. It would allow to apply the model to wider variety of applications, since in many practical settings the items being sold next to each other are imperfect substitutes.

A model of strategic behavior by the sellers is worth to be analyzed. Aside of strategic choice of a reserve price, sellers may also vary timing of their auctions, and make a choice between different selling strategies, such as auction and posted price. Incorporating posted price sector ("Buy It Now") would also alter the bidder's choice set, by introducing an outside option for the bidders.

The bidders strategy proposed in the Extended model is very simple and mechanistic. It does not explain a significant amount of "very early" bidding that is observed in the data. Introducing a model that would provide a realistic within-auction bid dynamics, when between-auction choice is available to the bidders would definitely contribute to our understanding of the market.

Finally, the performance of the market can be significantly improved by using an alternative mechanism described in Section 11. While resulting in an average price above competitive, this alternative mechanism would allow to significantly reduce the loss in consumer surplus with respect to the current mechanism. While the sellers surplus is difficult to evaluate due to lack of cost data the rise in sale price is a strong evidence that the sellers are also better off. As a result the proposed mechanism represents a Pareto improvement over the existing mechanism and significantly improves market efficiency.

References

[1] Ariely, Dan, Ockenfels, Axel and Roth, Alvin E. (2005) "An experimental analysis of ending rules in Internet Auctions", RAND Journal of Economics, Vol. 36, No. 4., pp. 890-907

[2] Athey, S. and Haile, P. (2001) "Identification of Standard Auction models", Econometrica, 70(b), 2107-2140

[3] Athey, S. and Haile, P. (2008) "Nonparametric Approaches to Auctions", in Handbook of Econometrics

[4] Bajari, Patrick, Hortacsu, Ali (2003) "The winner's curse, reserve prices, and endogenous entry: empirical insights from eBay auctions", RAND Journal of Economics, Vol, 34, No.2, Summer 2003, pp. 329-355

[5] Bajari, Patrick, Hortacsu, Ali (2004) "Economic insights from internet auctions", Journal of Economic Literature, Vol XLII (June2004), pp.457-486

[6] Cabral, Luis, , Hortacsu, Ali (2006) " Dynamics of Seller Reputation: Theory and Evidence from eBay", mimeo, U. Chicago

[7] Ely J, Hossain T (2006)"Slipping and Squatting in Auction Markets", working paper

[8] Guerre, E., Perrigne, I. and Vuong Q. (2000) "Optimal Nonparametric Estimation of First-Price Auctions", Econometrica, 86, 525-574

[9] Hendricks, Ken and Porter, Robert H. (2007) "An Empirical Perspective on Auctions", in Handbook of Industrial Organization, Volume 3.

[10] Jin, Ginger Z., Kato, Andrew (2002) "Blind Trust Online: Experimental Evidence from Baseball Cards", working paper, University of Mariland

[11] Jofre-Bonet, Mirela and Pesendorfer, Martin (2003) "Estimation of a Dynamic Auction Game", Econometrica, vol 71, 1443-1489

[12] Jofre-Bonet, Mirela and Pesendorfer, Martin (2008) “Optimal Sequential Auctions”, working paper , LSE

[13] Krasnokutskaya, Elena (2004) “Identification and Estimation in Highway Procurement Auctions Under Unobserved Auction Heterogeneity”, Working Paper, University of Pennsilvania

[14] Krishna, Vijay (2002) Auction Theory, San Diego, Academic Press

[15] Lewis, Gregory (2007) ”Asymmetric Information, Adverse Selection, and Seller Disclosure on eBay Motors”, working paper, University of Michigan

[16] Lucking-Reiley, David (1999) “ Using Field Experiments to Test Equivalence Between Auction Formats: Magic of the Internet”, American Economic Review, 89:5, pp. 1063-80

[17] Melnik, Mikhail I., Alm, James (2002) “Does a Seller’s eCommerce Reputation Matter? Evidence from eBay Auctions,” J. Ind. Econ., 50, pp. 337-50.

[18] Nekipelov, Denis (2007) ”Entry Deterrence and Learning Prevention on eBay”, working paper, Duke University

[19] Paarsh, H., Hong, H. (2006) An Introduction to the Structural Econometrics of Auction Data. MIT Press, Cambridge, MA

[20] Peters, Michael, Severinov, Sergei (2006) “Internet Auctions with Many Traders”, Journal of Economic Theory, 130, pp. 220-245

[21] Ockenfels, Axel and Roth, Alvin E. (2006) ”Late and multiple bidding in second price Internet auctions: Theory and evidence concerning different rules for ending an auction”, Games and Economic Behavior, 55, 297-329

[22] Reznick, Paul; Richard Zeckhauser, John Swanson and Kate Lockwood (2006) “ The Value of Reputation on eBay: A Controlled Experiment”, forthcoming in Experimental Economics

[23] Roth, A. and Ockenfels, A. (2000) “Last Minute Bidding and the Rules for Ending Second-Price Auctions Theory and Evidence from a Natural Experiment on the Internet”, NBER Working Paper no. W7729

[24] Sailer, Katharina (2006) “Searching the eBay Marketplace”, Working Paper, University of Munich.

[25] Smith, Vernon (1962) ”An Experimental Study of Competitive Market Behavior,” Journal of Political Economy, 70(3), 111-137

[26] Song, Unjy (2004) “Nonparametric Estimation of an eBay Auction Model with an Unknown Number of Bidders”, Working Paper, University of British Columbia.

A The operation of an Online Market

In this appendix I describe the operation of an online market that constitutes the source of my data[19]. This scheme is typical for many markets in particular eBay that is used as a data source by other authors such as Song (2006) etc.

The market brings together sellers and buyers of a wide variety of items. A simple registration procedure allows any person to become a registered user and bid in the auctions, the procedure of becoming a seller is just a little more complicated. There are no barriers to entry.

Items are sold through listings. Each listing is classified within a category hierarchy according to the type of the item offered. For example, a listing of an iPod classic, 60 GB might belongs to the following category branch: Consumer Electronics → MP3 Players → Apple iPod → Apple iPod, 60 GB. The category of an item is provided by the seller so there is a certain proportion of miscategorized items. For example in my data that comes from "Apple iPod, 60 GB" there are some listings for different iPod models, some listings for iPod accessories, some MP3 players by other firms etc. However most items are classifies according to their actual type.

There are several types of listings. One of them – the standard auction listing is discussed in detail below. Other types include: Posted Price listings, Multiple item auctions, Live Auctions and others.

The standard auction listing may apply to a single item or any bundle. The rules are the following. The listing has fixed expiration time. Typically, a listing lasts for 1, 3, 5, 7 or 10 days. Each listing constitutes a webpage, that contains the following information about the item being sold: a unique identifier, item title, item description (often with photos), seller's identity and feedback score, start time, expiration time, the current outstanding price and the list of bids submitted with the bid times and identities of the bidders.

Each seller of a standard auction may use the two options to set a reserve price for an item. First, there she may choose a minimal bid that would be accepted for the item. This minimal bid is posted on the item's page. Second, there is a possibility to set a "secret reserve" price that works as follows. If the "secret reserve" option is chosen and the price set by the seller, it is not displayed to the public. However as bids are

[19]The contract conditions do not allow me to reveal the name of the market in writing

submitted for an item the bidders are informed whether or not the secret reserve level have been exceeded by the current highest bid. If all the bids submitted happen to be below secret reserve the item is not sold.

The website is using "proxy bid" system that works as follows. When asked to submit a bid a bidder is informed that what he is submitting is a "proxy" bid, that will be used by the auction automated system in order to bid on the bidder's behalf. The system will be submitting the bids on behalf of the bidder that are just high enough for the bidder to remain the winner in the auction, but not exceeding the proposed proxy. There is a minimal bid increment that is determined externally by the system depending on the range of the bids.

The work of the system can be illustrated by an example. Suppose an item has no secret reserve and the minimal bid is set to be \$1. Suppose, the minimal bid increment is \$.5. Any bidder may submit a bid of \$1 or higher. Suppose, Adam submits a proxy bid of \$15 to the system. Then the system works on his behalf and submits the minimal bid on his behalf that would make him a winner: \$1. This bid is revealed to the public. The price of the good is now \$1 and all bidders may submit any bids at or above \$1.5. Suppose, Bertha submits a bid of \$5. Then the system rises the bid submitted on behalf of Adam to \$5.5 – the lowest value that keeps him a winner over Bertha. The highest bid the system may submit for Bertha is \$5 but it does not make her a winner. She receives a message that she have been overbidden in the auction. The item price turns to \$5.5. Now the bidders may submit any bid of \$6 or above. If the auction expires at this point – Adam wins the auction and pays the price of \$5.5. Notice, that this scheme is essentially a second-bid pricing scheme, since the price paid by the winner is determined by the second highest bid submitted (plus the minimal increment).

There is a feedback system operating on the website. After each completed transaction both parties are offered an opportunity to leave a feedback about each other. The feedback consists of the evaluation: positive, negative or neutral and a brief comment explaining it. The feedbacks of all users are publicly displayed on the website. Through many transactions each user gets a list of feedbacks that are summarized in a feedback score that is computed as the number of positive feedback comments minus the number of negative feedbacks. This number is displayed next to seller's identity and accompanies each listing. It is used to show the bidder how reliable is the seller of a given item and therefore helps to evaluate the item.

B Proof of Lemma 3 and Proposition 3

.

By Lemma 2 the optimal bid of the bidders is given by

$$B(\theta) = \theta + \beta X + \varepsilon - \tilde{\delta} V(\theta, Z). \tag{55}$$

Here Z is the vector of all the information about the future items available to the bidder. If the bidder can see N items forward, then $Z = (X_j, \varepsilon_j)_{j=1}^N$. Denote $\tilde{\theta} = \theta - \tilde{\delta} V(\theta, Z)$. Thus $\tilde{\theta}$ might be a function of Z.

The bidder's Bellman equation is

$$V(\theta, X, \varepsilon, Z') = \alpha W(\theta, X, \varepsilon) + \tilde{\delta}(\alpha P_{lose}(\theta, X, \varepsilon) + (1 - \alpha)) E_{Z''} V(\theta, Z', Z'') \tag{56}$$

Here Z' is all the information in Z except for the current item's characteristics (X, ε), Z'' is the information that becomes available between bidding for the current item and the next one. I assume that when the value function V is computed, the realization of the activity status of the bidder for the bid is not jet observed. This equation says that with probability α the bidder will be active. Then the payoff from the current item is determined by

$$W(\theta, X, \varepsilon) = \int\limits_0^{\tilde{\theta}+\beta X+\varepsilon} (\theta + \beta X + \varepsilon - p) dG(p, X, \varepsilon); \tag{57}$$

the continuation value $E_{Z''} V(\theta, Z', Z'')$ will be realized if either the bidder is being active, but loose the current auction and survives until the next one, or the bidder is not active in the current auction and thus receives the continuation value whenever she survives in the market. The probability to lose is

$$P_{loose}(\theta, X, \varepsilon) = 1 - G(\tilde{\theta} + \beta X + \varepsilon). \tag{58}$$

Notice, that the characteristics of the current item only affect (56) through the terms $W(\theta, X, \varepsilon)$ and $P_{lose}(\theta, X, \varepsilon)$. Thus, it is enough to show that the expressions $W(\theta, X, \varepsilon)$ and $P_{lose}(\theta, X, \varepsilon)$ do not depend on (X, ε), then by induction one can show that the further items in Z do not affect the expression as well.

Indeed, by Assumption 2

$$P_{lose}(\theta, X, \varepsilon) = 1 - G(\tilde{\theta} + \beta X + \varepsilon) = 1 - \tilde{G}(\tilde{\theta}), \tag{59}$$

might only depend on the future terms in Z.

Furthermore, by Assumptions 2 and 3

$$W(\theta, X, \varepsilon) = \int_{0}^{\tilde{\theta}+\beta X+\varepsilon} (\theta + \beta X + \varepsilon - p) dG(p, X, \varepsilon) = \int_{0}^{\tilde{\theta}} (\theta - p) d\tilde{G}(p) \tag{60}$$

thus this expression also does not depend on the current value of (X, ε).

Therefore we have shown that $V(\theta, X, \varepsilon, Z') = V(\theta, Z')$. Therefore the terms W and P_{lose} do not depend on the first coming auction characteristics, since $\tilde{\theta} = \theta - \tilde{\delta} V(\theta, Z)$ does not depend in them. Also the continuation value in (56) does not depend on this term. Etc. By induction no information in Z may affect the value function in (56).

To prove the proposition solve the remaining system of equations:

$$\tilde{\theta} = \theta - \tilde{\delta} V(\theta) \tag{61}$$

$$V(\theta) = \alpha W(\theta) + \tilde{\delta}(\alpha P_{lose}(\theta) + (1 - \alpha)) V(\theta) \tag{62}$$

$$P_{lose}(\theta) = 1 - \tilde{G}(\tilde{\theta}) \tag{63}$$

$$W(\theta) = \int_{0}^{\tilde{\theta}} (\theta - p) d\tilde{G}(p) \tag{64}$$

to get the equation (25).

C Identification in the Extended Model

The identification results parallel to those established for the Baseline Model hold true for the Extended Model under assumptions 9. I formulate the results for the more general case of heterogeneous items.

For the case of heterogeneous items, θ denote the bidder's type, so that an item valuation by the bidder is given by (16). The distribution of bidders' types in the population of bidders is denoted by $\mathcal{F}(\cdot)$. For a bidder of the type θ the optimal bid for an item with the characteristics (X, ε) is determined by the item characteristics and the bidder's transformed type $\tilde{\theta}$ given by (23). The transformed type can be determined as the bidder's bid for an item cleaned from the item characteristics.

The same reasoning as in the Baseline Model allows me to establish auction characteristics neutrality (given assumptions (2-3)).

The following results can be established.

Lemma 7. *Consider a stationary Extended Model with heterogeneous items, known parameters and zero reserve. Suppose, the bidders' strategy is determined by Assumption 9. Assume, the transformed type distribution function F is strictly increasing and continuous. There exist two strictly increasing transformation functions $\tilde{g} : [0, 1] \to [0, 1]$ and $\tilde{h} : [0, 1] \to [0, 1]$ such that for any b*

$$\tilde{G}(b) = \tilde{g}(F(b)), \tag{65}$$

$$\tilde{H}(b) = \tilde{h}(F(b)); \tag{66}$$

The transformation functions $\tilde{g}(\cdot)$ and $\tilde{h}(\cdot)$ are determined by the model structure, but do not depend on the distribution F.

Proof. Although the structure of the bidding process is different, the exact same reasoning as in Lemma 4 may be applied to prove this result □

Based on this Lemma the following result can be established.

Proposition 11. *Suppose, the bidders' strategy in the Extended Model is determined by Assumption 9. In a heterogeneous item model with known parameters the population transformed type distribution F is* *nonparametrically identified either from the first, or from the second highest bid data, and item observables.*

If the reserve price in the auctions is drawn from a known distribution $R(\cdot)$ one can identify the parent distribution $F(\cdot)$ as a fixed point.

Proposition 12. *Suppose, the bidders' strategy in the Extended Model is determined by Assumption 9. Let $R(\cdot)$ be the distribution of the reserve prices in the auctions.*

1. *For each distribution F' of the optimal bids of the bidders in the auctions the transformation functions $g(\cdot|F')$ and $h(\cdot|F')$ can be identified that relate the distribution F' and the resulting distributions G' and H' of the first and second highest bids in the auctions:*

$$G' = g(\cdot|F')$$
$$H' = h(\cdot|F').$$

2. *If the distribution G of the first highest bids in the auctions is identified then the parent distribution F can be identified as the fixed point of the transformation γ:*

$$\gamma(F) = g^{-1}(G|F). \tag{67}$$

3. *If the distribution H of the second highest bids in the auctions is identified then the parent distribution F can be identified as the fixed point of the transformation η:*

$$\eta(F) = h^{-1}(H|F). \tag{68}$$

The equations (67) and (68) can be solved numerically by iteration.

D Proof of Proposition 10

.

First, the optimal bid, cleaned of the observables effect, is determined as

$$B - \beta X - \varepsilon = \tilde{\theta} = \theta - \tilde{\delta} V(\theta).$$

From the other side, for each bidder the value of the problem is obtained from the occasional winning of an item, therefore in general case the value of the problem can be determined as

$$V(\theta) = \int_{\Omega} (\theta - p(\omega))\mathbb{1}(win)d\omega =$$

$$\theta \cdot P(win|\theta) - E(p|win, \theta) \cdot P(win|\theta) =$$

$$(\theta - E(p|win, \theta))P(win|\theta);$$

here *win* indicates the event of the bidder winning an item in one of the auctions before exit. Therefore,

$$\tilde{\theta} = \theta(1 - \tilde{\delta}P(win|\theta)) + \tilde{\delta}P(win, \theta)E(p|win, \theta)$$

and

$$\theta = \frac{\tilde{\theta} - \tilde{\delta}P(win|\theta)E(p|win, \theta)}{1 - \tilde{\delta}P(win|\theta)}. \tag{69}$$

Assuming a one-to-one relationship between $\tilde{\theta}$ and θ (that is later verified) I can change the conditioning:

$$\theta = \frac{\tilde{\theta} - \tilde{\delta}P(win|\tilde{\theta})E(p|win, \tilde{\theta})}{1 - \tilde{\delta}P(win|\tilde{\theta})}$$

Item_id	Item unique ID number
Start_price_usd	Minimal bid posted by the seller
Reserve_price_flag	Indicates whether or not Secret Reserve
	was used for this item
Success_yn	Indicated whether the item have been sold successfully
Color	1 for Black, 0 for White
seller_fb	Feedback score of the seller
end_period	Number of 10-minute period when the item ends
first_bid	The size of the highest bid in the auction
second_bid	The size of the second highest bid in the auction
new_sealed	1 for new, sealed items with no extras; 0 for others
Final_price_usd	The price of the item at the time of expiration
first_bidder	The identity of the winner (if applies) or 0.

Table 1: The field list in the item-level dataset

Variable	mean	std	min	max
Start_price_usd	77.2902	125.8965	0.01	801
Reserve_price_flag	0.1130	0.3167	0	1
Success_yn	0.8692	0.3372	0	1
Color	0.5635	0.496	0	1
seller_fb	10958.98	23209.64	-1	144011
first_bid	305.7898	108.2641	0	1000
second_bid	283.7928	107.3866	0	760
Final_price_usd	294.1150	101.0361	0	776
new_sealed	0.1332	0.3398	0	1

Table 2: Full sample summary statistics ($N = 8325$).

Variable	mean	std	min	max
Start_price_usd	72.1162	124.2848	0.01	444.99
Reserve_price_flag	0.1001	0.3003	0	1
Success_yn	0.8999	0.3003	0	1
Color	0.6492	0.4774	0	1
seller_fb	446.9197	1102.90	0	13561
first_bid	326.5333	97.6583	0	600
second_bid	309.0358	97.0219	0	600
Final_price_usd	316.7962	92.9398	0	600
new_sealed	1	0	1	1

Table 3: New-Sealed sample summary statistics ($N = 1109$).

Variable	mean	std	min	max
$Times\_active$	5.4332	12.2740	1	536
$Items\_active$	2.6189	6.2362	1	346
$Items\_won$	0.2467	0.6233	0	59
$Time\_in\_sample(hours)$	76	230	0	2302
$Time\_in\_sample \mid Repeated\_observation(hours)$	112	271	0	2302

Table 4: Bidders summary statistics ($N = 29239$).

bid	Coef.	Std. Err.	t	$P > \|t\|$	[95% Conf. Interval]	
color	10.04636	1.366918	7.35	0.000	7.365615	12.72711
seller_fb	.001882	.0005676	3.32	0.001	.000769	.0029951
time	-.0008554	.0001745	-4.90	0.000	-.0011976	-.0005133
first	14.98108	1.293274	11.58	0.000	12.44476	17.5174
cons	336.7828	1.775342	189.70	0.000	333.301	340.2645

Number of obs = 1985, R-squared = 0.0973

Table 5: OLS Results for the joint pool of second and first highest bids

Parameter	Mean	Standard Deviation
α	0.0172	0.0003
δ	0.9976	0.0001
λ	2.5104	0.0179
$\lambda_{i,sold}$	0.5459	0.0026
λ_i	0.6293	0.0008

Table 6: Parameter Estimates

	p	std	Δ%	q	std	Δ%	TCS	std	Δ%	CS	std
Strategic Bidders											
AU	341.88	2.20	0.0	7236	31	0.0	569,849	35,477	0.0	78.76	4.89
H2	356.19	2.63	−4.0***	7369	81	−1.8*	741,158	152,082	−22.1	100.54	20.44
H1	354.83	2.75	−3.7***	7345	86	−1.5	557,443	37,550	−0.2	75.90	5.04
AV2	357.10	2.94	−4.3***	7643	54	−5.3***	761,192	154,995	−22.9	99.58	20.24
AV1	355.77	3.25	−3.9***	7635	54	−5.2***	581,843	38,086	−2.7	76.21	4.95
LO2	356.29	2.64	−4.0***	7951	62	−9.0***	831,426	171,347	−26.4*	104.57	21.55
LO1	354.89	2.84	−3.6***	7925	67	−8.7***	635,853	41,191	−7.3**	80.24	5.24
LI2	347.93	2.63	−1.8***	8325	0	−13.1***	832,506	154,608	−29.9**	100.00	18.57
LI1	347.75	2.86	−1.7***	8325	0	−13.1***	643,722	36,872	−12.1***	77.32	4.43
Myopic Bidders											
AU	341.88	2.20	0.0	7236	31	0.0	100,617	10,957	0.0	13.91	1.52
H2	339.46	1.93	0.7***	7108	88	1.8	186,213	30,214	−44.5***	26.20	4.23
H1	339.55	1.97	0.7**	7110	87	1.8	143,917	8,542	−29.9***	20.24	1.18
AV2	338.25	2.33	1.0***	7499	60	−3.5***	202,149	32,502	−48.9***	26.45	4.25
AV1	338.51	2.40	1.0***	7501	60	−3.5***	158,771	9,429	−36.4***	20.80	1.23
LO2	339.13	1.91	0.9***	7689	69	−5.9***	199,387	32,694	−48.4***	25.93	4.26
LO1	339.21	1.98	0.8***	7690	69	−5.9***	156,583	9,492	−35.8***	20.36	1.22
LI2	333.42	2.31	2.5***	8325	0	−13.1***	235,682	31,971	−56.5***	28.31	3.84
LI1	334.23	2.46	2.3***	8325	0	−13.1***	188,982	9,717	−46.6***	22.70	1.17

Table 7: Baseline Model Estimation Results. See Table 9 for notation. The stars indicate significance level of the difference between the variable value in the auction market and the corresponding estimate of the competitive counterpart. *** indicates 1% significance, ** 5% significance, * 10% significance difference according to the bootstrap distribution.

	p	std	Δ%	q	std	Δ%	TCS	std	Δ%	CS	std
						Strategic Bidders					
AU	341.82	2.21	0.0	7237	31	0.0	615,406	89,007	0.0	85.04	12.29
H2	346.72	2.67	−1.4***	7250	85	−0.2	1,081,542	781,620	−35.7*	149.20	108.25
H1	344.76	2.99	−0.9**	7232	89	0.1	712,719	72,176	−13.3**	98.56	9.98
AV2	345.78	2.72	−1.2***	7570	55	−4.4***	1,088,889	823,671	−36.6*	143.85	108.69
AV1	344.34	2.91	−0.7*	7557	57	−4.2***	728,099	83,901	−16.7**	96.36	11.10
LO2	346.68	2.69	−1.4***	7837	60	−7.6***	1,191,044	926,707	−40.8**	151.99	117.95
LO1	344.77	2.95	−0.8*	7817	66	−7.4***	771,663	94,626	−19.9***	98.73	12.17
LI2	336.10	2.67	1.7***	8325	0	−13.1***	1,163,971	823,668	−41.6**	139.82	98.94
LI1	336.35	2.57	1.6***	8325	0	−13.1***	789,938	83,560	−23.2**	94.89	10.04
						Myopic Bidders					
AU	341.82	2.21	0.0	7237	31	0.0	100,787	10,741	0.0	13.93	1.48
H2	339.28	1.94	0.7***	7250	85	1.8	195,717	31,949	−47.1***	27.54	4.49
H1	338.51	2.21	1.0***	7232	89	1.9	153,541	9,437	−34.3***	21.62	1.29
AV2	337.92	2.33	1.1***	7570	55	−3.5***	210,102	33,843	−51.0***	27.76	4.48
AV1	337.13	2.54	1.4***	7557	57	−3.4***	170,124	9,656	−40.9***	22.51	1.28
LO2	339.21	1.82	0.8***	7837	60	−5.9***	211,449	34,727	−50.9***	27.49	4.52
LO1	338.46	2.13	1.0***	7817	66	−5.8***	166,396	9,955	−39.1***	21.65	1.29
LI2	331.64	2.43	3.1***	8325	0	−13.1***	256,709	33,693	−60.2***	30.84	4.05
LI1	331.81	2.42	3.0***	8325	0	−13.1***	209,461	10,205	−52.0***	25.16	1.23

Table 8: Extended Model Estimation Results. See Table 9 for notation. The stars indicate significance level of the difference between the variable value in the auction market and the corresponding estimate of the competitive counterpart. *** indicates 1% significance, ** 5% significance, * 10% significance difference according to the bootstrap distribution.

Table 9: Notation in Tables 7-8

AU	Auction market
H2	Upper bound for supply estimate; second bids used in demand estimation
H1	Upper bound for supply estimate; first bids used in demand estimation
AV2	Average between upper and lower bounds for supply estimate; second bids used in demand estimation
AV1	Average between upper and lower bounds for supply estimate; first bids used in demand estimation
LO2	Lower bound for supply estimate; second bids used in demand estimation
LO1	Upper bound for supply estimate; second bids used in demand estimation
LI2	Limit case estimate for strategic sellers second bids used in demand estimation
LI1	Limit case estimate for strategic sellers first bids used in demand estimation
p	Price
q	Quantity
TCS	Total consumer surplus
CS	Consumer surplus per item sold
Δ%	Average percentage difference between auction market value and the corresponding competitive estimate

Variable	Perfect Competition	Current Mechanism	Second Best Mechanism
Price	\$343.55	\$341.78	\$352.88 (0.42)
Quantity	7552	7237	7636 (23.57)
CS	\$739,771	\$622,761	\$662,941 (22,756)
PS	\$2,201,354	\$1,901,459	\$2,273,397 (13,038)
SS	\$2,941,125	\$2,524,220	\$2,936,337 (26,188)
$\%\Delta CS$	0	−15.82%	−10.73% (0.73)
$\%\Delta PS$	0	−13.62%	4.01% (0.38)
$\%\Delta SS$	0	−14.18%	−0.24% (0.09)

Table 10: Second Best Mechanism Simulation. CS stands for Consumer Surplus, PS for Seller Surplus, and SS for Social Surplus. $\%\Delta CS$ represents deviation from the Perfect Competition in Consumer Surplus, $\%\Delta PS$ in Seller Surplus, $\%\Delta SS$ in Social Surplus. Standard deviations for Second Best Mechanism are given in presentences.

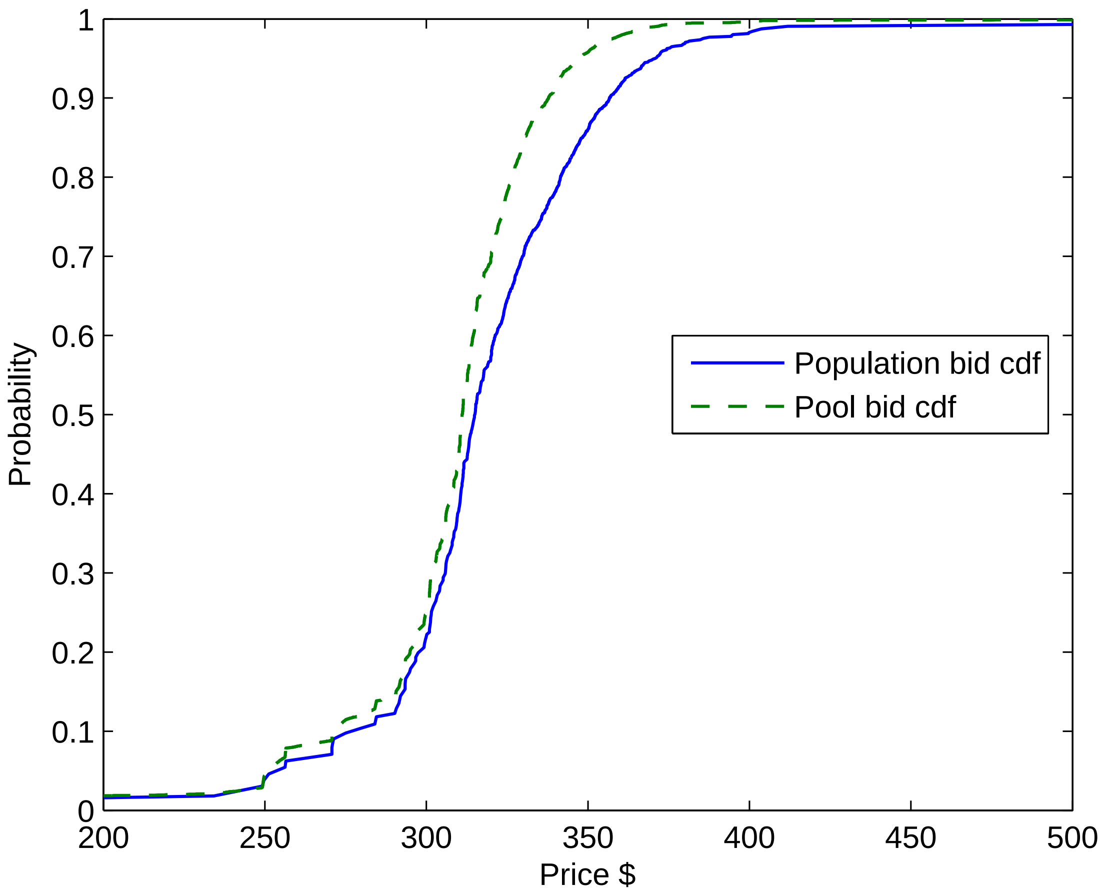

Figure 1: Population and Pool distributions

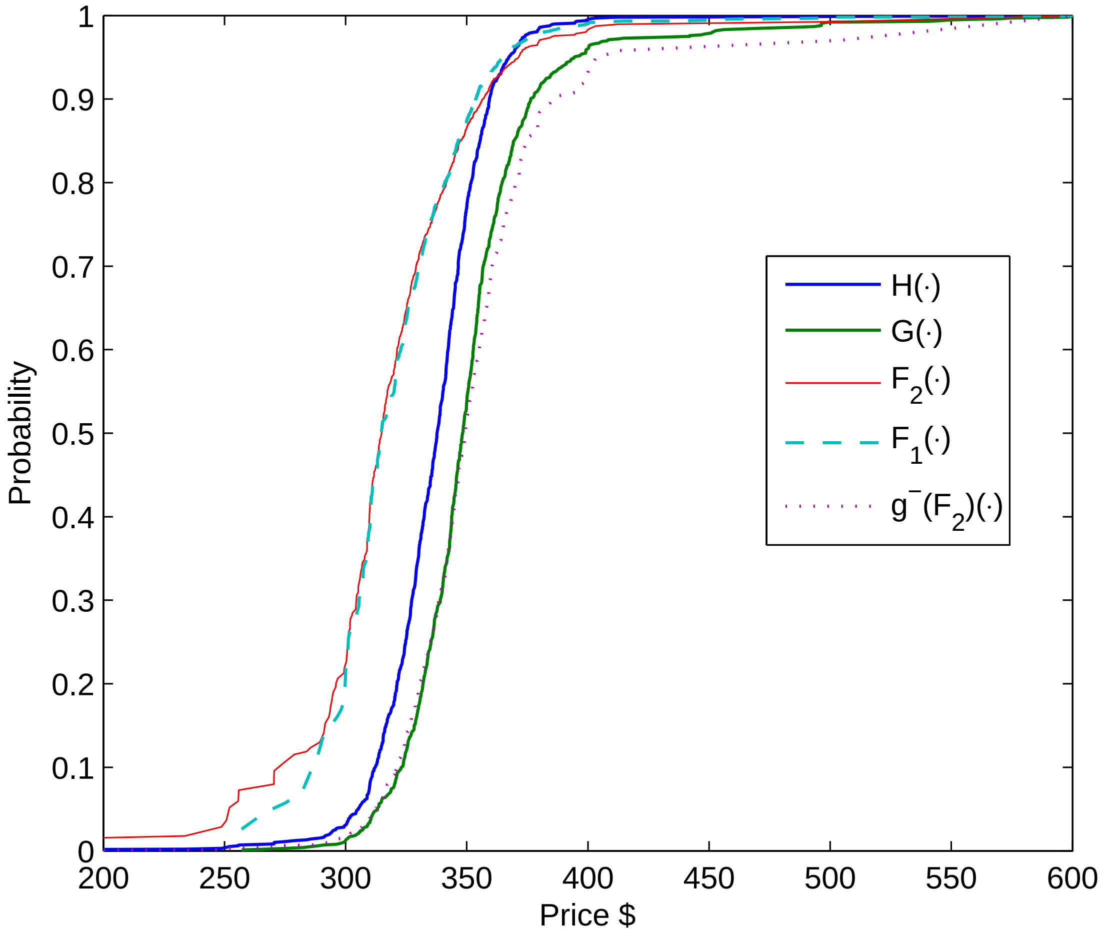

Figure 2: Estimated Distributions: Baseline Model

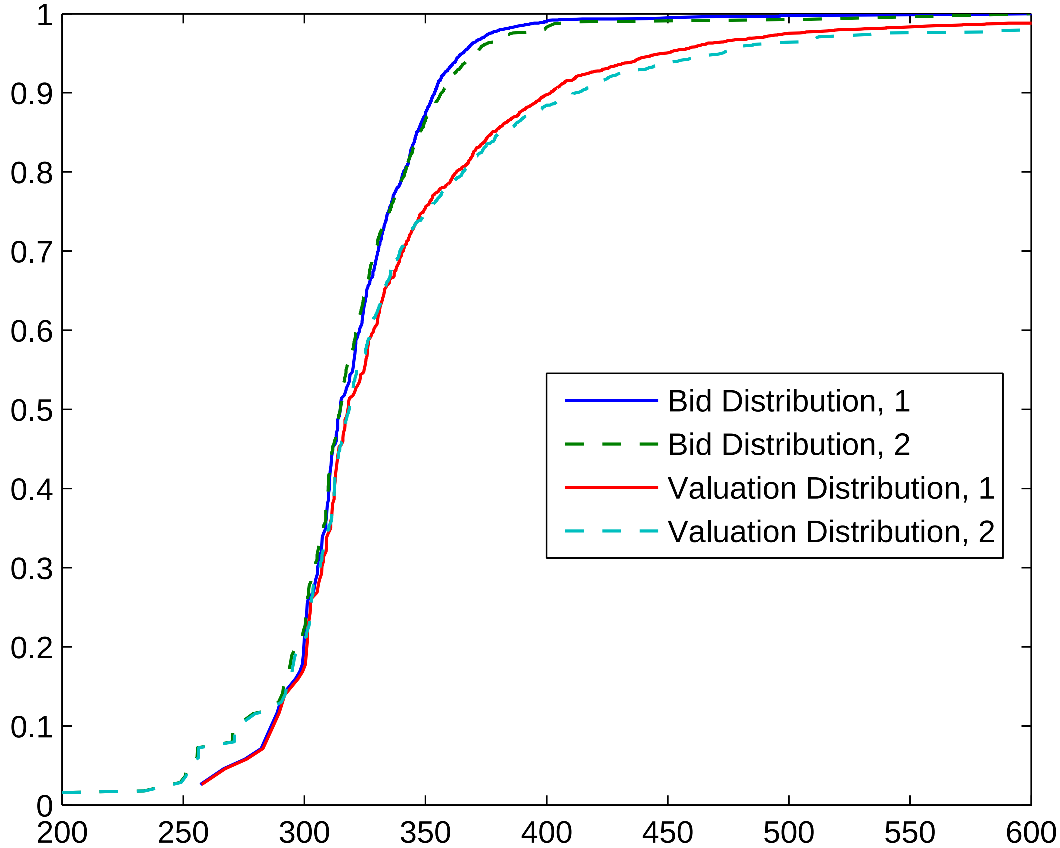

Figure 3: Estimated Parent Distributions: Baseline Model

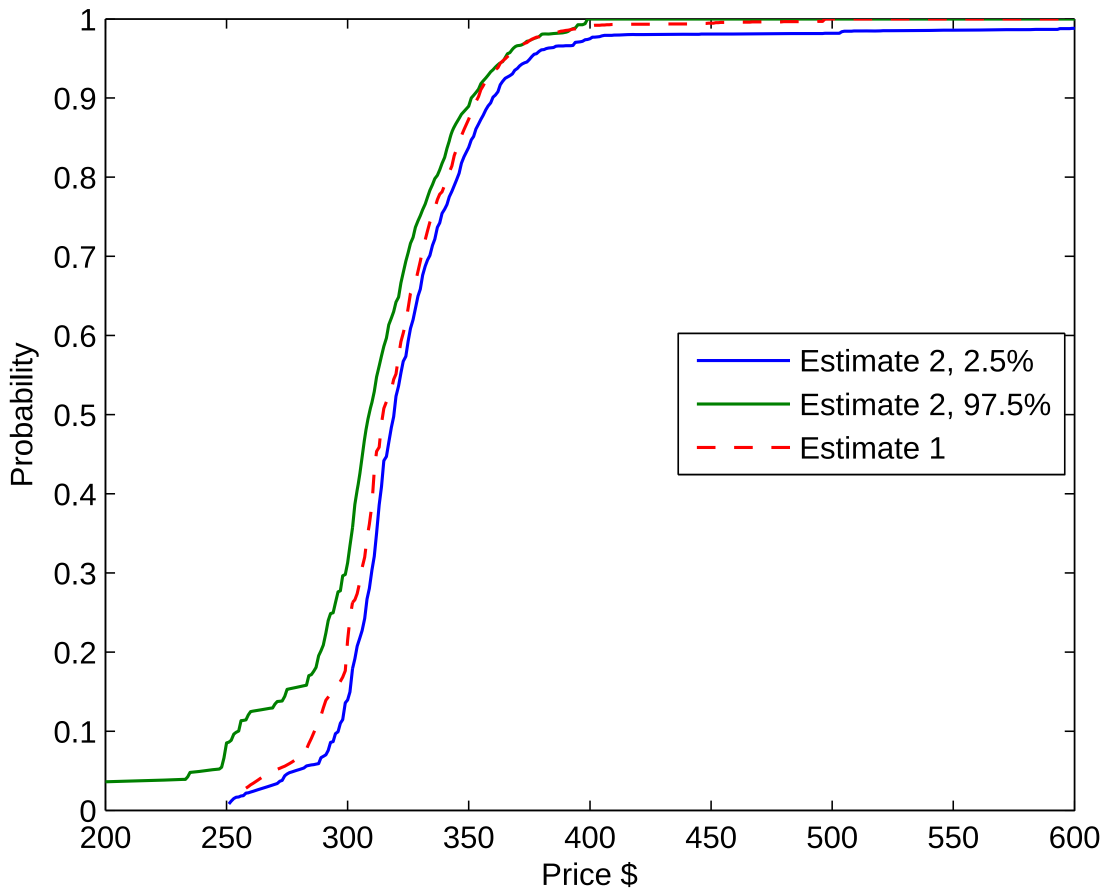

Figure 4: Bootstrap Estimate, Baseline Model: Bid Distribution

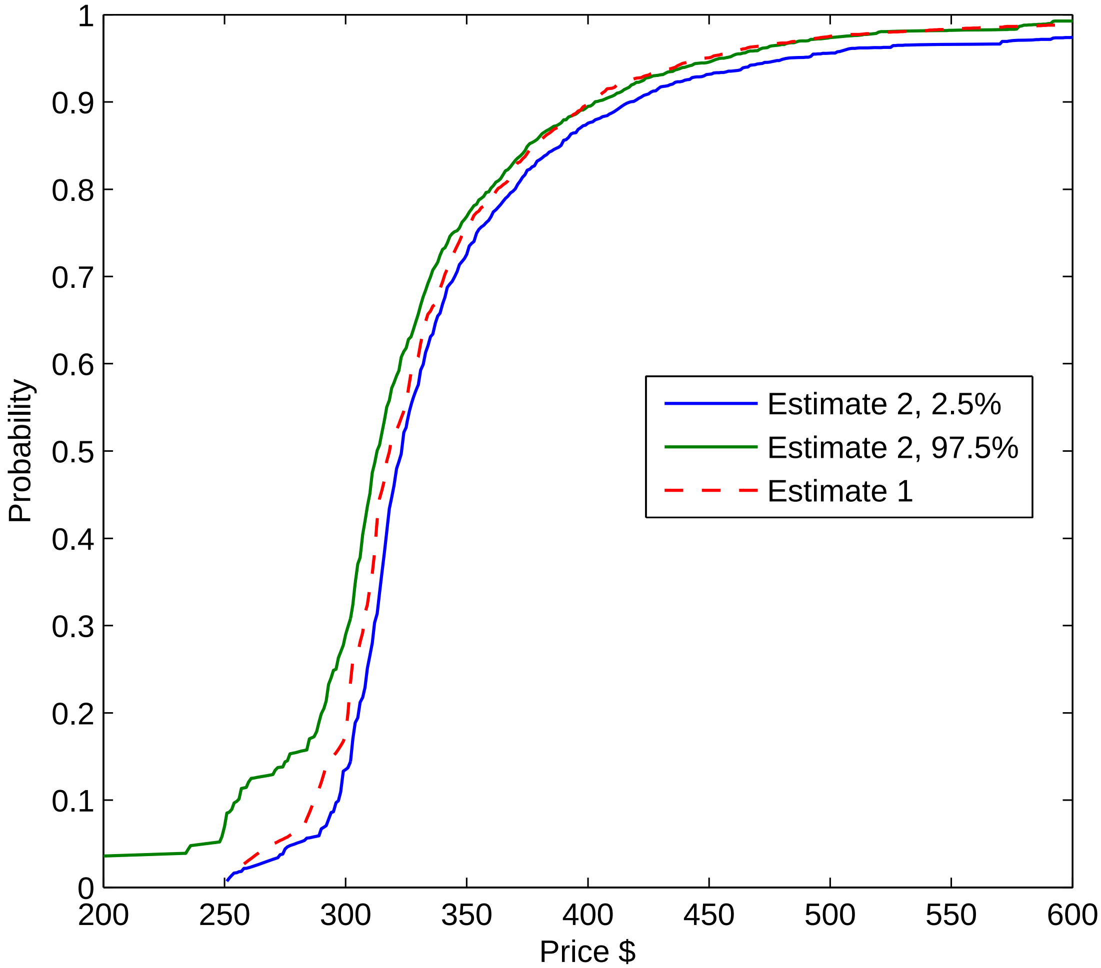

Figure 5: Bootstrap Estimate, Baseline Model: Valuation Distribution

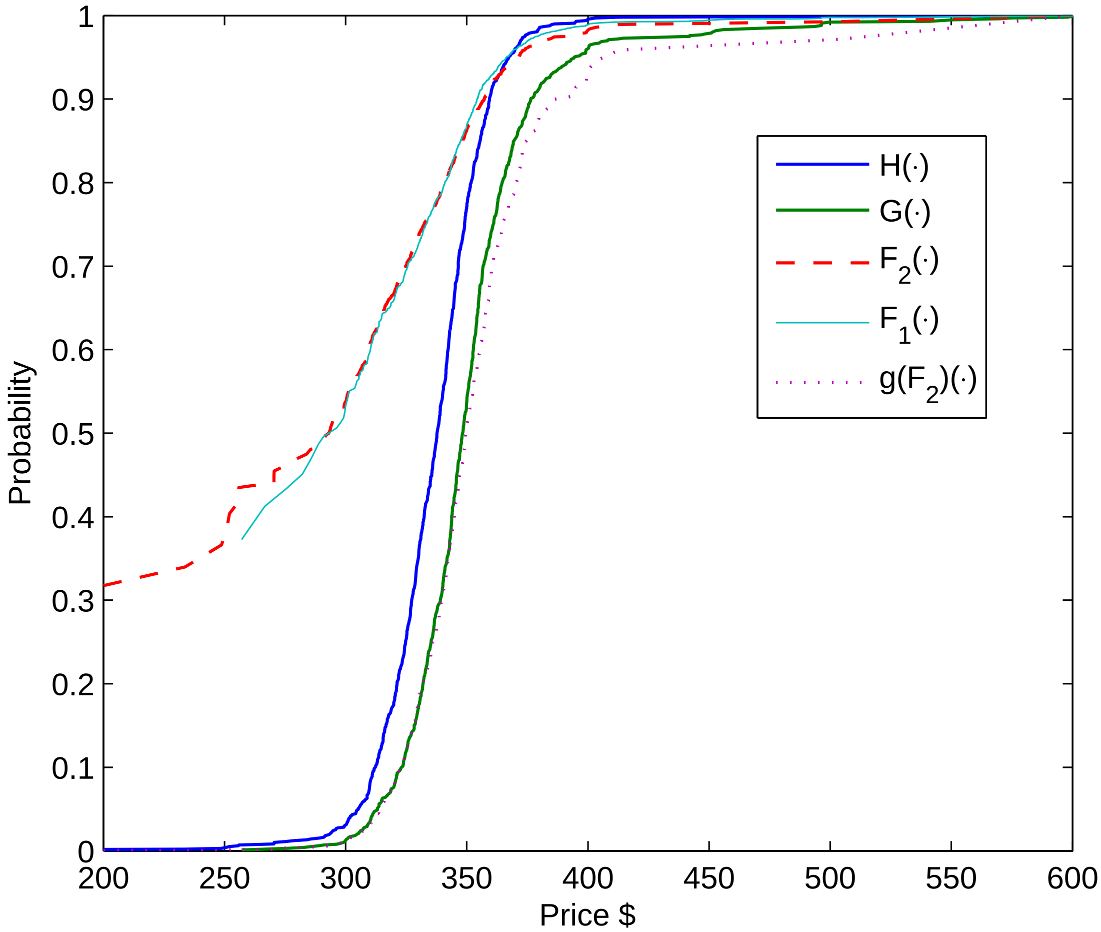

Figure 6: Estimated Distributions: Extended Model

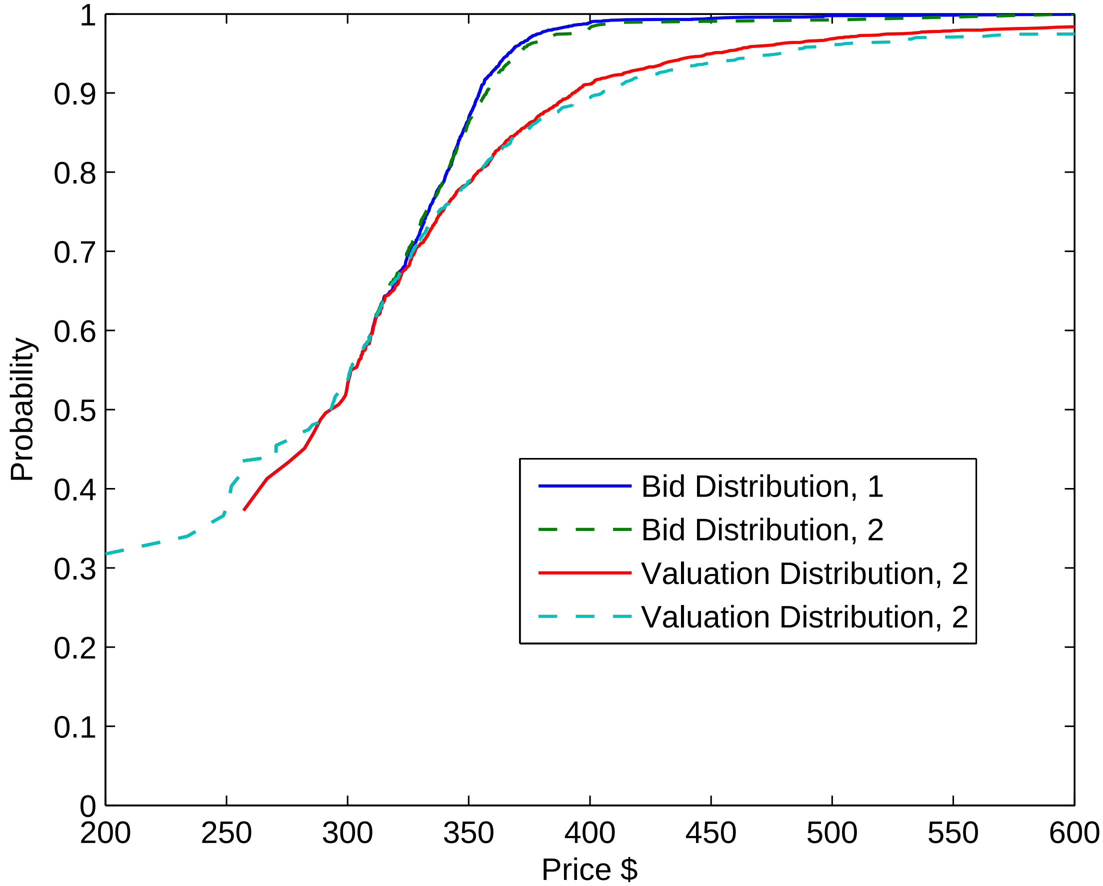

Figure 7: Estimated Parent Distributions: Extended Model

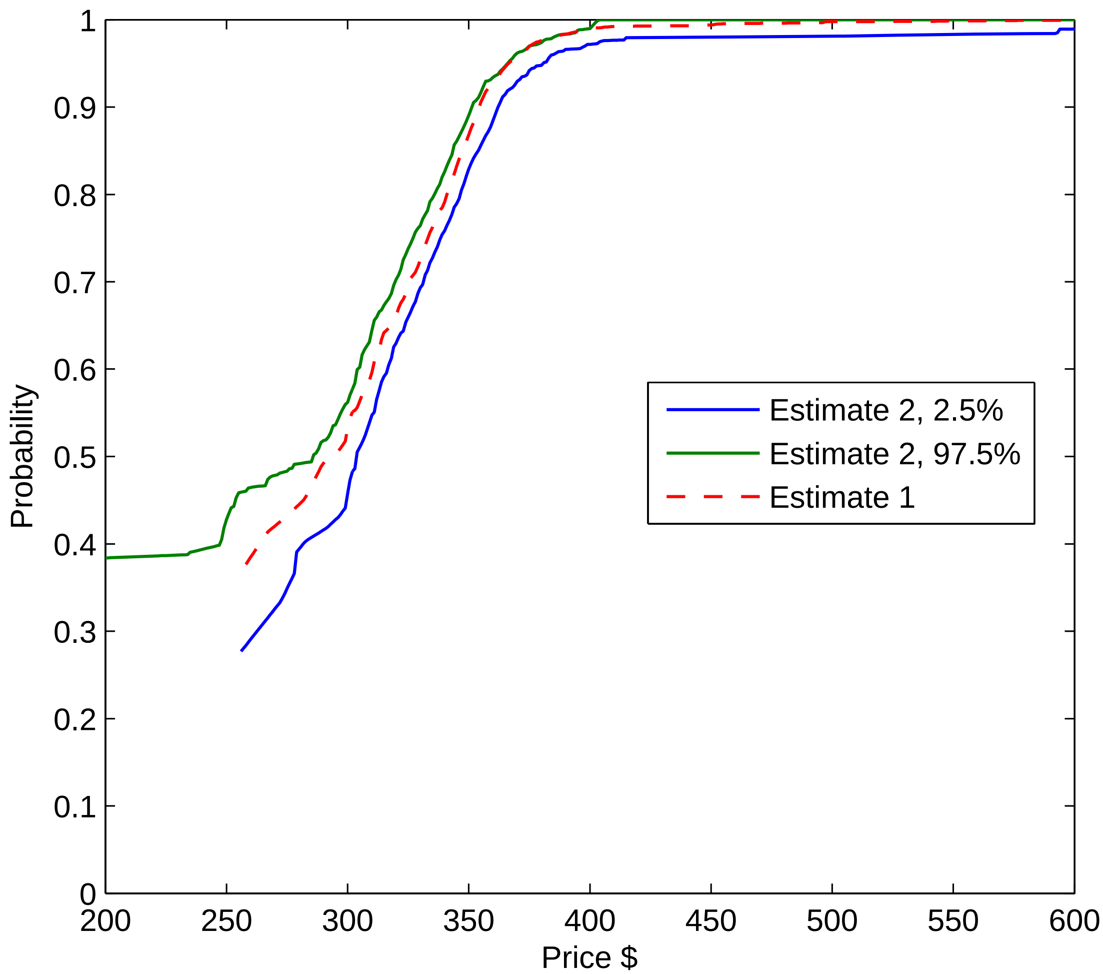

Figure 8: Bootstrap Estimate, Extended Model: Bid Distribution

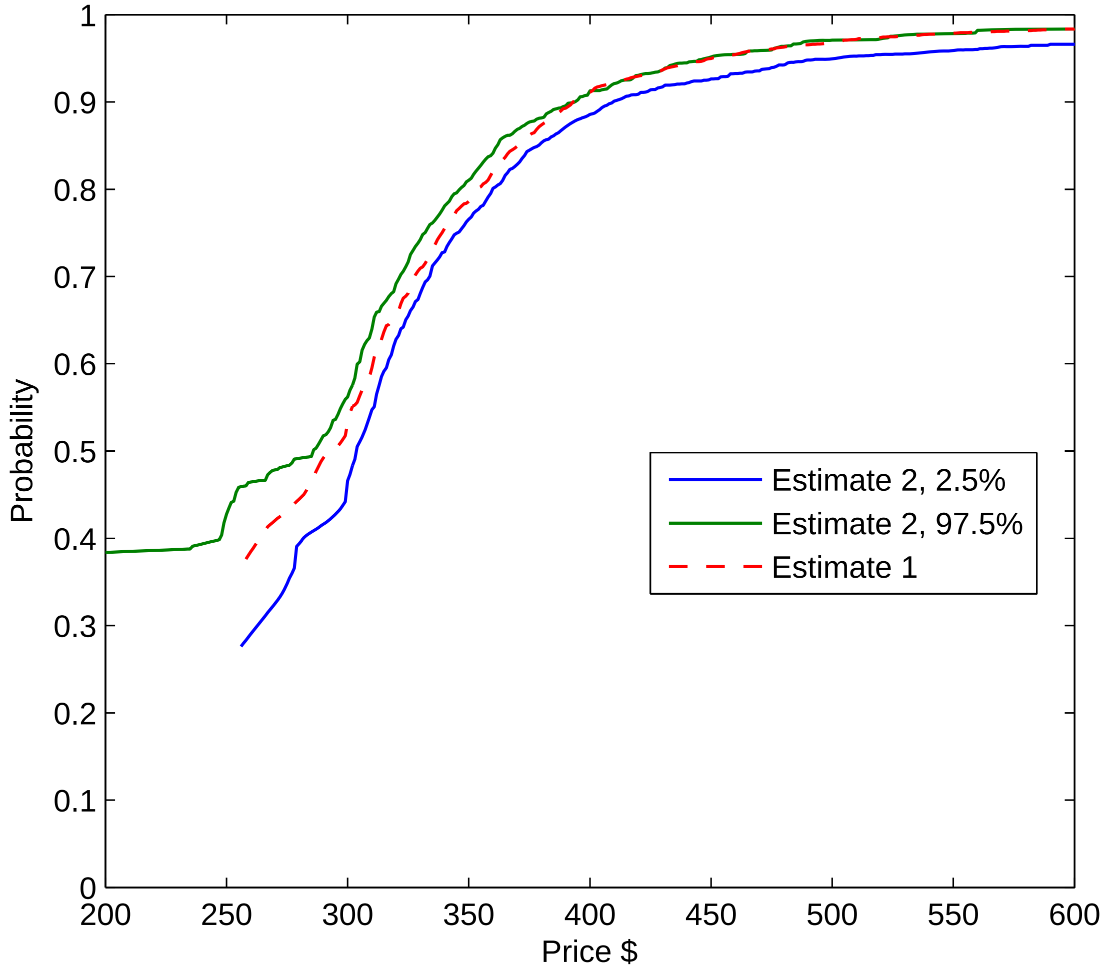

Figure 9: Bootstrap Estimate, Extended Model: Valuation Distribution